Πανεπιστήμιο Κρήτης
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

University of Crete

Faculty of Medicine

"ΑΛΕΞΑΝΔΡΟΣ ΦΛΕΜΙΓΚ"
Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών

PhD Thesis

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Generation of a conditional by inversion *Sox2* allele

«Δημιουργία ενός επαγώμενου με αναστροφή *Sox2* αλληλίου»

Νικόλαος Μάνδαλος

ΗΡΑΚΛΕΙΟ

ΝΟΕΜΒΡΙΟΣ 2012

This thesis is dedicated to my parents Theochari and Achilleas

and my late brother Miltiadis

The studies presented in this thesis were performed in the Stem Cell Biology Laboratory of Dr Eumorphia Remboutsika, Institute of Molecular Biology and Genetics, Biomedical Sciences Research Centre “Alexander Fleming”, 34 Fleming Str., 16672-Vari, Greece. This work was financially supported by BILATERAL COOPERATION GRANT (NON-EU82) from the General Secretariat of Research and Technology and it was performed under the auspices of a collaboration agreement with REGENERON Pharmaceuticals Inc. in New York, USA.

Επιβλέπουσα Ερευνήτρια: Ευμορφία Ρεμπούτσικα, Ερευνήτρια Γ΄ ΕΚΕΒΕ Φλέμινγκ

Τριμελής Συμβουλευτική Επιτροπή:

Θερμού Κ., Καθηγήτρια τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Ταβερναράκης Ν., Ερευνητης Α΄ (IMBB)

Καρδάσης Δ., Καθηγητής τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Επταμελής Συμβουλευτική Επιτροπή:

Θερμού Κ., Καθηγήτρια τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Ταβερναράκης Ν., Ερευνητης Α΄ (IMBB)

Καρδάσης Δ., Καθηγητής τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Πάχνης Β., Καθηγητής τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Γραβάνης Α., Καθηγητής τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Καραγωγέως Δ., Καθηγήτρια τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Χαραλαμπόπουλος Ι., Καθηγητής τμήματος Ιατρικής Πανεπιστημίου Κρήτης

Acknowledgements

This thesis could not be completed without the powerful help of Dr. Eumorphia Remboutsika who gave me the chance to work in her lab. She offered me the best quality of supervising I had through my career, a great project of high standards and more importantly the proper “mentoring” to learn how to think and act as a scientist. I truly wish more people to have the chance to study in Dr. Remboutsika’s lab. I will cherish the PhD years in Dr. Remboutsika’s lab as the best working years I could ever imagine.

Table of Contents

Table of Contents

Summary

Sox2 encodes a transcription factor that harbours a DNA binding HMG domain. It is considered one of the key gene players for the regulation of pluripotency and early embryonic development, and a determinant of cell fate during development and homeostasis. During early development, *Sox2* is required for the formation of the epiblast. However, its role in stem cells is still elusive. Here, we performed a surgical ablation of *Sox2* in the epiblast using a novel Conditional by Inversion *Sox2* allele (*Sox2*^{COIN}). The Conditional by Inversion (COIN) method for engineering conditional alleles relies on an invertible optimized gene trap-like element, the COIN module, for imparting conditionality. The COIN module contains an optimized 3' splice site-polyadenylation signal pair, but is inserted antisense to the target gene and therefore does not alter transcription, until it is inverted by Cre recombinase. In order to make COIN applicable to all protein-coding genes, the COIN module has been engineered within an artificial intron, enabling insertion into an exon. Therefore, theoretically, the COIN method should be applicable to single exon genes, and to test this idea we engineered a COIN allele of *Sox2*. This single exon gene presents design challenges, in that its proximal promoter and coding region are entirely contained within a CpG island, and are also spanned by an overlapping transcript, *Sox2Ot*, which contains *mmu-miR1897*. Here, we show that despite disruption of the CpG island by the COIN module intron, the COIN allele of *Sox2* (*Sox2*^{COIN}) is phenotypically wild type, and also does not interfere with expression of *Sox2Ot* and *miR1897*. Furthermore, the inverted COIN allele of *Sox2*, *Sox2*^{INV} is functionally null, as homozygotes recapitulate the phenotype of *Sox2*$^{ßgeo/ßgeo}$ mice, a well-characterized *Sox2* null. Lastly, the benefit of the eGFP marker embedded in the COIN allele is demonstrated as it mirrors the expression pattern of *Sox2*. Epiblast-inversion of the *Sox2*^{COIN} allele generates normal heterozygote *Sox2*$^{INV/+}$ adult animals and *Sox2*$^{INV/+\ (H)}$ haploinsufficient mutant embryos. *Sox2*$^{INV/+\ (H)}$ haploinsufficient and *Sox2*$^{INV/mosaic}$ embryos exhibit heart-looping defects, suffer from bradycardia, myocardium malformation and bleeding and die around E11. *Sox2*$^{INV/+(H)}$ haploinsufficient and *Sox2*$^{INV/mosaic}$ embryos exhibit hydrocephaly, exacerbated and aberrant migration of neural crest cells in the branchial arches and the frontonasal region, with no effect on cranial nerve formation but abnormal frontonasal development. We propose that Sox2 acts as a rheostat of the epithelial to mesenchymal transition during neural crest development.

Περίληψη

Το γονίδιο *Sox2* είναι ένας μεταγραφικός παράγοντας που περιέχει την περιοχή HMG-box, σαν δομική μονάδα πρόσδεσης στο DNA. Θεωρείται ένα από τα γονίδια-κλειδιά με πρωταγωνιστικό ρόλο στην ρύθμιση της πλειοδυναμίας της πρώιμης εμβρυικής ανάπτυξης και στο καθορισμό της κυτταρικής μοίρας στην διάρκεια της ανάπτυξης και της ομοιόστασης. Κατά την διάρκεια της πρώιμης ανάπτυξης, το γονίδιο *Sox2* είναι απαραίτητο για την δημιουργία της επιβλάστης. Όμως ο ρόλος του στα βλαστικά κύτταρα δεν είναι ακόμη γνωστός. Σε αυτήν την μελέτη δημιουργήσαμε μια χειρουργική απαλοιφή της λειτουργίας του *Sox2* στην επιβλάστη χρησιμοποιώντας ένα επαγόμενο με αναστροφή *Sox2* αλλήλιο. για την μετάδοση της επαγωγής, η μέθοδος του επαγόμενου με αναστροφή αλληλίου για την κατασκευή επαγόμενων αλληλίων βασίζεται σε μια βελτιστοποιημένη αναστρεφόμενη κασέτα παγίδευσης γονιδίων (Conditional by Inversion - COIN). Η κασέτα COIN περιλαμβάνει ένα ζευγάρι βελτιστοποιημένα σήματα πολυαδενυλίωσης για 3΄ μάτισμα, αλλά εισάγεται στην μη μεταγραφόμενη αλυσίδα του DNA του γονιδίου-στόχου, με αποτέλεσμα να μην μεταβάλλεται η μεταγραφή του γονιδίου, παρά μόνο όταν αναστρέφεται η κασέτα από το ένζυμο ανασυνδυασμού CRE. Για την εφαρμογή της τεχνολογίας του COIN σε όλα τα γονίδια που εκφράζουν πρωτεΐνες, το στοιχείο COIN έχει κατασκευαστεί μέσα σε ένα τεχνητό ιντρόνιο, επιτρέποντας την εισαγωγή της κασέτας μέσα σε ένα εξώνιο. Συνεπώς, θεωρητικά η μέθοδος COIN θα πρέπει να είναι εφαρμόσιμη και σε γονίδια που αποτελούνται από ένα εξώνιο. Για να ελέγξουμε αυτήν την ιδέα κατασκευάσαμε ένα αλλήλιο του γονιδίου *Sox2*. Το συγκεκριμένο γονίδιο που αποτελείται από ένα εξώνιο προσφέρει σχεδιαστικές προκλήσεις αφού ο εγγυής υποκινητής και η περιοχή κωδικοποίησης της πρωτεΐνης αποτελούν ένα CpG νησιδίο και γεφυρώνονται από ένα επικαλυπτόμενο μετάγραφο, *Sox2ot,* που περιλαμβάνει και το *mmu*-miR1897. Σε αυτήν την εργασία, αποδεικνύουμε ότι παρόλη την διατάραξη του CpG νησιδίου από το ιντρόνιο της κασέτας COIN, το $Sox2^{COIN}$ αλλήλιο είναι φαινοτυπικά ίδιο με το αντίστοιχο άγριου τύπου αλλήλιο, και επίσης δεν παρεμβαίνει στην έκφραση του *Sox2ot* και *mmu*-miR1897. Επιπλέον, το ανεστραμμένο COIN αλλήλιο του γονιδίου, $Sox2^{INV}$ είναι λειτουργικά ανενεργό, καθότι ομοζυγωτά ζώα εμφανίζουν τον ίδιο φαινότυπο όπως το $Sox2^{ßgeo/ßgeo}$, ένα πολύ καλά χαρακτηρισμένο *Sox2* ανενεργό αλλήλιο. Τέλος το πλεονέκτημα της πλήρους ενσωμάτωσης του γονιδίου-μάρτυρα eGFP στο αλλήλιο COIN αποδεικνύεται από το γεγονός του ότι η έκφραση του αντικατοπτρίζει την έκφραση του *Sox2* γονιδίου. Η αναστροφή του $Sox2^{COIN}$ αλλήλιου στην επιβλάστη παράγει φυσιολογικά ετεροζυγωτά $Sox2^{INV/+}$ ενήλικα ζώα και $Sox2^{INV/+(H)}$ απλοανεπαρκή μεταλλαγμένα έμβρυα. $Sox2^{INV/+ (H)}$ απλοανεπαρκή και επαγόμενα $Sox2^{INV/mosaic}$ έμβρυα παρουσιάζουν προβλήματα στην δημιουργία των κοιλιών κατά την ανάπτυξη της καρδιάς, υποφέρουν από βραδυκαρδία, από δυσμορφία του μυοκαρδίου και αιμορραγία και πεθαίνουν περίπου την εμβρυική ημέρα E11 (E11). $Sox2^{INV/+(H)}$ απλοανεπαρκή και επαγόμενα $Sox2^{INV/mosaic}$ έμβρυα παρουσιάζουν υδροκεφαλισμό, επιτείνουσα και παρεκκλίνουσα μετανάστευση των κυττάρων της νευρικής ακρολοφίας στα εμπροστορυνική περιοχή, χωρίς καμία επίδραση στην δημιουργία των κρανιακών νεύρων, παρά μόνο σε ανώμαλη εμπροστορυνική ανάπτυξη. Τα αποτελέσματα μας δείχνουν ότι το *Sox2* δρα ως ροοστάτης της μετατροπής κυττάρων με επιθηλιακό σε μεσεγχυματικό χαρακτήρα κατά την ανάπτυξη της νευρικής ακρολοφίας.

1. Introduction

1.1 *Sox* Gene Family

Throughout development transcriptional regulation of eukaryotic gene expression requires both the recruitment of tissue-specific transcription factors to a promoter and the proper establishment of the local chromatin structure (van Bakel, 2011). The *Sox* factors (Uchikawa et al., 2011) comprise a novel group of proteins characterized by the presence of an "Sex-determining region Y" (SRY) box (hence 'Sox'), a 79 amino acid motif that encodes an HMG-type DNA-binding domain. The genes encoding these factors, highly conserved across evolution, were originally identified through homology as they contain an HMG box closely related to that of the mammalian testis-determining gene SRY (Gubbay et al., 1990; Sinclair et al., 1990). The *Sox* family (Alatzoglou et al., 2009; Castillo and Sanchez-Cespedes, 2012; Harley and Lefebvre, 2010; Harris et al., 2010; Lovell-Badge, 2010; Malki et al., 2010; Uchikawa et al., 2011; Uy et al., 2012; Wegner, 2011) falls into a subclass of "High Mobility Group" (HMG) box proteins, the members of which show highly restricted tissue distribution (Alatzoglou et al., 2009; Bigalke and Burrill, 2007; Castillo and Sanchez-Cespedes, 2012; Chang et al., 2002; DeFina, 2008; Dong et al., 2004; Gould, 2003; Harley and Lefebvre, 2010; Harris et al., 2010; Kamachi et al., 2000; Kiefer, 2007; Koopman, 2005; Koopman et al., 2004; Lovell-Badge, 2010; Malki et al., 2010; Nagai, 2001; O'Brien and Degnan, 2000; Uchikawa et al., 2011; Uy et al., 2012; Wang et al., 2002; Wegner, 2011; Wilson and Koopman, 2002; Wilson and Dearden, 2008) and bind to specific sequences at high affinities (Ner, 1992). Most strikingly, on binding, Sox proteins cause DNA to bend at an acute angle (Ferrari et al., 1992; Giese et al., 1992). Compelling evidence for the developmental importance for individual members of the *Sox* gene family (Table I) (Pevny and Lovell-Badge, 1997) comes from mutational analyses in humans (Foster et al., 1994; Goodfellow and Lovell-Badge, 1993; Kwok et al., 1995; Schilham et al., 1996; Stevanovic et al., 1993; Wagner et al., 1994), mice (Schilham et al., 1996) Drosophila (Nambu and Nambu, 1996; Russell et al., 1996) and zebrafish (Germana et al., 2009).

Table I : The *Sox* family

Gene	Orthologue	Mapping	Expression
Group A *SRY*	Human, rodent, marsupial	Human Y	Genital ridge and testis
Group B *Sox1* *Sox2*	Human, mouse, chicken Human, mouse, chicken	Human 8 Human 3	*Sox*l: embryonic CNS, lens *Sox2*: ICM, primitive ectoderm, CNS, PNS, embryonic gut, endoderm; c*Sox2*: embryonic CNS and PNS
Sox3 *Soxl4* ZfSoxl9 *Sox15* Dichaete (*Sox7*OD, fish-hook)	Human, mouse, chicken Mouse Zebrafish Mouse Drosophila	Human X - - - -	*Sox3* and c*Sox3*: embryonic CNS - Embryonic CNS - Entire trunk of syncitial blastoderm; seven irregular stripes in cellular blastoderm; ventral and cephalic neuroectoderm
Group C *Sox4* *Sox11*	Human, mouse Human, mouse, chicken	Human 6 Human 2	Neural plate, proliferating progenitors of CNS throughout ontogeny Embryonic heart and spinal cord, adult pre-B and pre-T cells c*Sox7* 7: embryonic CNS, post-mitotic neurons

Group C			Neural plate, proliferating progenitors of CNS throughout ontogeny
Sox4	Human, mouse	Human 6	Embryonic heart and spinal cord, adult pre-B and pre-T cells
Sox11	Human, mouse, chicken	Human 2	c*Sox7* 7: embryonic CNS, post-mitotic neurons
Sox72	Human, mouse	-	-
Sox20	Human	Human 17	-
Group D			
Sox5	Human, mouse	Human 6	*Sox5*: adult testis
Sox6	Human, mouse	-	*Sox6*: embryonic CNS, adult testis
Group E			
Sox8	Human, mouse	-	-
Sox9	Human, mouse, chicken	Human 17, Mouse 2	Chondrocyte, genital ridge and adult testis, CNS, notochord
Sox10	Human, mouse	-	-
Group F			
Sox7	Mouse	-	Endoderm

ICM, inner cell mass; PNS, peripheral nervous system.

To date, HMG box proteins have been classified into two major subgroups according to sequence specificity of the DNA-binding and the number of HMG DNA-binding domains within a single protein (Laudet et al., 1993). The first sub- group consists of HMG box proteins, which contain more than one DNA binding domain, are usually expressed ubiquitously and bind preferentially to bend DNA. Members of this group include the HMG non-histone chromatin associated proteins HMG1 and HMG2 (Jantzen et al., 1990). In contrast, members of the second subgroup of proteins contain only a single HMG box,

showing highly restricted expression patterns, and may bind prestructured DNA with either little or no sequence specificity members of this class. However, they can also bind to linear DNA at high affinities in a sequence-specific manner. Examples of this group include the T-cell factor (TCF)(Waterman and Jones, 1990), MATA locus (Kelly et al., 1988) and the *Sox* families (Schepers et al., 2002). HMG-box proteins or genes can be further categorized into families on the basis of groups, which share a high degree of homology within the HMG box. An HMG box, which has at least 50% sequence identity with the founding member of this group, defines the Sox family of proteins, mouse SRY. At present, this group includes at least 20 members and can be further divided into subfamilies on the basis of degree of homology both within and outside the HMG box (Coriat et al., 1993; Denny et al., 1992; Weiss, 2001; Wright et al., 1993). Categorizing the *Sox* genes only on known sequence, many of which are incomplete, however, is at present ambiguous and may need to be refined on the basis of results from recent cloning and functional studies. Individual *Sox* family members also show a high degree of conservation across species both within and outside their HMG domain. It is difficult, however, to identify orthologues of a specific *Sox* gene in invertebrates but in some cases the similarities are intriguingly high. An example is human *Sox2* which shares 98% overall similarity with the mouse protein (Collignon et al., 1996; Stevanovic et al., 1994). 95% with the chicken homologue c*Sox2* (Uwanogho et al., 1995b) and 88% with the putative Drosophila homologue Dicheate. *Sry* is a major exception because very little sequence homology can be detected outside the HMG domain even between closely related species (Hacker et al., 1995; Tucker and Lundrigan, 1993; Whitfield et al., 1993).

1.2 Functional analysis of Sox genes

The first demonstration of a functional role for *Sox* genes in a genetic disease and a developmental process came from studies on human sex reversal (Alatzoglou et al., 2009; Bigalke and Burrill, 2007; Castillo and Sanchez-Cespedes, 2012; Chang et al., 2002; DeFina, 2008; Dong et al., 2004; Harley and Lefebvre, 2010; Harris et al., 2010; Kiefer,

2007; Koopman, 2005; Lovell-Badge, 2010; Malki et al., 2010; Uchikawa et al., 2011; Uy et al., 2012; Wegner, 2011; Wilson and Dearden, 2008). Genetic and molecular analysis of XY sex reversal (Alcock et al., 2009), arising from mutations in *SRY* and studies involving autosomal sex reversal in campomelic dysplasia (CD) patients, resulting from mutations in *Sox9* suggest that Sox genes may act to initiate or bias early cell fate decisions. This hypothesis is supported by the finding that misexpression of *Sry* in the mouse diverts cells destined to become ovaries to differentiate as testes (Cao et al., 1995; Capel, 1995). Expression of *Sry* alone, however, may be insufficient for cell fate conversion as, in a number of naturally occurring and experimental conditions, the sex reversal shows varying degrees of penetrance (Capel, 1995). Studies suggest that *Sox* genes may act in a dose-dependent manner. An idea that is supported by the fact that the dominant phenotype campomelic dysplasia patients must be caused largely by haploinsufficiency rather than a dominant negative effect. Detailed molecular analyses carried out in these studies provide in vivo evidence that Sox proteins can function as both classical transcription factors as well as architectural proteins.

1.3 *Sox* genes function in CNS and PNS

The study of transcription factors and their role in controlling gene expression during nervous system development has been extensively studied. Most of these studies (Aybar et al., 2002; Mason, 1996; Nakayama and Inoue, 2006; Rogers et al., 2009; Yamada et al., 1993) mention that although epigenetic factors affect the developmental fate of cells, transcription factors act for both the decoding of incoming messages and in controlling changes in the expression of other genes (Gu and Wang, 2012). For example, *retinoic acid* and *Noggin* and *Sonic hedge*hog genes affect the expression of a cluster of transcription factors such as *Hox*, *Sox*, *Pax* and *Pou* genes, while these factors alter the expression of banks of downstream genes, thereby controlling the developmental fate of stem cells (Scotting and Rex, 1996). Although little is fully understood about the intrinsic factors that regulate stem cell maintenance, and how they decide whether stem cells will generate

neurons or glia during differentiation (Wegner and Stolt, 2005). Thus, *Sox* gene function provides important clues about the control of these events. It has been widely shown that *Sox1*, *Sox2* and *Sox3* are required for progenitor maintenance and they are controlled by *Sox21* expression (Cunningham et al., 2008; Kuzmichev et al., 2012; Mallanna et al., 2010; Sandberg et al., 2005). *Sox9*, on the other hand, changes the potential of stem cells from neurogenic to gliogenic (Lee and Saint-Jeannet, 2011; Scott et al., 2010), while, *Sox10* is essential for terminal oligodendrocyte differentiation (Pozniak et al., 2010).

Interestingly so, each Sox protein has different functions in the peripheral nervous system (PNS) indicating also the important developmental differences between the CNS and PNS. More precisely, in the premigratory neural crest of the trunk, *Sox9* is switched on before *Sox10* (Finzsch et al., 2008). Moreover, melanocyte progenitor neural crest cells that migrate along the ventral and dorsolateral pathway turn off *Sox9* expression, while they still express *Sox10* (Cook et al., 2005). Many premigratory neural crest cells undergo apoptosis in mutant *Sox9* mice embryos (Scott et al., 2010). Once ventral neural crest cells have been located in their final position, they differentiate into neurons. This differentiation is triggered by the loss of *Sox10* expression, while when *Sox10* continues to be expressed in these cells they differentiate into glia or oligodendrocytes, as reffered. In the absence of *Sox10*, no glia or oligodendrocytes are generated (Mollaaghababa and Pavan, 2003). Finally, *Sox2* is turned on in Schwann cell precursors along nerves (Le et al., 2005; Parrinello et al., 2010).

1.4 *Sox*B genes

*Sox*B genes (*Sox*1, 2, 3, 14, 21) (Table II) show diverse and dynamic patterns of expression throughout embryogenesis and in a variety of adult tissue types. *Sry* is expressed in the undifferentiated male gonad and is quickly downregulated once the decision is made to initiate male development (Wright et al., 1995; Wunderle et al., 1996). Similarly, the expression patterns of many of the other Sox family members throughout development also appear to correlate with early cell fate decisions. For example, during the early phases of neural induction in both mouse and chick, the neuroepithelium shows a

dramatic up-regulation of *Sox1* and *Sox3* expression and a high level of *Sox2* expression (Streit et al., 1997) but then, coincidentally with the differentiation of neural precursors, expression of all three genes is rapidly downregulated. The three genes therefore, seem to respond both to signals involved in neural induction as well as signals which trigger the loss of potential and, thus, their expression largely defines the uncommitted neuroblast population in mice. The expression of a particular *Sox* gene is not necessarily restricted to a particular cell type or lineage. The expression patterns of *Sox*B genes implicate that these genes play a role in establishing cell fates in various tissues, implying that there must be alternative cooperating factors and different sets of target genes for each *Sox*B gene (Table II) (Pevny and Placzek, 2005).

Table II *: Sox* family members in CNS and PNS progenitor cells.

Sox group	Gene	Species	Location of expression of Sox factors involved in maintaining progenitor fate
*Sox*B	*Sox*B	Amphioxus	Neural primordium
*Sox*B1	*Sox1*	Mouse	Neural plate, proliferating progenitors of CNS throughout ontogeny
		Drosophila	Proliferating progenitors of embryonic CNS
			Neuroblasts of CNS
	Sox2	Mouse	Neural plate, proliferating progenitors of CNS throughout ontogeny
		Chick	
		Xenopus Dr osophila (Di chaete)	Neural plate, proliferating progenitors of CNS and PNS throughout ontogeny
			Neural plate, proliferating progenitors of CNS and PNS throughout ontogeny
			Neuroblasts of CNS and midline glia
	Sox3	Mouse	Neural plate, proliferating progenitors of CNS throughout ontogeny
		Chick	
		Xenopus	Neural plate and embryonic proliferating progenitors
			Neural plate and embryonic proliferating progenitors
*Sox*B2	*Sox14*	Mouse	Subclass of interneurons in the developing spinal cord
		Chick	Subclass of interneurons in the developing spinal cord
	*Sox2*1	Mouse	Ventricular zone of the developing nervous system
		Chick	Ventricular zone of the developing nervous system
*Sox*E	*Sox9*	Mouse	Neural crest progenitors, crest derivatives of branchial arches
		Chick	Neural crest progenitors, crest derivatives of branchial arches
		Xenopus	Neural crest progenitors
	Sox10	Mouse	Neural crest progenitors, melanoblasts and glia
		Chick	Neural crest progenitors, melanoblasts and glia
		Xenopus	Neural crest progenitors, melanoblasts and glia

1.5 *Sox2*: a stem cell gene

Sox2 encodes a transcription factor protein that harbours a DNA binding HMG domain. It is considered one of the key gene players for the regulation of pluripotency and early embryonic development, and a determinant of cell-fate during development and homeostasis. *Sox2* is a member of the *Sox* (SRY-related HMG box) gene family. It is a single exon gene that encodes *Sox2*, which belongs to the *Sox*B1 subgroup that also includes *Sox1* and *Sox3*, based on homology within and outside the HMG box (Kamachi et al., 2000). HMG domain of *Sox2* regulates its target genes by being paired off with tissue specific partners.

Amaral and co-workers have reported that the *Sox2* gene sequence is placed downstream of a long noncoding RNA (ncRNA) sequence, termed *Sox2* overlapping transcript (*Sox2*ot). It is transcribed at the same orientation as *Sox2* and its molecular and biological function remains elusive. *Sox2*ot transcript is expressed in mouse embryonic stem cells and in other tissues, including the nervous system where *Sox2* is also expressed. Furthermore, an isoform of *Sox2*ot (*Sox2*dot) located around to 500 bp upstream of Sox2 was detected exclusively highly in mouse brain (Amaral et al., 2009).

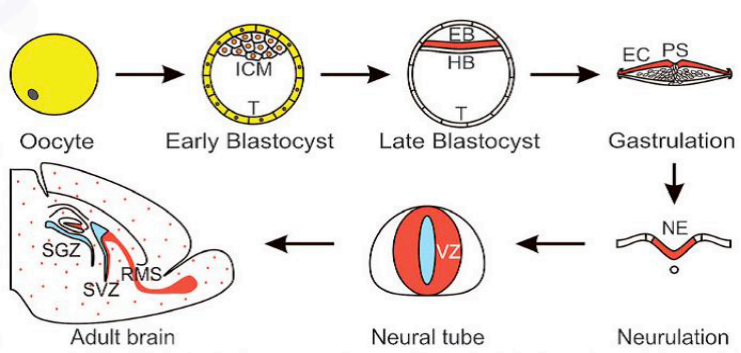
Embryonic *Sox2* expression (Guth and Wegner, 2008) appears at morula stage in mouse embryos, while maternal Sox2 protein accumulates in the cytoplasm of fertilized oocytes and persists until the blastocyst stage (Figure 1A). Before the formation of definitive neuroectoderm, *Sox2* is expressed in the primitive ectoderm (epiblast) of the pre-streak stage embryos in the mouse (Avilion et al., 2003b). During early gastrulation, expression of *Sox2* becomes restricted to the ectoderm destined to give rise to the neuroectoderm and anterior surface ectoderm, whilst *Sox2* expression is also maintained in the gut endoderm. Lovell-Badge and co-workers demonstrated the importance of *Sox2* in the formation of multipotential cells of the embryo and the ExEn (extraembryonic endoderm) (Avilion et al., 2003b). Furthermore, *Sox2*, along with the members of the B1 group *Sox* genes (*Sox1* and *Sox3*), is expressed within the developing neuroepithelium of vertebrate embryos where it is

implicated in the maintenance of neuroepithelial cell character (Figure 1B) (Pevny and Lovell-Badge, 1997) (Wood and Episkopou, 1999).

At E9.5 otocyst, *Sox2* is expressed in the region in which neuroblast will delaminate to form the statoacoustic ganglion. At E9.5, Sox2 is localized in all endoderm cells of the undivided foregut. It is also expressed in the adjacent hindbrain. At E10.5 *Sox2* is also expressed and overlaps with *Sox9* expressing cells found on two domains on the medial side of the otocyst. However, *Sox9* expression signal is relatively weaker in the ventral–medial *Sox2* expressing region than in the rest of the otocyst. At E12.5, *Sox2* expression is marked at the sensory primordia of the cristae, maculae and cochlear duct (Que et al., 2007). At this stage, *Sox2* is also expressed in developing lingual and palatine glands and in the foregut giving rise to the esophagus, trachea and stomach (Que et al., 2009).

Sox2 plays a direct role in establishing the prosensory domain within the cochlea where it is essential for the development of hair follicle cells. *Sox2* blocks inner hair cell formation in a dose dependent manner, suggesting that the relative ratio of *Sox2* to Atoh1 is important for the regulation of inner, and possibly outer, hair cell fate (Dabdoub et al., 2008). *Sox2* expression persists in adulthood, in brain areas where neural progenitor cells are located. Neural stem cells (NSCs) in adult brain exist mainly in the subventricular zone of the lateral ventricles and in the subgranular layer of the hippocampal dentate gyrus. *Sox2* is an essential factor for the maintenance of self-renewal capacity of adult NSC even when they have acquired cancer properties in glioblastomas (Gnjatic et al., 2009).

A



B

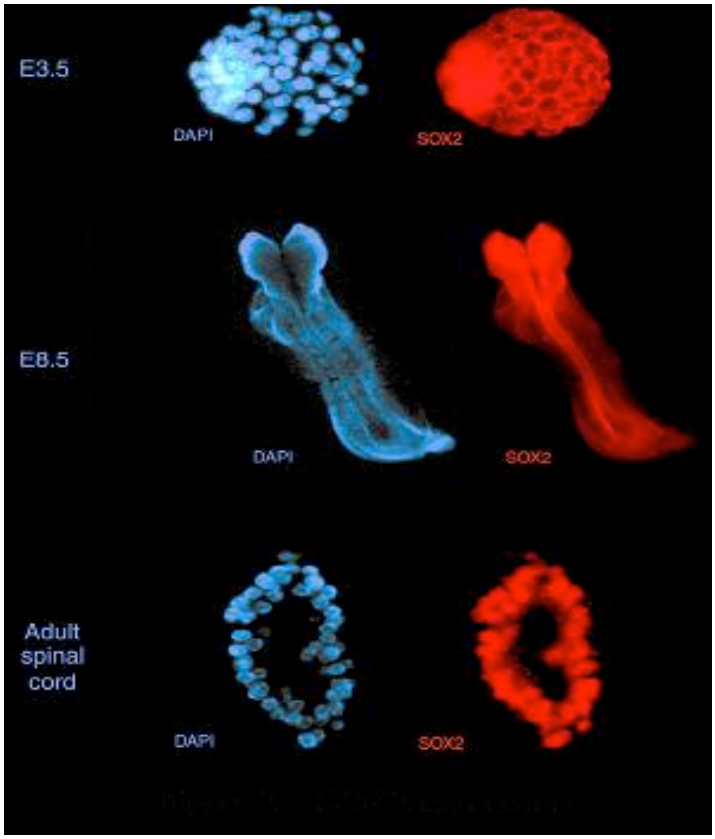


Figure 1. A) *Sox2* expression during mammalian development. Maternal Sox2 protein (yellow) from the oocyte is still dominant in the early blastocyst, although embryonic expression (red) in cells of the inner cell mass (ICM) has set in. Note that Sox2 is nuclear in cells of the inner cell mass (ICM) and cytoplasmic in trophoblast cells (T). In the late blastocyst, Sox2 occurs in epiblast (EB), but is absent from the hypoblast (HB). Upon gastrulation, Sox2 becomes restricted to the ectoderm (EC) and is no longer present in those cells of the epiblast that enter through the primitive streak (PS) to become endoderm or mesoderm. Further confinement to the neuroectoderm (NE) occurs upon neurulation. With the onset of differentiation in the neural tube, Sox2 is found in the ventricular zone (VZ), which harbors the proliferating stem cell population. Stem cells in the adult brain retain *Sox2* expression and are predominantly found in the subgranular zone of the hippocampus (SGZ), the subventricular zone (SVZ) of the lateral ventricles (blue) and the rostral migratory stream (RMS). B) Select populations of mature neurons also express Sox2 proteins, as indicated by dots (kindly provided by E. Remboutsika, blastocysts staining kindly provided by A. Avilion).

1.6 *Sox2*: gain of function

Sox2 is zygotically expressed during the 2-cell stage and its expression increases between the morula and blastocyst stages. Overexpression of *Sox2* has little effect on the ability of 2-cell embryos to cleave to the 4-cell stage. 15% of zygotically expressed genes are dramatically repressed in 2-cell embryos overexpressing *Sox2*. Furthermore, overexpression of *Sox2* leads to developmental failure after the 2-cell stage but before the 8-cell stage. The amount of Sox2 is critical for successful execution of this transition (Pan and Schultz, 2011). Remarkably, overexpression of *Sox2* decreases expression of its own gene, as well as four other *Sox2*:Oct-3/4 target genes (Oct-3/4, Nanog, Fgf-4, and Utf1) in which cells. Thus overexpression of *Sox2* in ES cells would trigger their differentiation (Kopp et al., 2008). As a result it has been speculated that the self-renewal of ES cells requires that Sox2 levels be maintained within narrow limits. Thus, *Sox2* appears to function as a molecular rheostat that controls the expression of a critical set of embryonic genes, as well as the self-renewal and differentiation of ES cells. Recent studies in this laboratory demonstrated that overexpression of *Sox2* in EC cells and ES cells leads to the suppression of the promoter activity of five *Sox2*:Oct-3/4 target genes. Moreover, transient increases in the expression of *Sox2* in F9 EC cells were shown to reduce the endogenous expression of *Sox2* and Oct-3/4 mRNA, plus the mRNA for Nanog and two other *Sox2*:Oct-3/4 target genes, *Fgf-4* and *Utf1* (Kim et al., 2008) (Figure 2). *Sox2* is a key gene that controls transcriptional networks required for pluripotency (Pan and Schultz, 2011). Adult pluripotent cells do not undergo all pluripotent characteristics typical for embryonic stem cells. Adult cells could be completely reprogrammed to embryonic stem cell-like cells by overexpression of some key transcription factors for pluripotency (Oct4, *Sox2*, Klf4 and c-Myc) (Geraerts and Verfaillie, 2009). Overexpression of *Sox2* results in the production of neurons, suggesting that *Sox2* is sufficient for the induction of neuronal fate in non-sensory epithelial cells (Puligilla et al., 2010)

Sox2 overexpression both induces a proximal phenotype in the distal airways/alveoli (Sasaki et al., 2012) and leads to cancer (Lu et al., 2010). *Sox2* is amplified in a subset of

squamous cell lung and esophageal cancers (Wilbertz et al., 2011). Increases in the expression of *Sox2* during brain tumour progression are likely to be linked closely with changes in other critical genes that work in concert with *Sox2* to enhance the tumorigenicity of brain tumours (Medulloblastoma and Glioblastoma) (Cox et al., 2012). Finally, *Sox2* overexpression does not affect the differentiation capacity of cultured ES cells, but it is essential for the survival and maintenance of the pluripotent phenotype (Nichols et al., 1998) (Niwa, 2000).

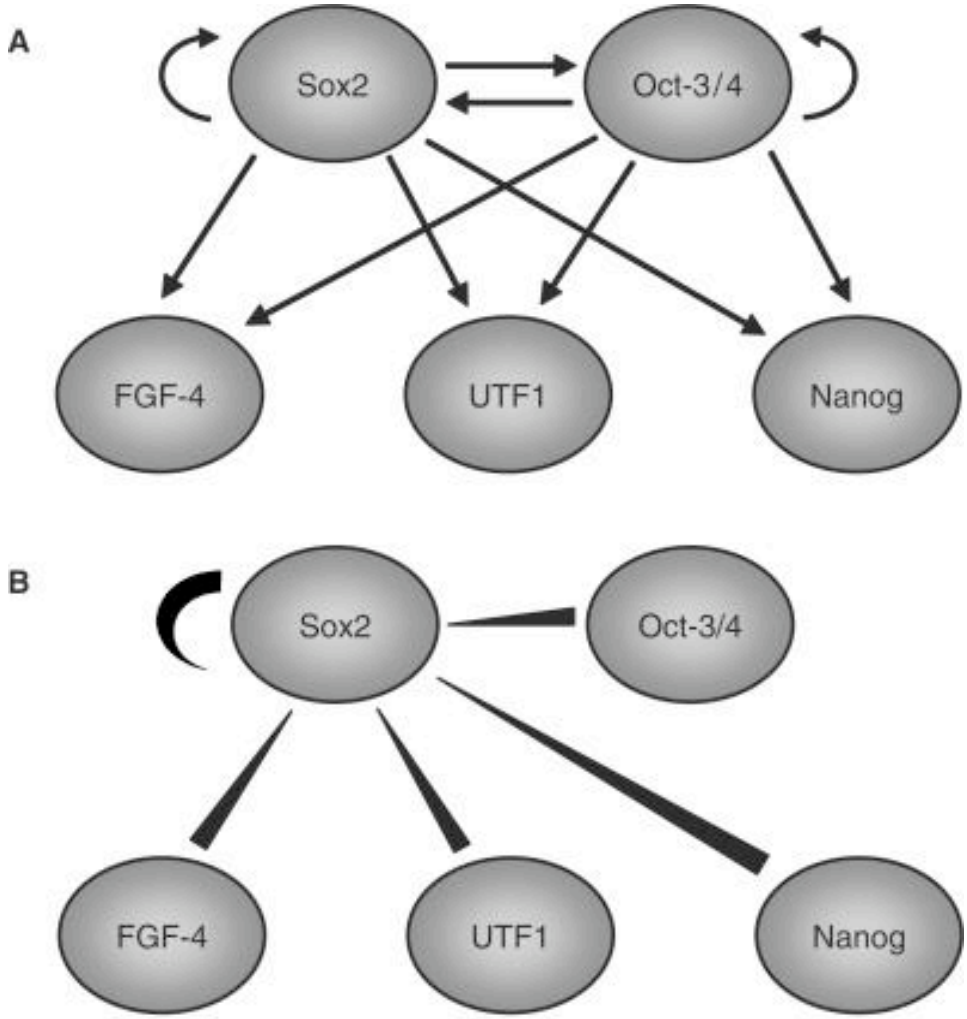


Figure 2. Model of *Sox2*, Oct-3/4 and Nanog gene regulation. *Sox2* is a central node in a network of genes of embryogenesis as well as stem cell self-renewal and pluripotency. A) In normal conditions *Sox2* and Oct-3/4 activate transcription of all known *Sox2*:Oct-3/4 target genes. B) Overexpression of *Sox2* leads to inhibition of target gene expression.

1.7 *Sox2*: Loss-of-function mutation

Mice carrying a loss of function mutation in the *Sox2* locus have been generated by the insertion of a βgeo cassette into the *Sox2* locus by Lovell-Badge and co-workers (Zappone et al., 2000). *Sox2* null embryos fail to form the epiblast and die at E5.5 (Figure 3) (Avilion et al., 2003b). Episkopou and co-workers have engineered a similar transgene allele, called $Sox2^{\beta geo2}$. $Sox2^{\beta geo2/\beta geo2}$ null mice show the same phenotype (Ekonomou et al., 2005). Heterozygote mice derived from these two targetings are viable with mild fertility defects, with no obvious phenotypic abnormalities reported, compared to wild type littermates.

Sylvia Nicolis and co-workers have engineered another *Sox2* mutated allele called $Sox2^{\beta geo\Delta E}$ (Ferri et al., 2004). This *Sox2* allele has been generated by the combination of a null mutation of *Sox2* ($Sox2^{\beta geo/+}$) with a regulatory knock out mutation ($Sox2^{deltaENHdeltaneo}$), which is obtained by deleting a neural-specific enhancer from the *Sox2* locus. In this study, it was indicated that $Sox2^{deltaENHdeltaneo}$ allele decreases the expression levels of *Sox2*, although it does not abolish it completely. Neither the heterozygous mutant ($Sox2^{\beta geo/+}$), nor the homozygous knock out mutants ($Sox2^{DENH/DENH}$) show any pathology, with the exception of some mild ventricle enlargement occasionally. This particular pathological phenotype is mostly observed in $Sox2^{\beta geo/DENH}$ compound heterozygotes, which express *Sox2* at a percentage of 25-30%, relative to the wild type expression, i.e. below the level of $Sox2^{\beta geo/+}$ heterozygote. Significantly, this suggests that *Sox2* acts in a gradient dependant manner (Ferri et al., 2004). Hypomorphic, i.e. partial loss of *Sox2* function, *Sox2* allele, in $Sox2^{\beta geo/DENH}$ mutant mice, has been generated by Sylvia Nichols and co-workers (Cavallaro et al., 2008) as referred. This hypomorphism of *Sox2* causes neurodegeneration in the cortical region and hippocampus due to the reduction of the neuronal progenitor proliferation rate and consequently abnormal neuronal function in these areas, resulting in circling mice. Hypomorphic *Sox2* mutant mice have also been studied in the retina by Larysa Penvy and co-workers, in which hypomorphic alleles (less than 40% activity) cause hypoplasia of optic nerves and chiasmata and variable micropthalmia (Taranova et al., 2006). Similar phenotype has been reported in *Sox2*/Oct-1 compound mutant alleles, which result in a complete failure

of nasal placode induction and impaired induction of the lens placode resulting in anophthalmia. Additionally, this phenotype resembles the one reported in Pax6$^{Sey/+}$ transgenic mice (Donner et al., 2007). Furthermore, in the hypomorphic *Sox2*$^{EGFP/LP}$ mutant embryos the placodes develop normally but taste buds fail to mature. They suggested that Sox2 functions in a dose-dependent manner to regulate the differentiation of endodermal progenitor cells of the tongue into taste bud sensory cells (Okubo et al., 2006). They showed recently that *Sox2* ablation, using *Sox2*$^{loxP/\ loxP}$ null mice, at E14.5 causes reduction of *wnt3a* expression in the cortical hem, which includes the hippocampal primordium, particularly in posterior regions. In addition, at the same embryonic day, *Shh* mRNA is reduced in *Sox2*-mutants in telencephalon and diencephalon but not in the midbrain and spinal cord (Favaro et al., 2009).

Four conditional hypomorphic alleles (*Sox2*^{tm1Lpev}, *Sox2*^{tm2Lpev}, *Sox2*^{tm3Lpev}, *Sox2*^{tm4Lpev}) were also generated by Larysa Pevny and co-workers (Ellis et al., 2004) (Ellis et al., 2004; Taranova et al., 2006). Using Cre mediated recombination for the removal of the floxed sequence, which includes the promoter and the coding sequence (cassette flanked by a loxP site in the NheI site immediately after the 3 first bases of the *Sox2* coding region), they generated hypomorphic alleles for *Sox2*. The *Sox2*-EGFP, *Sox2*^{COND}, *Sox2*^{LP} and *Sox2*^{IR} respectively, termed alleles were initially used for the conditional knock out of *Sox2* in the retinal neural progenitors. Conditionally ablated *Sox2* retinal progenitor cells lose the ability to proliferate and terminally differentiate. Hypomorphic *Sox2* allele, labeled *Sox2*^{tm2Lpev}/*Sox2*$^{tm2.1Lpev}$Tg(Pax6-cre,GFP)2Pgr, causes reduction of *Sox2* expression up to 40% of Wt levels and results in variable microphthalmia at E13.5 embryos due to aberrant neural progenitor differentiation and proliferation (Taranova et al., 2006). Hypomorphic *Sox2*^{tm1Lpev} allele was also used for the generation of transgenic mice expressing enhanced green fluorescent protein (eGFP) under the control of the endogenous regulatory regions of the *Sox2* locus to prospectively locate neural stem/progenitor cells in vivo and in vitro (Ellis et al., 2004). They also mentioned that inhibition of *Sox2* signaling in chick results in the

delamination of neural progenitor cells from the ventricular zone and their exit from cell cycle, which is associated with the loss of progenitor markers and the onset of early neuronal differentiation markers. The phenotype elicited by inhibition of *Sox2* signaling can be partially rescued by *Sox1*, providing evidence for the redundancy of SoxB1 proteins in CNS progenitors. In the same study it was indicated that the conditional inhibition of *Sox2* signaling in a chick's developing brain gives rise to embryos recovered at Mendelian ratio at all embryonic stages, but survived no longer than 12h postpartum (Graham et al., 2003).

Que and co-workers have used the Nkx2.5-Cre transgene and a floxed *Sox2* allele in order to delete *Sox2* conditionally in the ventral epithelial domain of the early anterior foregut, which gives rise to the future trachea and lung buds. They studied the $Sox2^{EGFP/COND}$ hypomorphic compound mutants which showed difficulties in breathing and die immediately at birth with air in the stomach. All the $Sox2^{EGFP/LP}$ hypomorphic mutants also die soon after birth, showing largely the same symptoms. They also observed that hypomorphic $Sox2^{EGFP/COND}$ mutants have isolated esophageal atresia (EA) or esophageal atresia /tracheoesophageal fistula (EA/TEF), with variable penetrance (Que et al., 2007). Among conditional mutants 60% of them have a short trachea. In the tracheal epithelium of all conditional mutants there are significantly more mucus-producing cells compared with wild type, and fewer basal stem cells, ciliated and Clara cells. Additionally, they observed that the differentiation of the epithelium lining the conducting airways in the lung is malformed, suggesting that *Sox2* also plays a role in the differentiation of embryonic airway progenitors into specific lineages. Conditional deletion of *Sox2* was also used to study its role in adult epithelium maintenance. Epithelial cells, including basal stem cells, lacking *Sox2* have a reduced capacity to proliferate in culture and to repair after injury in vivo. They combined two hypomorphic *Sox2* alleles ($Sox2^{EGFP/COND}$ null), generating compound mutant embryos in which the level of Sox2 varied by 40-50%. Most mutant embryos have a short trachea due to the shifting of the primary budding site of the lung anteriorly. Tracheal epithelium of mutant embryos has a larger number of mucus-producing cells compared with wild type, and fewer basal stem cells, ciliated and Clara cells. Additionally, differentiation of the epithelium lining

the conducting airways in the lung is abnormal, suggesting that *Sox2* also plays a role in the differentiation of embryonic airway progenitors into specific lineages. They concluded that the range of phenotypes they observed in posterior pharynx, trachea, esophagus, and fore stomach acts in a dose dependent manner of *Sox2* expression (Que et al., 2009).

Finally, Okuda and co-workers created a null allele, by breeding *Sox2*$^{WT/FL}$ mice with the nestin-Cre transgenic mice to obtain *Sox2*$^{FL/FL}$ mice; the level of *Sox2* mRNA in homozygote E14.5 mutant embryos was decreased to about 5% of Wt levels. Replacing the *Sox2* ORF with IREShygTKpA cassette flanked by loxP sites created *Sox2*FL mice. They showed that the loss of *Sox2* causes enlargement of the lateral ventricles at E14.5 and a decrease in the number of neurosphere-forming cells, but still *Sox2* mutant neural stem cells retained their stem multipotency and self-renewal capacity. They concluded that the expression levels of *Notch1*, *Hes1*, and *Hes5* in the forebrain and in neurospheres were essentially unaffected by loss of *Sox2* expression. Phenotype observed in *Sox2* targeting mutants suggest that it acts as a cell-intrinsic regulator that coordinates together with the microenvironment through various signaling pathways to regulate neural stem cell maintenance, self-renewal, and fate determination in both the embryonic and the adult nervous system (Ferri et al., 2004). Figure 4 summarizes all of the above *Sox2* conditional alleles. It is worth to note that in all cases *Sox2* sequences are missing, they lack a reporter and it is not mentioned of any effects in possible regulatory element of the locus, such as in the *Sox2ot* and *miR1897* expression.

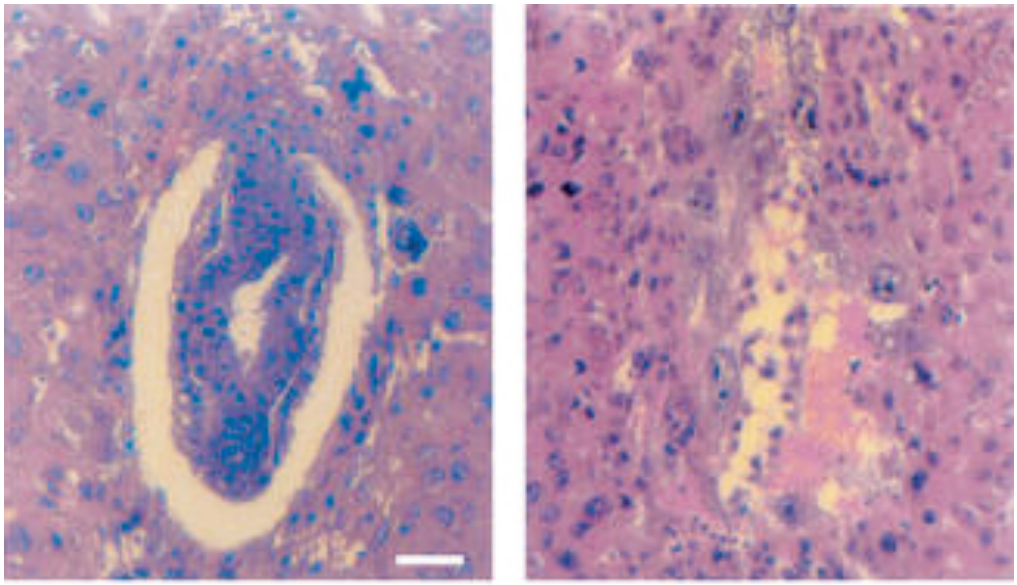


Figure 3. ***Sox2***$^{\beta geo}$ **mutant embryos lack epiblast.** Histological analysis of 6.0-dpc embryos from *Sox2*$^{\beta geo}$ intercrosses. Normal (left) and full knock-out (right) embryos. Sections were stained with haemotoxylin and eosin (H & E). Scale bar: 50μm (reprinted from Avilion et al., 2003).

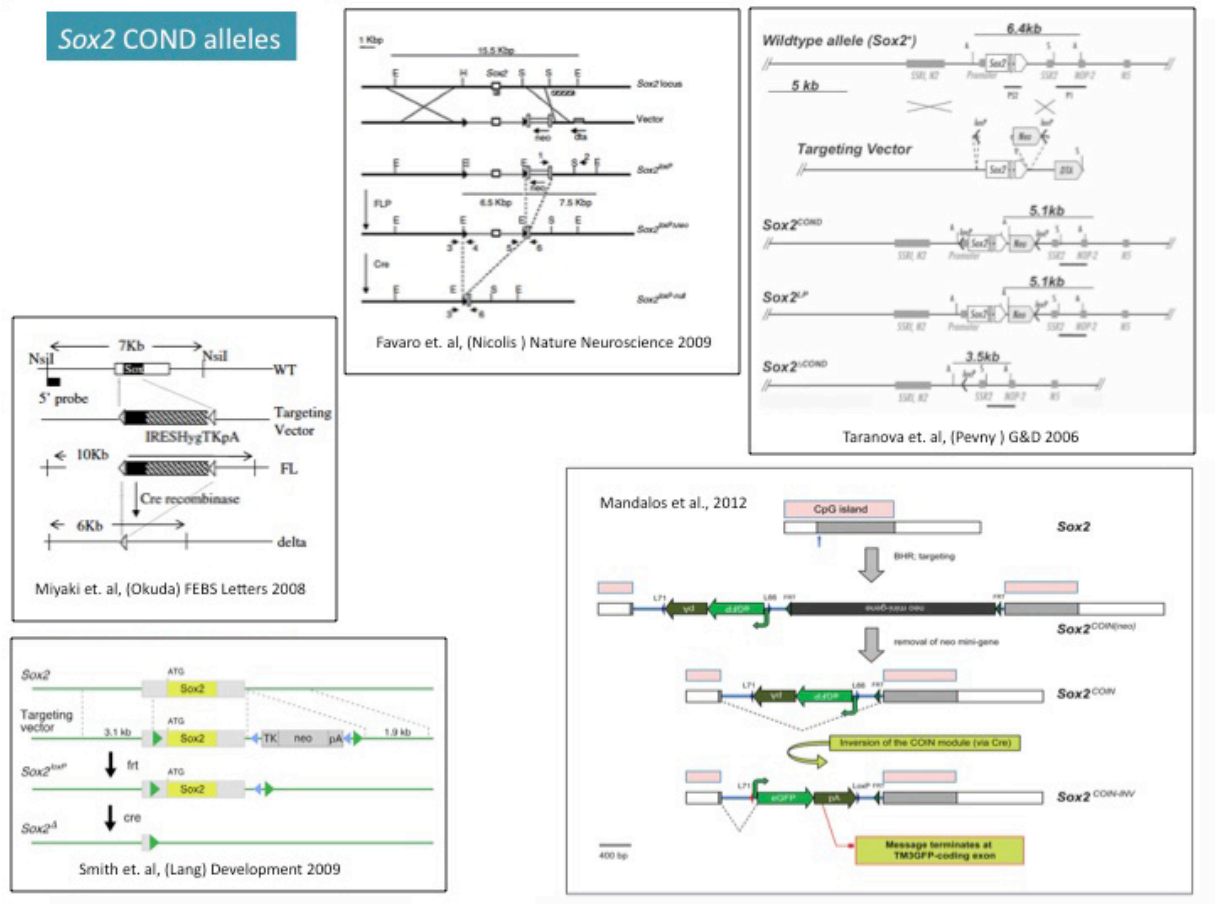


Figure 4. *Sox2* Conditional alleles

1.8 *Sox2* role in reprogramming to Induced pluripotent stem (iPS) cells

Somatic cells have been reprogrammed into pluripotent stem cells after the introduction of a combination of several transcription factors, such as *Oct3/4*,*Sox2*, *Klf4* and *c-Myc*. From a patient's somatic cells could be a useful source for drug discovery and cell transplantation therapies. However, most human iPS cells are made by viral vectors, such as retrovirus and lentivirus, which integrate the reprogramming factors into the host genomes and may increase the risk of tumour formation. Several non-integration methods have been reported to overcome the safety concern associated with the generation of iPS cells, such as transient expression of the reprogramming factors using adenovirus vectors or plasmids, and direct delivery of reprogramming proteins. Although these transient expression methods could avoid genomic alteration of iPS cells, they are inefficient. Several studies of gene expression, epigenetic modification and differentiation revealed the insufficient reprogramming of iPS cells, thus suggesting the need for improvement of the reprogramming procedure not only in quantity but also in quality.

Regenerative medicine using pluripotent/multipotent stem cells holds a great promise in developing therapies for treating developmental abnormalities, degenerative disorders, and aging-related illness. However, supply and safety of the stem cells are two major problems with today's regenerative medicine. Recent development of induced pluripotent stem cells (iPSCs) has overcome the supply shortages by allowing the reprogramming of patients' body cells to embryonic stem cell (ESC)-like pluripotent cells. Still, the potential tumorigenicity of iPSCs remains as an obstacle. During early embryogenesis ESCs can be generated without tumour formation; therefore, understanding the mechanisms underlying ESC generation may help us to prevent iPSC tumorigenicity. Previous studies have shown that an ESC-enriched noncoding RNA, miR-302, induces somatic cell reprogramming (SCR) to form iPSCs, suggesting its pivotal role in stem cell generation. Recent research further revealed that miR-302-induced SCR involves an epigenetic reprogramming mechanism similar to the natural zygotic reprogramming process in the two- to eight-cell-stage embryos. These findings indicate that miR-302, as a cytoplasmic gene silencer, inhibits the translation of multiple key epigenetic regulators, including AOF1/2, methyl-CpG binding proteins 1 and

2, and DNA (cytosine-5-)-methyltransferase 1, to induce global DNA demethylation, which subsequently triggers the activation of the previously defined factors Oct4, *Sox2*, and Nanog to complete the reprogramming process. The same mechanism was also found in the event of somatic cell nuclear transfer.

1.9 Application of a novel strategy of engineering conditional alleles to a single exon gene, *Sox2*

Intronless (single exon) genes are thought to be evolutionary innovations, whose formation via reverse transcription mediated mechanisms represents an important route of evolution for tissue-specific functions in animal cells (Brosius and Gould, 1992; Kamachi et al., 2009). Approximately 12% of the human and 13.4% of the mouse protein-coding genes are intronless (Gentles and Karlin, 1999; Sakharkar et al., 2006; Sakharkar et al., 2005; Venter et al., 2001), and include genes that encode for regulatory proteins and components of signal transduction pathways (Hill and Sorscher, 2006), histones (Friend et al., 2007; Huang and Carmichael, 1997; Maquat and Li, 2001; Medlin et al., 2005), G Protein-coupled Receptors (Gentles and Karlin, 1999) and transcription factors such as the *Sox* (SRY-related HMG box) family (Elkouris et al., 2011; Remboutsika et al., 2011a; Schilham et al., 1993).

Sox2 is a well-characterized and important example of a single exon gene. It pairs with tissue-specific partners (Kirby et al., 2002) to impart and maintain pluripotency (Behbahaninia et al., 2011) and multipotency (Remboutsika et al., 2011a; Yoon et al., 2011) during development and homeostasis (Parrinello et al., 2010). *Sox2* null embryos fail to form the epiblast and die at E5.5 (Avilion et al., 2003a). However, even reduction in *Sox2* levels to 25-30% relative to the wild type leads to pathological phenotypes in mice. These include neurodegeneration in the cortical region and hippocampus (Cavallaro et al., 2008), hypoplasia of optic nerves and chiasmata and variable microphthalmia (Taranova et al., 2006), failure of nasal placode induction (Donner et al., 2007), failure of taste buds to mature (Okubo et al., 2006), malformation of the epithelium lining the conducting airways in the lung

(Que et al., 2007), enlargement of the lateral ventricles at E14.5 (Ferri et al., 2004), and immature differentiation of cochlea hair follicles (Dabdoub et al., 2008) as referred above.

From a gene structure standpoint, *Sox2* presents a complex locus rich in genetic elements, including an overlapping transcript (Amaral et al., 2009), a putative microRNA (Griffiths-Jones et al., 2006), and a CpG island (Alonso et al.; Barrand et al.; Barrand and Collas) (Farthing et al., 2008). The combination of a well-conserved compact locus with overlapping transcripts and regulatory elements (Funabashi et al., 2010; Iwafuchi-Doi et al., 2011; Miyagi et al., 2006; Miyagi et al., 2004; Sikorska et al., 2008; Tomioka et al., 2002; Wiebe et al., 2000), together with the apparent need to maintain proper levels of Sox2 for organogenesis and homeostasis, underscore the difficulties associated with designing conditional alleles for *Sox2*. We hypothesized that developed method for generating conditional alleles – Conditional by Inversion (COIN) – might present a better choice over simple floxing of *Sox2*, and generated the corresponding conditional-null allele, $Sox2^{COIN}$. We show that this method is successful in that the $Sox2^{COIN}$ allele starts as wild type, and it is converted into a null by the action of Cre, at which point, the expression of *Sox2* is replaced by that of a marker, *eGFP*. This work indicates that the COIN method can be applied to single exon genes and provide a new design modality that can be adopted for other genes like *Sox2*.

Objective

Sox2 is a single exon gene essential for pluripotency, embryonic development and adult life. During early development, *Sox2* is required for the formation of the epiblast and the multipotent lineages in the early embryo. However, its role in stem cells is still elusive. Here, we performed a surgical ablation of *Sox2* in the epiblast using a novel Conditional by Inversion *Sox2* allele ($Sox2^{COIN}$) to study the role of *Sox2* in stem cells.

2. Materials and Methods

2.1 Gene Targeting

Targeted *Sox2*$^{COINneo/+}$ ES cells were generated using VelociGene™ methodology, essentially as described (Valenzuela et al., 2003). Briefly, the BAC-based targeting vector was assembled on bacterial artificial chromosome (BAC) RP23_406a6 that encompasses the single protein-coding exon of *Sox2* flanked by approximately 95 and 71 kb upstream and downstream, respectively. The COIN module intron was introduced by bacterial homologous recombination (Zhang et al., 1998) after the 30th nucleotide of *Sox2*'s coding region (i.e. coordinate 34549367 on Chromosome 3, as annotated on Ensembl release 67), splitting the single coding exon of *Sox2* into two exons of 441 bp and 2016 bp respectively (Figure 5B).

2.2 Experimental Animals

Sox2$^{COINneo/+}$ mice were bred with *Tg(ACTB:FLPe)* mice (Flp-deleter mice) to excise the Neo cassette and generate *Sox2*$^{COIN/+}$ *Tg(ACTB:FLPe)* mice. These were bred with C57BL6 mice to bring the *Sox2*^{COIN} allele into the germline. *Sox2*$^{COIN/+}$ mice were in turn bred with Tg(*Sox2*:CRE) mice to generate *Sox2*$^{COIN-INV/+}$ mice. All animals were handled in strict accordance with good animal practice as defined by the Animals Act 160/03.05.1991 applicable in Greece, revised according to the 86/609/EEC/24.11.1986 EU directive regarding the proper care and use of laboratory animals and in accordance to the Hellenic License for Animal Experimentation at the BSRC "Alexander Fleming" (Prot. No. 767/28.02.07) issued after protocol approval by the Animal Research Committee of the BSRC "Alexander Fleming" (Prot. No. 2762/03.08.05).

2.3 Embryo processing, tissue preparation and histological analysis

For staging of the embryos, midday of the vaginal plug was considered as embryonic day 0.5 (E0.5). E6.5 decidua and E14.5 embryos were collected and dissected in cold PBS. Tissues were fixed with 10% formalin for 24 hours at room temperature and then washed several times with 1% PBS, then placed in embedding cassettes. Paraffin sections (10 µm) were stained with Hematoxylin and Eosin (H&E) using standard procedures and mounted with xylene based mounting medium. E14.5 *Sox2*$^{\beta geo2/+}$ LacZ staining was performed following standard protocol (Avilion et al., 2003a).

Deciduas were removed from pregnant female mice and placed in a petri dish containing fresh DMEM medium (Gibco) supplemented with antibiotics (penicillin/streptomycin). Embryos were harvested and fixed with 4% PFA overnight at 4^{o}C and thoroughly washed with 0.12M phosphate-buffered saline (PBS) pH 7.5. Fixed embryos were incubated with 20% sucrose overnight at 4^{o}C for cryoprotection before they were embedded with O.C.T. compound (VWR International), snap-frozen in dry ice and finally stored at -80^{o}C. Sagittal sections were prepared using cryostat. Cryosections (10-12 µm thick) were collected on Superfrost Plus microscope slides (VWR International) and stored at -20^{o}C before analysis. For In Situ Hybridization, embryos were fixed overnight in PFA 4% in PBS then rinsed three times in PBS/Tween (0.1%) followed by three times wash in methanol before storage at -20°. Conventional bright field and fluorescence microscopy was performed under a Leica MZ16FA stereoscope.

2.4 Genotyping

Tail, yolk sack or embryonic tissues were isolated and processed according to previously described methodology. Deletion of the EM7p-neo-polyA FRT-flanked (FRTed) in

the germline dual-purpose antibiotic/drug selection cassette by FLP recombinase was documented by genotyping PCR from adult mice tails. Neo gene was amplified with *Neo* Frw (5'-CTGAATGAACTGCAGGACGA-3'), *Neo* Rvs (5'-ATACTTTCTCGGCAGGAGCA-3') (172bp); FLPe with *FLP* Frw (5'-GAGAAGAACGGCATAGTGCG-3'), *FLP* Rvs (5'-GACAAGCGTTAGTAGGCACAT-3') (600bp). Genomic DNA from E14.5 embryos was isolated from yolk sack. In E6.5 embryos genomic DNA was isolated, under the stereoscope, by carefully scraping off glass slides after staining with H&E and mounting. Detachment of cover slides was done by embedding mounted slides in xylene. Detection of *Sox2*$^{COIN/+}$ mice that carry the assay for the COIN cassette was performed using the following set of primers: 5' end primer combination (338 bp), INVdiaF2 (F2): 5' CACTTTCTACTCTGTTGAC 3', INVdiaR2 (R2): 5' CCTTACATGTTTTACTAG 3', 3' end primer combination (470 bp), INVdia(eGFP).F3 (F3): 5'CTGAAGCACTGCACGCCGTAG 3', INVdia.R4 (R4): 5' CTCAGAGTATTTTATCCTCATCTC 3' (Figure 6a). PCR amplification of COIN cassette in mice and embryos E6.5 and E14.5 was also performed with the following primers: Sox2401F Frw: 5' GCGCGCATGTATAACATGATGGAGAC 3', R2*R 5' TTCCTCCTCTCCTGACTACTCCCAGT 3' (465bp). PCR amplification of *Sox2* in WT and heterozygous mice was performed with the following primers: Sox2401F Frw: 5' GCGCGCATGTATAACATGATGGAGAC 3', Sox21050R Rvs: 5' TAGGTCTGCGAGCTGGTCATGGAGTT 3' (650 bp). *Sox2*$^{INV/+}$ mice and embryos that carry inverted COIN cassette, were genotyped with INVdiaR2 (R2), INVdia.R4 (R4) (600bp) and INVdiaF1: 5' GTTTTCAGGGTGTTGTTTAG 3', INVdia(eGFP).F3 (F3) (300bp) set of primers. Sequences for PCR amplification of *Tg(Sox2-cre)* and *Sox2*$^{\beta geo2}$ were found in MGI (ID: 103270 and 1915777 respectively). All PCR reactions were performed using 0.2 U/μL Taq Polymerase, standard PCR conditions, and 1 M betaine. All PCR reactions were performed in a BioRad C1000 thermal cycler. PCR conditions for Sox2 WT and Sox2COIN cassette are: 92°C for 60 sec 1 cycle, 92°C for 30 sec 35 cycles, 55°C for 30 sec, 72°C for 45 sec, 72°C for 240 sec 1 cycle, 10°C for 1 cycle and room temperature/shut off hold.

2.5 Imaging analysis

Conventional bright field and fluorescence microscopy was performed under a Leica MZ16FA stereoscope, while the dissection of the embryos took place either in 1X cold PBS or in DMEM medium supplemented with 2mM glutamine and 0.5mM penicillin and streptomycin.

2.6 Western Blotting

E14.5 embryos were lysed by sonication and the resulting pellets were washed with PBS and dissolved in cold buffer A (20mM Tris-HCl, 420Mm NaCl, 0.2Mm EDTA, 0.5mM DTT, 25% glycerol, 0.5mM PMSF, 1.5mM MgCl2, 0.5% NP40). Incubation at 4^{o}C for 15min and centrifugation for 15 min at 10,000xg followed. The supernatant was recovered and the protein concentration was determined by BCA Protein Assay Reagent (bicinchoninic acid) according to the instructions of the manufacturer (Thermo Scientific Pierce BCA Protein Assay Kit). Proteins (50μg per lane) were separated on 10% SDS-polyacrylamide gel, transferred to nitrocellulose membrane and membrane was blocked in western blot blocking buffer (5% milk, 10mM Tris-HCl pH7.6, 0.15mM NaCl, 0.05% Tween-20) for 2h at RT, incubated o/n with the primary antibody at 4^{o}C. Sox2 rabbit Polyclonal IgG, goat anti-GFP polyclonal antibody (Santa Cruz Biotechnology, Inc, Santa Cruz, CA, USA) and goat β-actin polyclonal antibody (Cell Signaling Technology, Inc, Danvers, MA, USA) were used (1:1000). After extensive washing in TBST.1 (10mM Tris-HCl, 0.15mM NaCl, 0.1% Tween-20), goat anti-rabbit HRP conjugated secondary antibody was applied (1:10,000) for 2h at RT. Proteins were visualized by chemiluminescence detection using ECL (Cell Signaling Technology, Inc., Danvers, MA, USA).

2.7 RNA analysis

RNA was extracted from E14.5 mouse embryos and subjected to Taqman. Real-time PCR analysis typically, *Gapdh* was used as a control house-keeping gene, although analysis was also performed using *Cyclophlin* and *β-actin* with similar results. For miR1897 analysis, miR16 and Sno135 were used as internal controls. All probes are hydrolysis probes with 5' Fam Fluorophore and 3' quencher (BHQ) (Biosearch Technologies). Probes codes and sequences for each gene are: for mSox2, Applied Biosystems, Inc, TaqMan assay ID: Mm00488369_s1, for m*Sox2ot*, Applied Biosystems, Inc, TaqMan assay ID: Mm01291217_m1, for Lac-Z, FRW: TTTCAGCCGCGCTGTACTGGA, RVS: TGTTGCCACTCGCTTTAATGATG, for eGFP: FRW: TCTTCA AGTCCG CCATGCCCG, RVS: CTACCCCGACCACATGAAGC, for miR1897: Applied Biosystems, Inc, TaqMan assay, Probe sequence: 121199.

2.8 RNA In Situ Hybridization

RNA Probes for in situ hybridization reactions were prepared by PCR. PCR reaction mix was composed of 1 µg buffer, 1µg DNA, 2µl 0.1M DTT stock 10mM, 1µl RNAsin (40units/ µl) Promega, 2µl 10x Dig mix (Boehringen) 1x=1mM for A, C, G + 0.55mM UTP or 0.35mM dig UTP, 4µl 5x Tc buffer* (Promega), Dep C H2O and 2µl RNA polymerase (Promega) T3, T7, Sp6.

Reaction (Wilkinson, 1995) took place at 37°C for 2 hours and then DNA was checked on an RNAse free gel/cold (cold cabinet or on ice), 150mA for 10 - 30min, in an agarose gel with TBE. After running 1µl of reaction, DNA was denatured with 7µl formamide at 80°C for 3min and then 2µl DNAse (RNAse free) were added (Promega) and finally incubated them at 37°C for 15min. Volume was brought up to 50µl and probe was purified through an RNA purification column (Clonetech) (chromaspin column) (100) and checked in gel again (1µl) and the integrity of the probe was tested in an agarose gel.

3. Results

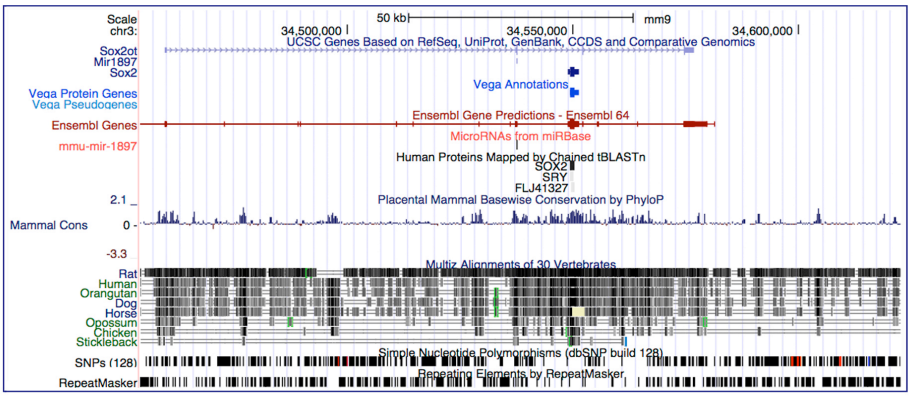
3.1 Generation of the *Sox2*^{COIN} allele

Sox2 (ENSMUSG00000074637) is a single exon gene encoding a 319 amino acid protein. The *Sox2* locus contains several features that render it complex from the standpoint of engineering modified alleles (Figure 5A). To begin with, *Sox2's* proximal promoter and coding region comprise a CpG island (Kirby et al., 2002). Furthermore, the *Sox2* exon is contained with the intron of a long non-coding RNA (ncRNA), termed *Sox2* overlapping transcript (*Sox2ot*) or "non-protein coding RNA 43", which also contains mmu-miR1897 (miR1897) (Griffiths-Jones, 2006). *Sox2ot* is transcribed from the same strand as *Sox2* but its molecular and biological functions remains elusive. *Sox2ot* transcript is expressed in mouse embryonic stem cells and in other tissues, including the nervous system, where *Sox2* is also highly expressed (Fantes et al., 2003), while an isoform of *Sox2ot*, *Sox2dot*, located around 500 base pairs upstream of *Sox2*, was detected exclusively in adult mouse brain (Amaral et al., 2009). Because of this complexity, *Sox2* is a challenging locus to apply conditional mutagenesis, and therefore presents a stringent test for new methods of allele design, such as COIN.

The COIN method relies on an optimized gene trap-like element, referred to as the COIN module (Economides et al., 2012). The COIN module is comprised of a *3' splice region-reporter cDNA-polyadenylation region* optimized to function as an efficient transcriptional block, and it is flanked by *Lox71* and *Lox66* sites are in a mirror image configuration to enable Cre-mediated inversion (Albert et al., 1995). In order to generate conditional-null alleles, the COIN module is placed in a position antisense to the target gene, either within a native intron, or an exon. The latter is made possible by embedding the COIN module within an artificial intron – the COIN module intron – and using that intron to split the target exon into two operational halves (Economides et al., 2012).

To generate *Sox2*^{COIN}, the COIN module intron was inserted directly into the single exon of *Sox2* (Figure 5B), after the 30th nucleotide of *Sox2*'s open reading frame, splitting the single *Sox2* exon into two exons. The COIN module lies inertly within the antisense strand of *Sox2*, stealth to transcription. Upon Cre-mediated inversion into the sense orientation, the *Sox2*^{COIN} allele is converted into a null allele, *Sox2*^{INV}. This is accomplished by the COIN module abrogating transcription of full length Sox2, effectively replacing it with expression of the COIN module's eGFP reporter. The expression of *eGFP* in place of *Sox2* is controlled by *Sox2*'s promoter and regulatory elements, and enables visual identification of the inversion event at the tissue and cellular level. The functionality of the allele was assessed *in vivo* in a series of experiments that assessed whether *Sox2*^{COIN} is a truly wild type allele, and whether *Sox2*$^{INV/INV}$ recapitulate the null phenotype, while providing a useful marker that faithfully reproduces the expression profile of *Sox2*.

A. *Sox2* genomic locus



B. Targeted allele

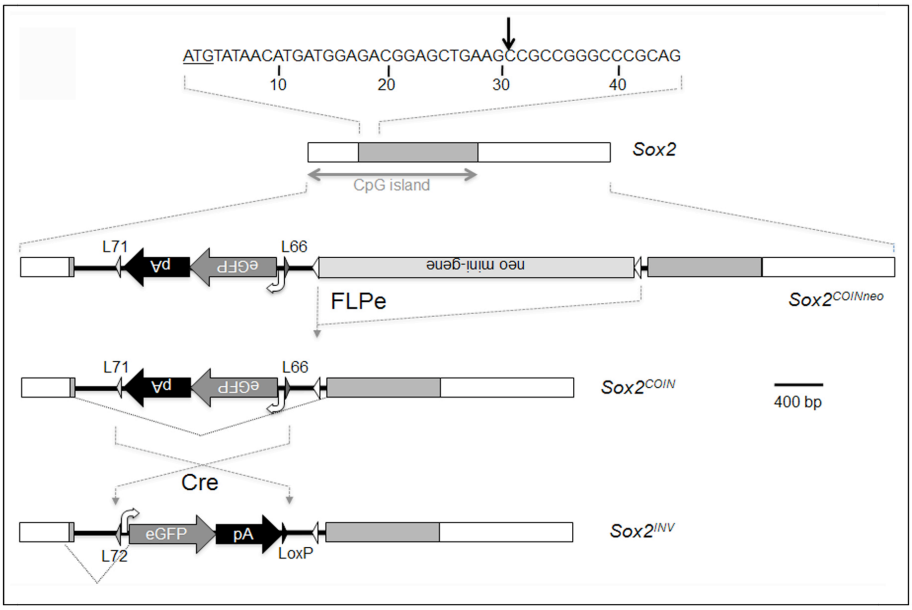


Figure 5. Targeting strategy generating a COIN allele of *Sox2*. A) Schematic representation of the mouse *Sox2* locus indicating the relative location of the exon on chromosome 3, as well as that of *mir1897*, the non-coding RNA *Sox2ot*, and CpG islands in the genomic region. The degree of conservation of the locus sequence between mammalian species (ECRs) is indicated. Adapted from http://genome.ucsc.edu. B) Schematic representation of the *Sox2*^{COIN} allele. The COIN module intron is inserted after the 30th nucleotide of *Sox2*'s coding region (i.e. coordinate 34549367 on Chromosome 3) splitting the single exon of Sox2 into two exons and also dividing the CpG island. The COIN module is comprised of an optimized gene trap-like element composed of the 3' splice region of the rabbit beta globin gene (HBB_RABIT), followed by *eGFP* (lacking an initiating ATG) and the polyadenylation region from HBB_RABIT, all placed in the antisense strand. The COIN module has been flanked with *Lox71* and *Lox66* sites in a mirror image orientation, thereby enabling inversion by Cre. For BHR and targeting, a FRT-flanked *neo* cassette has been incorporated into the COIN intron. After targeting, *neo* is removed to give rise to the *Sox2*^{COIN} allele. The COIN module is antisense to *Sox2*, and hence it predicted not to interfere with expression of *Sox2*. However, after inversion of the COIN module to the sense strand, transcription terminates around the polyadenylation region of the COIN module, and as a result expression of *Sox2* is replaced by *eGFP*.

3.2 *Sox2*^{COIN} is wild type in homozygosis

Offspring of *Sox2*$^{COIN/+}$ intercrosses were born in Mendelian ratios and no lethality was observed in embryos, newborn pups and adults (Table III). Homozygote mice fed normally, showed no abnormal behaviour and they had normal weight in adulthood (data not shown). Macroscopic analysis of E14.5 *Sox2*$^{+/+}$, *Sox2*$^{COIN/+}$, and *Sox2*$^{COIN/COIN}$ mice showed that the COIN module does not affect normal embryonic mouse development (Figure 6C). These phenotypic observations are further corroborated by the result that *Sox2*$^{COIN/\beta geo2}$ E6.5 embryos were morphologically indistinguishable from *Sox2*$^{+/+}$ or *Sox2*$^{\beta geo2/+}$ embryos derived from a *Sox2*$^{\beta geo2/+}$ with *Sox2*$^{COIN/COIN}$ cross (Figure 6B, F), where *Sox2*$^{\beta geo2}$ is a null allele of *Sox2* (Avilion et al., 2003a) (see below). Furthermore, examination of *Sox2* mRNA (Figure 6F) and Sox2 protein (Figure 6H) expression levels show no apparent difference between the three genotypic classes, *Sox2*$^{+/+}$, *Sox2*$^{COIN/+}$, and *Sox2*$^{COIN/COIN}$, demonstrating that the COIN module has no effect on the expression of *Sox2*. Thus, by all of these criteria – heritability, phenotype, expression of mRNA and protein –*Sox2*^{COIN} behaves as a wild type allele.

3.3 The expression of *Sox2ot and miR1897* are unaffected in *Sox2*$^{COIN/COIN}$ mice

To assess whether the COIN module affects *Sox2ot* RNA expression, we isolated RNA from E14.5 mouse embryos from different intercrosses and quantified *Sox2ot* RNA levels by Taqman Real-Time PCR analysis (Figure 6G). No significant difference was detected in the expression of *Sox2ot* in *Sox2*$^{+/+}$, *Sox2*$^{COIN/+}$, and *Sox2*$^{COIN/COIN}$, demonstrating that the COIN module has no effect on the expression of *Sox2ot*, at least prior to inversion. Identical observations where made for *miR1897*, which is embedded in *Sox2ot* (Figure 6H).

3.4 *Sox2*^{COIN} is efficiently inverted by Cre to generate *Sox2*^{INV}

To assess whether we could trigger COIN inversion upon Cre expression, *Sox2*$^{COIN/+}$ adult mice were intercrossed with *Sox2*Tg (Sox2:CRE) transgenic mice to generate *Sox2*$^{INV/+}$ embryos and adult mice (Figure 6D, E). In contrast to the partial infertility phenotype that has been observed with *Sox2*$^{ßgeo2/+}$ mice (Avilion et al., 2003a), *Sox2*$^{INV/+}$ adult mice exhibited no obvious phenotypes and transmitted the inverted allele in Mendelian ratios (Table IV), irrespective of whether the *Sox2*^{INV} allele is transmitted via the male or female germline (data not shown). More importantly, E14.5 *Sox2*$^{INV/+}$ embryos (Figure 6E) displayed vivid eGFP expression in the cerebral cortex, retina, olfactory bulb, hair follicles, olfactory epithelium and spinal cord (Figure 6E), mirroring what has been observed with X-gal stained E14.5 *Sox2*$^{ßgeo2/+}$embryos (Ekonomou et al., 2005). In addition, the presence of eGFP protein can be detected by Western blotting in protein extracts derived from *Sox2*$^{INV/+}$ embryos, and appears to be accompanied by a reduction in the levels of Sox2 protein, similar to what has been observed in the *Sox2*$^{ßgeo2/+}$ embryos (Figure 6I).

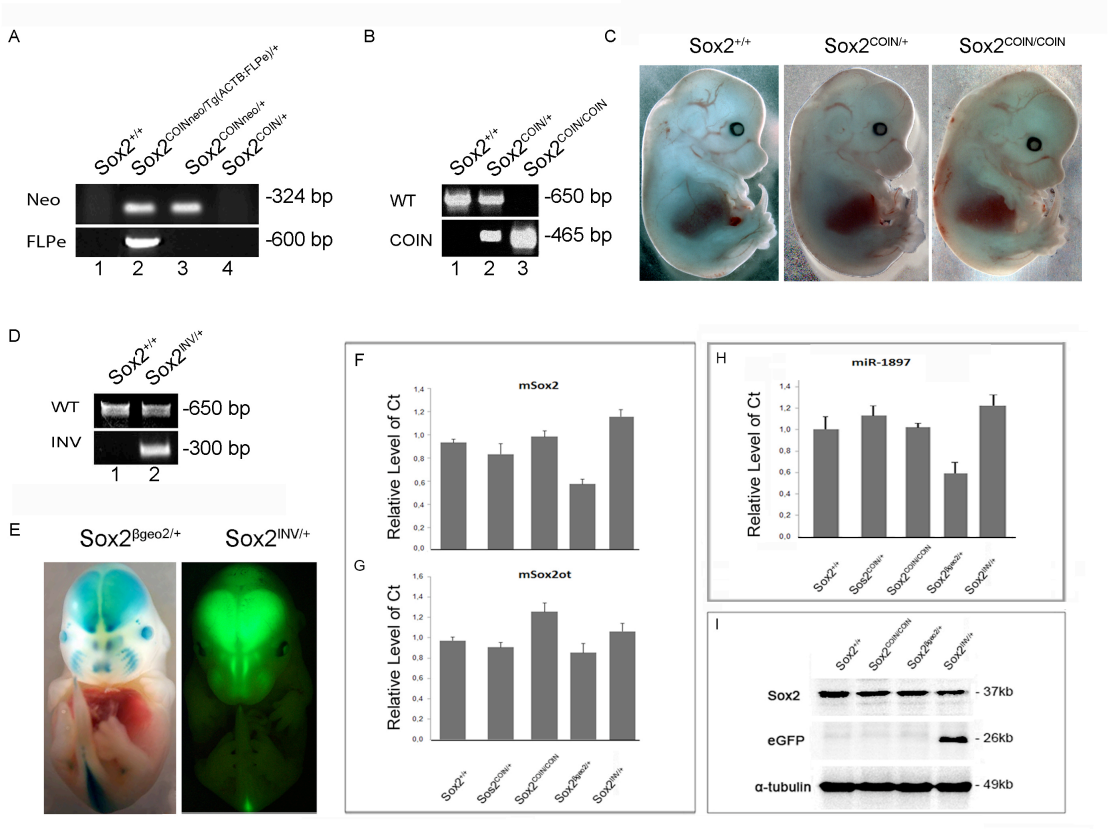


Figure 6. ***Sox2^{COIN}* is a functional conditional allele.** A) Efficient removal of the *neo* cassette by Flpe recombinase to generate a *Sox2$^{COIN/+}$* allele. *Sox2$^{+/+}$* (lane 1), *Sox2$^{COIN/+}$* (lane 2), *Sox2$^{COIN/COIN}$* (lane 3), *Sox2$^{INV/+}$* (lane 4), *Sox2$^{βgeo2/+}$* (lane 5) E14.5 mice. PCR genotyping of (1) *Sox2$^{+/+}$*, (2) *Sox2$^{COIN/+}$ Tg(ACTB:FLPe)*, (3) *Sox2$^{COINneo/+}$*, (4) *Sox2$^{COIN/+}$*. B) PCR genotyping of *Sox2$^{+/+}$* (lane 1), *Sox2$^{COIN/+}$* (lane 2), *Sox2$^{COIN/COIN}$* (lane 3) (genomic DNA from tail biopsies). C) E14.5 *Sox2$^{COIN/COIN}$* embryos are morphologically indistinguishable from *Sox2$^{+/+}$* and *Sox2$^{COIN/+}$*. D) Efficient inversion of the COIN module to generate *Sox2$^{INV/+}$* mice. PCR-based genotyping of (1) *Sox2$^{+/+}$* and (2) *Sox2$^{INV/+}$* (genomic DNA from tail biopsies). F, G) *Sox2* and *Sox2ot* qPCR analysis in *Sox2$^{+/+}$*, *Sox2$^{COIN/+}$*, *Sox2$^{COIN/COIN}$*, *Sox2$^{βgeo2/+}$* and *Sox2$^{INV/+}$* E14.5 embryos. Cyclophilin was used as a control label. Data are presented as the mean±SEM (n=3-6) for each genotype. The COIN module does not affect expression of *Sox2* prior to inversion. Additionallly, the COIN module does not appear to affect the expression level of *Sox2ot* either in *Sox2$^{COIN/COIN}$* or in *Sox2$^{INV/+}$* embryos. H) *MiR1897* qPCR analysis in *Sox2$^{+/+}$*, *Sox2$^{COIN/+}$*, *Sox2$^{COIN/COIN}$*, *Sox2$^{βgeo2/+}$* and *Sox2$^{INV/+}$* E14.5 embryos. I) Sox2 protein analysis. Western blot showing Sox2 and eGFP protein detected with specific antibodies in whole protein extracts from *Sox2$^{+/+}$*, *Sox2$^{COIN/+}$*, *Sox2$^{COIN/COIN}$*, *Sox2$^{βgeo2/+}$*, and *Sox2$^{INV/+}$* E14.5 embryos. The COIN module does not affect Sox2 protein levels prior to inversion.

3.5 *Sox2^{INV}* is a null allele of *Sox2*

Mice carrying a loss of function mutation in the *Sox2* locus have been generated by the insertion of a βgeo cassette into the *Sox2* locus (*Sox2*$^{\beta geo}$ (Avilion et al., 2003a), and *Sox2$^{\beta geo2}$* (Ekonomou et al., 2005)). Upon homozygosis, both alleles yield *Sox*2-null embryos that fail to form the epiblast and die around implantation. To test whether *Sox2$^{INV/INV}$* phenocopy *Sox2$^{\beta geo2/\beta geo2}$*, we performed *Sox2*$^{INV/+}$ intercrosses (Table V). No *Sox2*$^{INV/INV}$ offsprings were born. More specifically, *Sox2$^{INV/INV}$* mutants failed to survive shortly after implantation (Figure 7I), phenocopying *Sox2$^{\beta geo/\beta geo}$* and *Sox2$^{\beta geo2/\beta geo2}$* embryos. Only *Sox2*$^{+/+}$ and *Sox2$^{INV/+}$* embryos reach the embryonic stage of E6.5 (Figure 7A, G). Histological examination of whole decidual swellings harvested at 6.5 dpc revealed that 25% of deciduas carried abnormal implants, which had no egg cylinder structure and lacked the epithelial cells typical of epiblast (Figure 7I). Instead, many trophoblast giant cells could be identified (Figure 7C, H, I). The same phenotype is observed in *Sox2$^{INV/\beta geo2}$* embryos (Figure 7H). These results demonstrate the failure of *Sox2$^{INV/INV}$* embryos to develop an epiblast similarly to the *Sox2$^{\beta geo/\beta geo}$* and *Sox2$^{\beta geo2/\beta geo2}$* mutants. Thus, the inverted COIN cassette generates a true *Sox2* null phenotype.

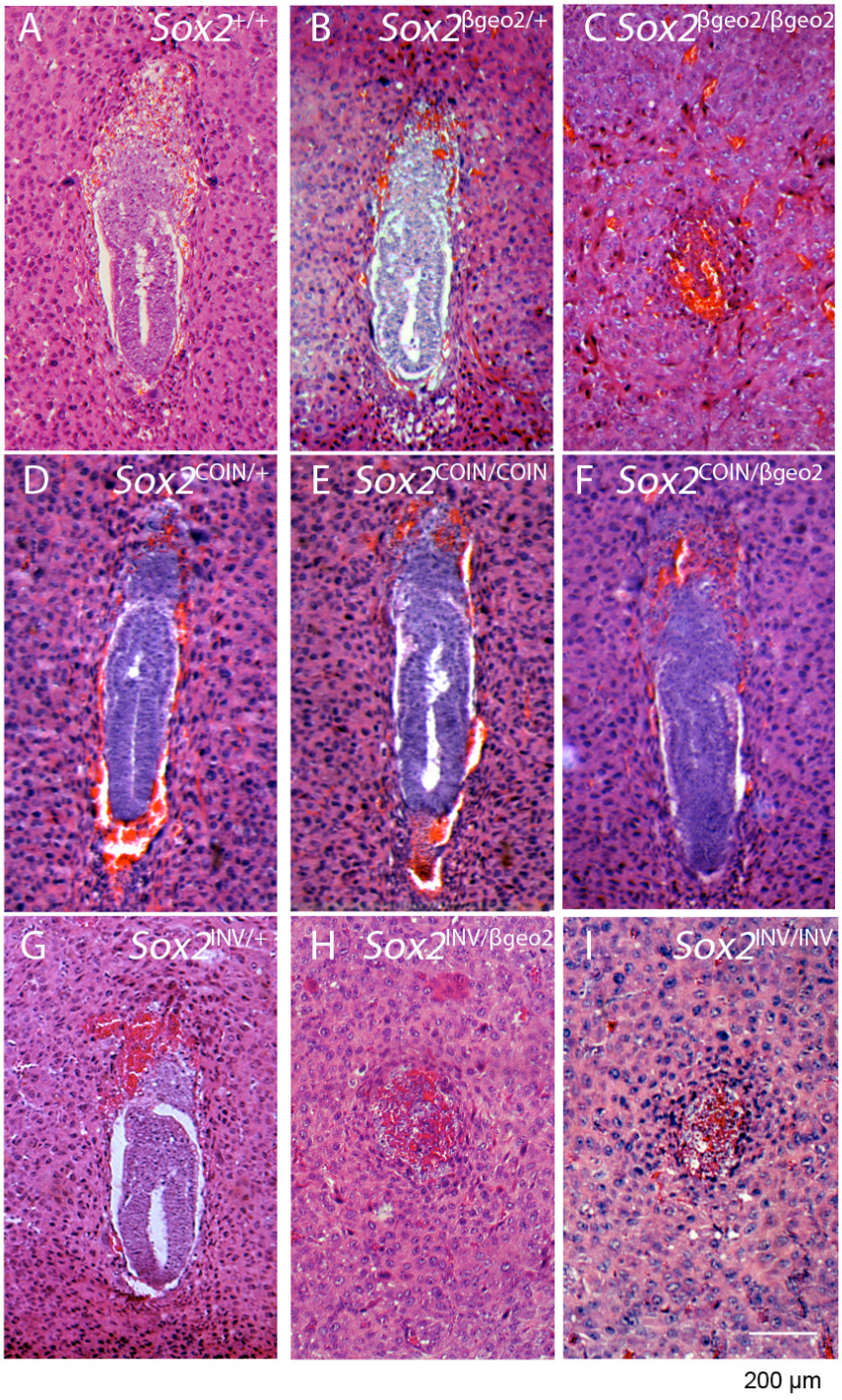


Figure 7. *Sox2*^{INV} is a null allele. Histological analysis of 6.5-dpc embryos from *Sox2*$^{βgeo2/+}$, *Sox2*$^{COIN/+}$, *Sox2*$^{INV/+}$ intercrosses. *Sox2*$^{+/+}$ A), *Sox2*$^{βgeo2/+}$ B), *Sox2*$^{COIN/+}$ D), *Sox2*$^{COIN/COIN}$ E), *Sox2*$^{COIN/βgeo2}$ F), and *Sox2*$^{INV/+}$ G) are all phenotypically normal. In contrast, *Sox2*$^{INV/INV}$ I) and *Sox2*$^{INV/βgeo2}$ H) embryos form disorganized extraembryonic tissues, but fail to form the epiblast, effectively phenocopying *Sox2* $^{βgeo2/βgeo2}$ C), and displaying the phenotype previously described for *Sox2* homozygous-null embryos. Sections were stained with hematoxylin and eosin (H & E).

Table III. Analysis of progeny from *Sox2*$^{COIN/+}$ intercrosses$^{\#}$

Genotypic distribution in live progeny*

	Live	Dead	*Sox2*$^{+/+}$	*Sox2*$^{COIN/+}$	*Sox2*$^{COIN/COIN}$
Age	No	No	No (%)	No (%)	No (%)
P21	95	0	27 (25%)	40 (50%)	28 (25%)

Genotyping of *Sox2*$^{COIN/+}$ heterozygous intercross progeny

#Data collected from mice in C57BL6 background

*Genotypes were assessed by PCR of genomic tail DNA

Table IV. Analysis of progeny from *Sox2*$^{INV/+}$ x *Sox2*$^{+/+}$ intercrosses[#]

Genotypic distribution in live progeny*

	Live	Dead	*Sox2*$^{+/+}$	*Sox2*$^{INV/+}$
Age	No (%)	No (%)	No (%)	No (%)
E6.5	17	0	9 (53%)	8 (47%)
P21	73	0	40 (55%)	33 (45%)

Genotyping of *Sox2*$^{INV/+}$ heterozygous intercross progeny

[#]Data collected from mice in C57BL6 background

*Genotypes were assessed by PCR either from tail biopsies or whole embryos

Table V. Analysis of progeny from *Sox2*$^{INV/+}$ x *Sox2*$^{INV/+}$ intercrosses[#]

Genotypic distribution in live progeny*

	Live	Dead	*Sox2*$^{INV/+}$	*Sox2*$^{INV/INV}$	*Sox2*$^{+/+}$
Age	No	No (%)	No (%)	No (%)	No (%)
E6.5	18	6 (33%)	5 (28%)	0	7 (39%)
P21	43	0 (0%)	21 (49%)	0	22 (51%)

Genotyping of *Sox2*$^{INV/+}$ heterozygous intercross progeny

[#]Data collected from mice in C57BL6 background

*Genotypes were assessed by PCR either from tail biopsies or embryo tissue

3.6 Epiblast inverted *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ mutants exhibit multiple embryonic defects

Sox2 null mutants (*Sox2*$^{\beta geo/ \beta geo}$ and *Sox2*$^{\beta geo2/ \beta geo2}$) are embryonic lethal around implantation (Avilion et al., 2003a; Ekonomou et al., 2005), thus masking the role of *Sox2* in the generation of tissue-specific stem cells during development and homeostasis. We made use of the novel conditional by inversion technology to generate a *Sox2*^{COIN} allele, taking into account the challenges of the *Sox2* locus*,* in that its proximal promoter and coding region are entirely contained within a CpG island, and are also spanned by an overlapping transcript, *Sox2Ot,* which contains *mmu-miR1897.* The inverted COIN allele of *Sox2, Sox2*^{INV} is functionally null, as homozygotes recapitulate the phenotype of *Sox2*$^{\beta geo/ \beta geo}$ and *Sox2*$^{INV/ \beta geo}$ embryos (Mandalos et al., 2012). In our previous work we generated epiblast inverted *Sox2*$^{INV/+}$ mice by *Sox2*$^{COIN/+}$ with Tg(*Sox2*CRE) intercrosses (Hayashi et al., 2002a, b; Hayashi et al., 2003; Vincent and Robertson, 2003). First, we wondered whether epiblast inverted null mice were viable, so we performed *Sox2*$^{COIN/+}$ (males) with *Sox2*$^{\beta geo2/+}$;Tg(*Sox2*Cre) (females) intercrosses. The Tg(*Sox2*CRE) transgene is active in the female germline, as it has reported to exist in the cytoplasm of the mouse female gametocyte. Consequently, all offspring that arise from female mice heterozygous for the Tg(*Sox2*CRE) transgene have demonstrable Cre activity, irrespective of whether they inherit the transgene itself (Hayashi et al., 2003). The Tg(*Sox2*CRE) mouse exert efficient Cre-mediated recombination in the epiblast, but not in extraembryonic tissues (Hayashi et al., 2002c, d; Hayashi et al., 2003; Vincent and Robertson, 2003). Excision of the floxed sequences by Tg(*Sox2*CRE) mice efficiently results in the visualization of eGFP in the epiblast (Figure 8). Analysis of the progeny showed that *Sox2*$^{\beta geo2/INV}$ embryos were not born confirming that *Sox2*^{INV} to *Sox2*$^{\beta geo2}$ intercrosses (Table VI) leads to embryonic lethality.

Epiblast is destined to derive all multipotent lineages in the mouse embryo. To study the role of *Sox2* in tissue-specific stem cells, we generated a conditional loss of *Sox2* function

mutation in the epiblast using *Sox2*^{COIN} X Tg(*Sox2*CRE) mouse intercrosses. We first performed intercrosses between *Sox2*$^{COIN/+}$ males with Tg(*Sox2*CRE) females to generate *Sox2*$^{INV/+}$ and *Sox2*$^{INV/+}$;Tg(*Sox2*CRE) adult progeny. *Sox2* is known to function as a cell fate determinant (Yamaguchi et al., 2011), so we harvested embryos derived from these crosses at E11.5, when organogenesis starts. Heterozygote E11.5 *Sox2*$^{INV/+}$ embryos do not show any obvious abnormalities and are indistinguishable from control *Sox2*$^{COIN/COIN}$ and *Sox2*$^{+/+}$ littermates (Figure 9Ai-iii)..

Table VI. Analysis of progeny from *Sox2*$^{COIN/+}$ x *Sox2*$^{\beta geo/+}$/Tg(*Sox2*:CRE) intercrosses[#]

	Genotypic distribution in live progeny							
	Live	Dead	*Sox2*$^{+/+}$	*Sox2*$^{COIN/+}$	*Sox2*$^{\beta geo/+}$	*Sox2*$^{INV/+}$	*Sox2*$^{\beta geo/INV}$	*Sox2*$^{+/+}$/ Tg(*Sox2*:CRE)
Age	No (%)	No (%)	No (%)	No (%)	No (%)	No (%)	No (%)	No (%)
P21	27(100%)	0	19(70.3%)	0	6(22.2%)	1(3.75%)	0	1(3.75%)

[#]Data collected from mice in CBAxC57 (F1) background

*Genotypes were assessed by PCR either from tail biopsies or from embryonic yolk sac or whole embryos.

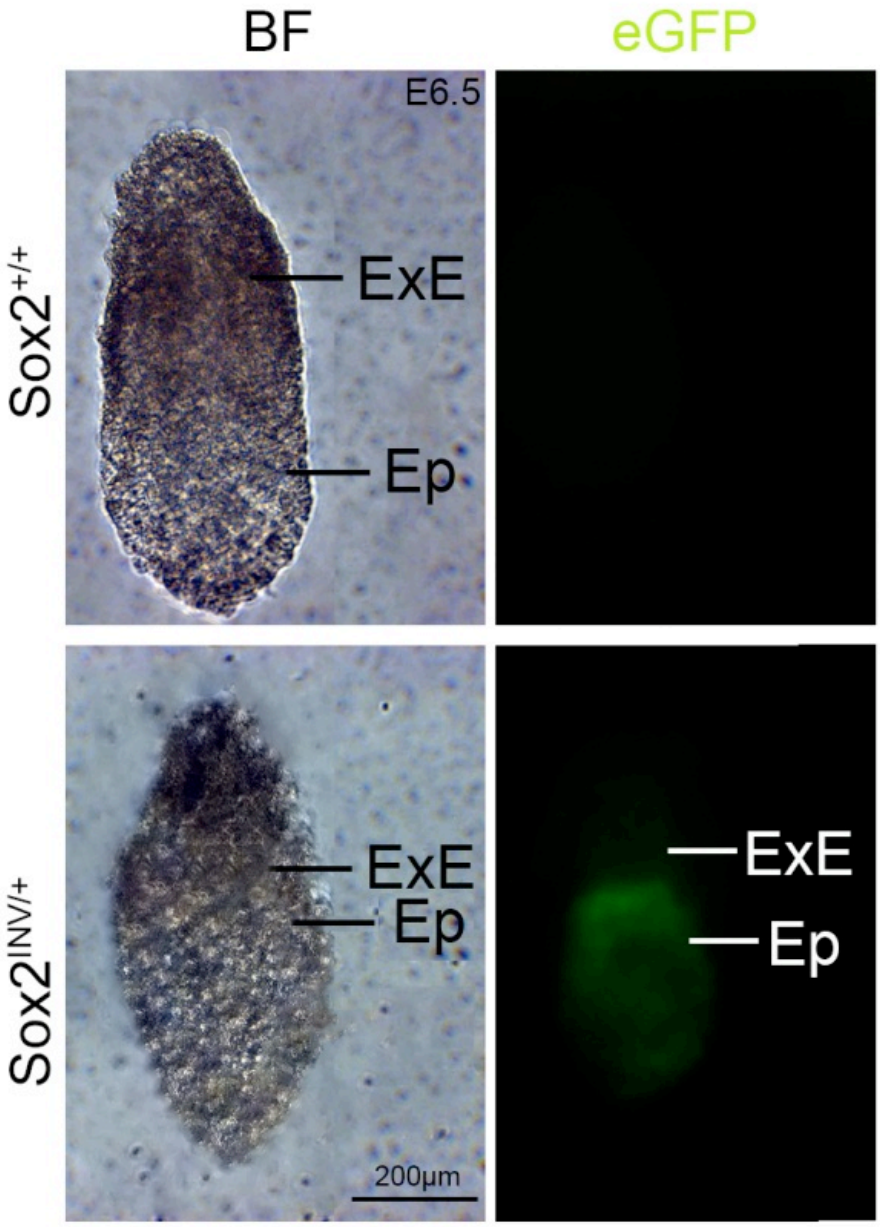


Figure 8. *Sox2* is expressed in the epiblast. *Sox2*$^{INV/+}$ E6.5 embryos show normal morphology compared with *Sox2*+/+ E6.5 embryos. eGFP field shows that *Sox2* is only expressed in the embryonic stem cells of the epiblast, which give rise to the three embryonic germ layers, and not in the extraembryonic tissue cells.

Additionally, E11.5 *Sox2*$^{INV/+}$ embryos express eGFP precisely at the areas where *Sox2* has been previously reported to be expressed (Avilion et al., 2003a) (Figure 9Avi). *Sox2*$^{INV/+}$ adult mice are fertile, feed normally, have normal body weight and normal lifespan. Thus, heterozygote *Sox2*$^{INV/+}$ mice, which theoretically express 50% of Sox2, show the same phenotypic properties during embryogenesis and adulthood as heterozygote *Sox2*$^{\beta geo2/+}$. Until now, heterozygote *Sox2*$^{INV/+}$ adult male mice do not show infertility problems as *Sox2*$^{\beta geo2/+}$ micehave. However, it is plausible to suggest that infertility problems in *Sox2*$^{\beta geo2/+}$ mice could appear due to the removal of regulatory regions at the 3' sequence of *Sox2* coding region, while in *Sox2*$^{INV/+}$ mice the whole sequence of the locus remains intact even after inversion.

To generate epiblast inverted *Sox2*$^{INV/+}$ embryos, we proceeded to *Sox2*$^{INV/+}$ intercrosses (Table VII) and harvested the embryos again at E11.5. Normally, if all embryos would survive at E11.5, we would expect that 75% of the embryos would be eGFP positive. However, we observed that only 50% of the embryos harvested were eGFP positive, suggesting that 25% of the embryos die in the deciduas at an early stage, (Table VII) most possibly before E6.5 as previously observed in Mandalos et al, 2012. Half the number of eGFP positive embryos had normal phenotype (Figure 9Bii, vi), while the remaining eGFP positive embryos (Figure 9Bvii, viii) represented haploinsufficient *Sox2*$^{INV/+}$ (*Sox2*$^{INV/+\ (H)}$) mutants, which exhibit multiple abnormalities (Figure 9Biii, iv). As haploinsufficiency in *Sox2* embryos has been previously observed when *Sox2* levels fall below 50% we assumed that *Sox2*$^{INV/+\ (H)}$ mutants would most possibly express *Sox2* in levels lower than 50%. This though remains to be proven experimentally Observed haploinsufficient embryonic phenotypes fall primarily in two categories: among littermates, a 10% has a significantly smaller size (Figure 9Biv), compared with *Sox2*$^{+/+}$ (Figure 9Bi, v) and *Sox2*$^{INV/+}$ littermates (Figure 9Biii, vii). The smaller size of these haploinsufficient embryos suggests that low levels of *Sox2* could affect proliferation during embryogenesis. From the remaining embryos, 13% of haploinsufficient

embryos exhibit significant hemorrhage defects, enlarged head structures in conjunction with craniofacial defects (Figure 9Biii). In order to assess whether *Sox2*$^{INV/+\ (H)}$ mutants survive beyond E11.5 during embryogenesis, we collected E12.5 and E15.5 embryos derived from *Sox2*$^{INV/+}$ intercrosses. As expected, only *Sox2*$^{INV/+}$ normal embryos were harvested. We also observed that in late E11.5 embryos which were stained with H&E (Figure 12), tissues have already started to deteriorate. Thus, we believe that these *Sox2*$^{INV/+\ (H)}$ embryos die probably at around E11 (Figure 9Biv, viii)(Table VII). Phenotypic differences among heterozygote *Sox2*$^{INV/}$ and *Sox2*$^{INV/+\ (H)}$ littermates, derived from *Sox2*$^{INV/+}$ intercrosses suggest there is indeed a narrow *Sox2* expression threshold, below which various phenotypic abnormalities appear during embryogenesis.

Table VII. Analysis of progeny from *Sox2*$^{INV/+}$ x *Sox2*$^{INV/+}$ *intercrosses*[#]

			Genotypic distribution in live progeny*			
	Live	Dead	*Sox2*$^{INV/+}$	*Sox2*$^{INV/+ (H)}$	*Sox2*$^{INV/INV}$	*Sox2*$^{+/+}$
Age	No	No (%)	No (%)	No (%)	No (%)	No (%)
E6.5	18	6(33%)	5(28%)	0	0	7(29%)
E11.5	43	11(25%)	11(25%)	10 (23%)	0	11(25%)

[#]Data collected from mice in CBAxC57 (F1) background

*Genotypes were assessed by PCR either from tail biopsies or from embryonic yolk sac or whole embryos.

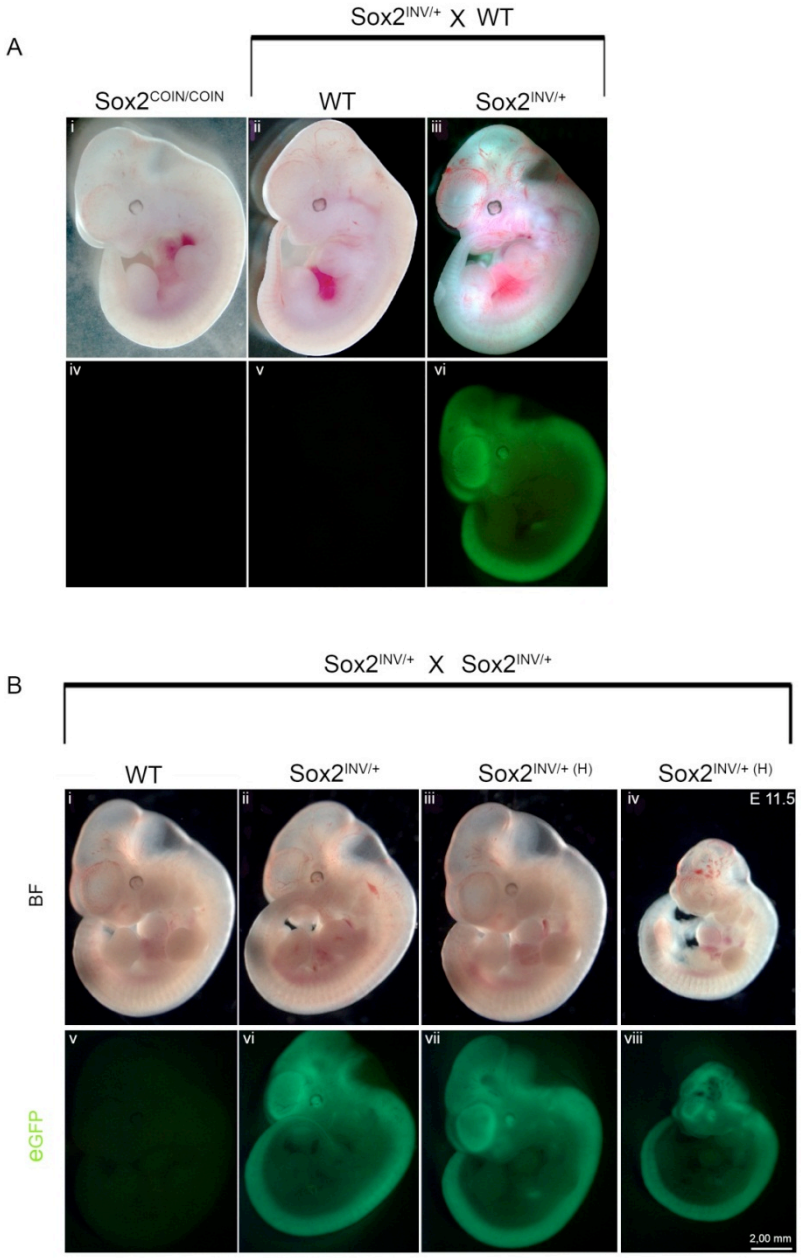


Figure 9. *Sox2*$^{INV/+\ (H)}$ embryos die around E11.5. A(i-vi), E11.5 *Sox2*$^{COIN/COIN}$ and *Sox2*$^{INV/+}$ embryos show no morphological differences when compared to *Sox2*$^{+/+}$. B(i-viii), *Sox2*$^{INV/+}$ intercrosses give rise to 15% of *Sox2*$^{INV/+}$ (H) E11.5 embryos which have normal size (iii, vii) and to 10% of *Sox2*$^{INV/+\ (H)}$ embryos which have smaller size (iv, viii). Both types of haploinsuffiecient mutant embryos show defects in the heart and in the frontonasal cavities. eGFP indicates that smaller size *Sox2*$^{INV/+\ (H)}$ E11.5 embryos (viii) have a disorganized ventricle in the forebrain, smaller midbrain and hindbrain and malformed branchial arches. The 25% of *Sox2*$^{INV/+}$ E11.5 embryos (ii, vi) from these crosses show no morphological defects compared with *Sox2*+/+ embryos (i, v).

To generate conditional mutants of *Sox2* in the epiblast, we performed *Sox2*$^{COIN/COIN}$ to *Sox2*$^{INV/+}$;Tg(*Sox2*CRE) intercrosses. We named these conditional embryos *Sox2*$^{INV/mosaic}$. These express the highest levels of eGFP and thus lower levels of *Sox2* than *Sox2*$^{INV/+\ (H)}$ embryos, with each litter harvested represent a gradient of *Sox2* expression (Figure 9A, B). We harvested *Sox2*$^{INV/mosaic}$ at E8.5, E9.5, E10.5 and E11.5. We were not able to harvest embryos beyond E11.5 from these crosses as well as *Sox2*$^{INV/mosaic}$ because embryos, similarly to *Sox2*$^{INV/+\ (H)}$, die around E11.5. *Sox2*$^{INV/+\ (H)}$ mutants and *Sox2*$^{INV/mosaic}$ mutant embryos exhibited severe abnormalities when compared to *Sox2*$^{+/+}$ littermates. Embryos are generated from *Sox2*$^{INV/+}$;Tg(*Sox2*-CRE) males with *Sox2*$^{COIN/COIN}$ female intercrosses at E11.5 (Figure 10a-d). As revealed from visualization, 90% of the embryos from these crosses were eGFP positive; 25% of them died early in development (empty deciduas), as previously described (Mandalos et al., 2012). Among littermates, only 10% of them were *Sox2*$^{INV/+}$, 20% are *Sox2*$^{INV/+\ (H)}$, while the remaining 70% of them were *Sox2*$^{INV/mosaic}$. *Sox2*$^{INV/+}$ fluorescent embryos do not show any abnormal phenotype (Figure 10a, e, I, n). E11.5 *Sox2*$^{INV/mosaic}$ embryos show severe malformations in anterior structures (Figure 10b-d) and develop an edema in the midbrain area (Figure 10b, f), possibly due to brain fluid homeostasis defects (Benarroch, 2009; Shevaga, 1981). Embryos also show vascular defects in the midbrain region and occasionally more severe ts in the forebrain (Figure 10c, g, d, h). Vascular defects occasionally can be caused due to defective pharyngeal arch development, which in turn would further affect the development of the pharyngeal arch arteries and heart (Que et al., 2009; Vitelli et al., 2002), causing abnormally high blood pressure (Koyanagi et al., 2010). We also observed ectopic tissue development, which express *Sox2* (eGFP positive cells), in the spinal cord (Figure 10l, g) and in the forebrain/midbrain junction (Figure 10d, h), indicating that downregulation of *Sox2* causes the formation of ectopic tissue masses.

Then, we again crossed *Sox2*$^{INV/+}$;Tg(*Sox2*-CRE) females with *Sox2*$^{COIN/COIN}$ males and harvested again embryos at E11.5 (Figure 10i-m). All embryos isolated from both types of

crosses were eGFP positive; while 15% of the embryos died early in development (empty deciduas). Among littermates, only 10% of them were *Sox2*$^{INV/+,}$ and 75% of them were *Sox2*$^{INV/mosaic}$. Again as before, high levels of eGFP indicate low levels of *Sox2* expression (Figure 10n-r). E11.5 embryos with low eGFP expression show no obvious defects (Figure 10n, o). As gradually *Sox2* becomes inactivated (with higher eGFP expression) (Figure 10p), we observe obvious mild enlargement of the forebrain ventricle. Embryos which show medial expression of eGFP show milder craniofacial defects (Figure 10q), while the ones with the highest expression of eGFP among littermates show severe malformation in the developing brain and craniofacial development (Figure 10r). All mutants develop heart abnormalities with a prominent one chambered heart and abnormal heart beat rates.

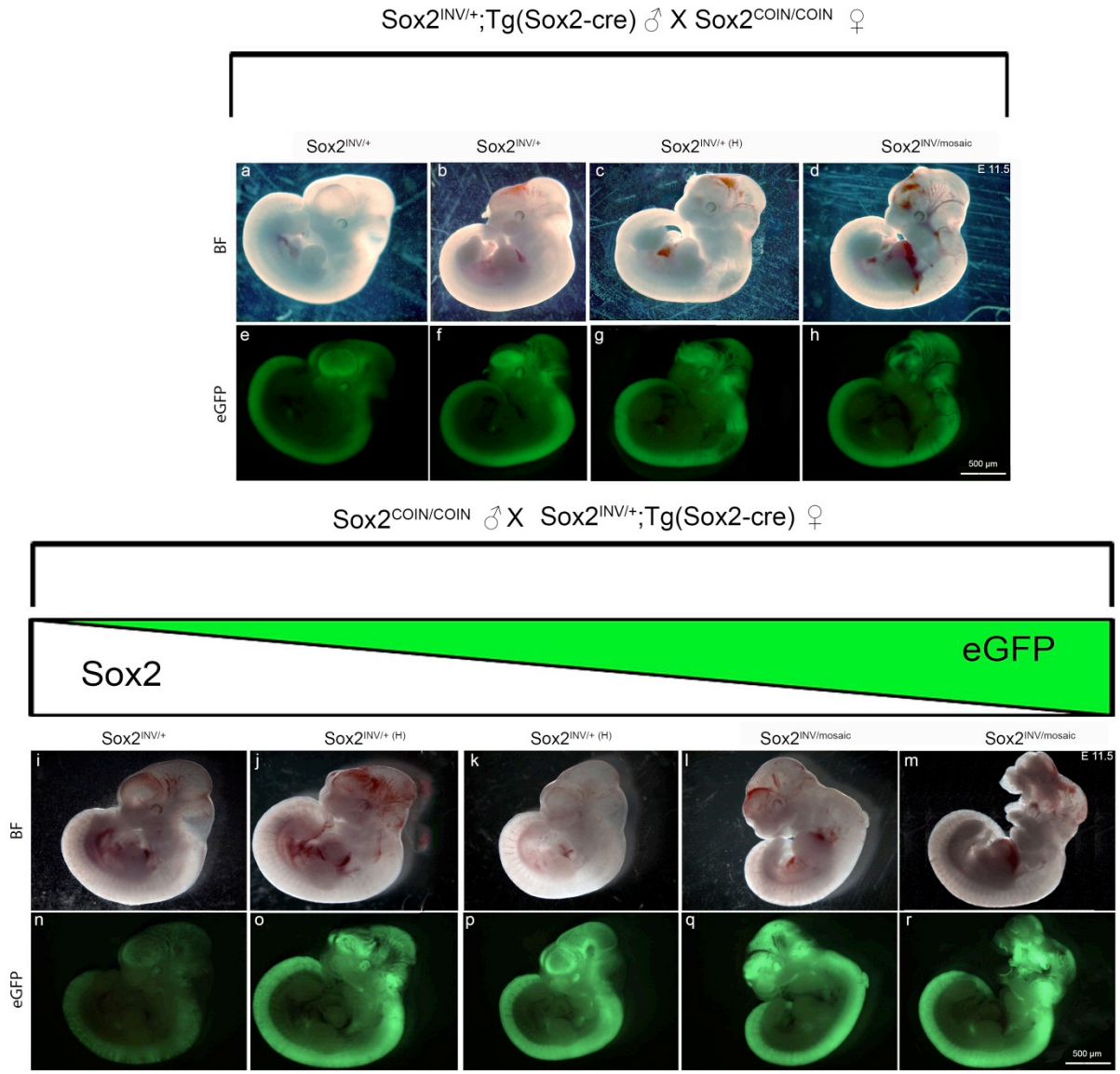


Figure 10. Dosage-dependent heart, brain and craniofacial abnormalities in *Sox2*$^{INV/mosaic}$ conditional mutants. (a) *Sox2* haploinsufficient embryos exhibit defects in the frontonasal area. (b-c) or severe frontonasal abnormalities, edema or development of ectopic tissue. (d) in the midbrain, plus failure of the ventricles in the forebrain to form. (e-h) Downregulation occurs in the forebrain, midbrain, hindbrain regions along spinal cord and in the retina. (i-l) E11.5 *Sox2*$^{INV/mosaic}$ embryos show more severe phenotypes when *Sox2*CRE transgene originates from the mother. (m) while 10% of the littermates are have a closure defect in the craniofacial area. (n-r)

3.7 *Sox2*$^{INV/+}$ haploinsufficient and *Sox2*$^{INV/mosaic}$ mutants develop heart looping defects and die around E11

Sox2$^{INV/+\ (H)}$ and *Sox2*$^{INV/mosaic}$ mutants die around E11. In these mutants; heart fails to develop due to abnormal looping, and undergoes remodeling of the primitive heart tube into a one-chambered organ (Figure 11a-c). Embryos die soon after, apparently from circulatory insufficiency (Abu-Issa and Kirby, 2007; Targoff et al., 2008; van den Berg et al., 2009). That was initially surprising as *Sox2* is not expressed in the developing heart, at any stage during development. To investigate the origin of this defect, we performed RNA in situ hybridization for genes involved in heart development, starting with *Nkx2.5*, one of the essential genes expressed during gastrulation in the lateral plate mesoderm responsible for the formation of precardiac cells (Durocher et al., 1997) (Harvey, 1996). In the mouse, targeted ablation of Nkx2.5 allows for the heart tube formation, but blocks heart development during looping morphogenesis (Lyons et al., 1995). Homozygous *Nkx2.5* null embryos show arrest of cardiac development at looping stage, poor development of blood vessels and severe defects in vascular formation and hematopoiesis (Kim et al., 2011). *Sox2*$^{INV/+\ (H)}$ and *Sox2*$^{INV/mosaic}$ mutant hearts expressed *Nkx2.5*. In the control embryo, the atrioventricular canal had a narrow luminal diameter and the ventricle (the future left ventricle), the bulbus cordis (the future right ventricle) and the outflow tract was already formed at this stage (Stallmeyer et al., 2010; Wang et al., 2011; Warren et al., 2011). We observed that the outflow tract in mutant embryos was misplaced, and the atrial region was not located posteriorly and on the left side, indicating that the rightward looping of the heart tube does not occur (Figure 11d, e). A reason for that may be that the atrioventricular canal was still wide open and a single ventricle was abruptly connected to a poorly developed outflow tract in mutant embryos (Tanaka et al., 1999). These results suggest that *Sox2* loss does not alter the initial heart development. As a result, defective looping in *Sox2*$^{INV/+\ (H)}$ and *Sox2*$^{INV/mosaic}$ mutant embryos is independent of *Nkx2.5* function.

Mutant hearts showed reduced expansion of the left ventricle consistent with hemodynamic obstruction. Similar abnormal heart phenotypes give rise to small thin-walled ventricles and disrupted myocardium at E10.5 (Risebro et al., 2009). One of the genes that play a crucial role during cardiac morphogenesis and performance is *Prox1* (Risebro et al., 2009). Prox1 ablation causes, among other abnormal embryonic phenotypes, impaired cardiac muscle growth and aberrant myocyte arrangement at E10.5 (Risebro et al., 2009). At E9.5, *Prox1* is expressed in a subpopulation of blood ECs (BECs) in the anterior cardinal vein and ECs bud from the veins and form embryonic lymph sacs and the lymphatic vasculature (Srinivasan et al., 2007). *Prox1* mRNA in situ hybridization was used to assess whether E9.5 $Sox2^{INV/+\ (H)}$ mutants suffer from heart failure. *Sox2* haploinsufficient mutant embryo hearts ectopically express *Prox1* when compared to $Sox2^{+/+}$ embryos. In mutant hearts, the presumptive right atrium was reproducibly expanded (up to 2-fold) and blood-filled, never appearing to empty appropriately, during the cardiac cycle, suggesting that there might be impaired blood flow through the heart, something that may explain the slow heart rate in mutant hearts compared with the $Sox2^{+/+}$ ones (Figure 11f-h). These results suggest that *Prox1*, could be similarly to *Sox1* in neural progenitors (Elkouris et al., 2011), a downstream target of *Sox2* negatively regulated during heart looping development to control blood circulation during heart development.

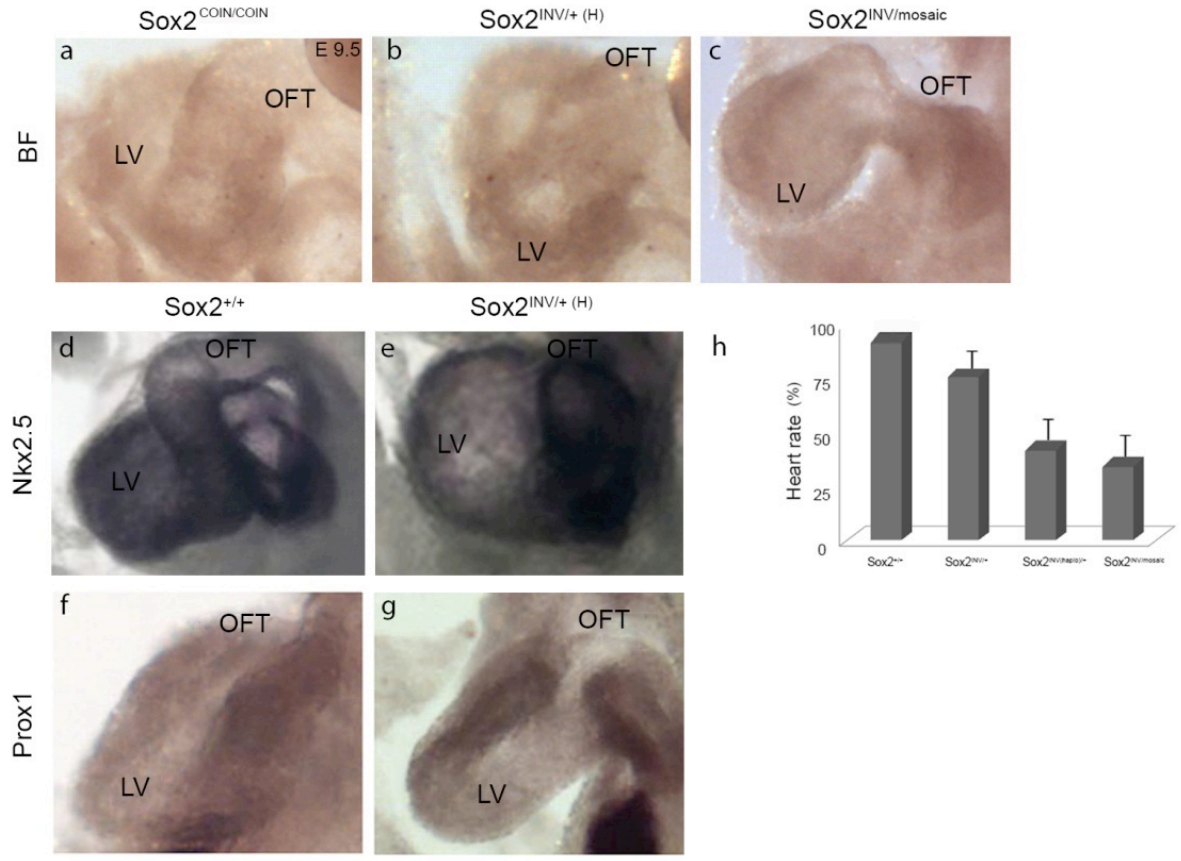


Figure 11. *Sox2* haploinsufficient and *Sox2*$^{INV/mosaic}$ mutant embryos exhibit heart looping defects. (a-c) E9.5 *Sox2*$^{INV/mosaic}$ hearts suffer from a heart-looping defect. E9.5 *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ hearts have a shorter outflow tract and only the left ventricle has developed. (d-e) In situ mRNA hybridization for Nkx2.5 in *Sox2*$^{+/+}$ and *Sox2*$^{INV/+ (H)}$ at E9.5 hearts. Nkx2.5 is expressed in the outflow tract and in the developing left ventricle of *Sox2*$^{INV/+ (H)}$ heart. (f-g) In situ mRNA hybridization for Prox1 in *Sox2*$^{+/+}$ and *Sox2*$^{INV/+ (H)}$ E9.5 hearts. Prox1 is ectopically expressed in *Sox2*$^{INV/+ (H)}$ hearts. (h) *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ heartbeat at E9.5 is reduced compared with *Sox2*$^{+/+}$ and *Sox2*$^{INV/+}$.

3.8 *Sox2*$^{INV/+(H)}$ and *Sox2*$^{INV/mosaic}$ mutants exhibit serious craniofacial abnormalities

As reported above, defects due to the conditional ablation of *Sox2* in the heart are mainly observed in SHF cells, which share common progenitor cells in pharyngeal mesoderm with craniofacial skeletal muscles (Zaffran and Kelly, 2012). Thus , it may not be surprising that *Sox2*$^{INV/+(H)}$ and *Sox2*$^{INV/mosaic}$ mutant embryos exhibit serious craniofacial abnormalities (Figure 10d. h, i, q) as well. Several studies have shown that regulation of epithelial to mensenchymal transition during NCC development plays a crucial role for the normal development of the cleft palates and nasal cavity (Kang and Svoboda, 2005). In order to examine how *Sox2* loss affects craniofacial development, we initially harvested and sectioned E11.5 *Sox2*$^{INV/+ (H)}$ mutants and stained them with hematoxylin and eosin stain. These mutants display severe developmental anomalies in the craniofacial area (Figure 12), that simulate the development of cleft palate in humans (Panetta et al., 2008) and account for the loss of the lower jaw. Furthermore, an enlargement of thehindbrain area and of the mesencepahlic and telencephalic vesicles in the same embryos was observed with consequent loss of cell mass. This observation suggests that there is a disruption of epithelial cells lining the ventricles in E11.5 mutant embryos (Figure 11) when compared to *Sox2*$^{+/+}$ embryos that show normal formation of frontonasal area, telencepahlic and forth vesicle and buccal cavity. We harvested E10.5 *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ mutants and detected *Sox2* by RNA in situ hybridization, to examine the levels of *Sox2* responsible for these defects. Then, we detected that *Sox2* is absent in the spinal cord at E10.5 of both type of mutants, while there is a high level of expression at the tail tip of *Sox2*$^{INV/+ (H)}$ mutant and even lower expression at the tail of *Sox2*$^{INV/mosaic}$ mutant (Figure 13). However, as mentioned above, we see no obvious morphological defects in the spinal cord at E9.5 and E11.5 embryos, indicating that downregulation of *Sox2* could be rescued by either *Sox1* or *Sox3* expression due to redundancy (Archer et al., 2011; Rex et al., 1997; Uwanogho et al., 1995a; Wood and Episkopou, 1999). Nevertheless, we observed a significant downregulation of

Sox2 in the frontonasal area, in the pituitary and telencephalic region, in the midbrain, hindbrain junction and in the retina (Hever et al., 2006). These areas exhibit most of the defects in *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ mutant embryos (Figure 13), indicating that *Sox2* is essential for the normal development of craniofacial and brain regions, as observed in older embryos and adult mice,. In *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ mutant embryos we have absence of the retina, a drastic enlargement of the telencephalic ventricle and severe edema along rhombencephalon rhombomeres 1-6. Finally, we observed absence of *Sox2* expression the neuroepithelium destined to generate migrating NCC cells in branchial arches 1 and 2, which give rise to a maxillary and mandibular process, and in the otic stalk, giving rise to the ear, of *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ mutant embryos. These results indicate that *Sox2* expression is essential for the proper development of the brain, mouth and nasal cavities, the eye and the ear and that defects observed on these structured in the absence of *Sox2* cannot be rescued by the expression of other *Sox*B1 family genes.

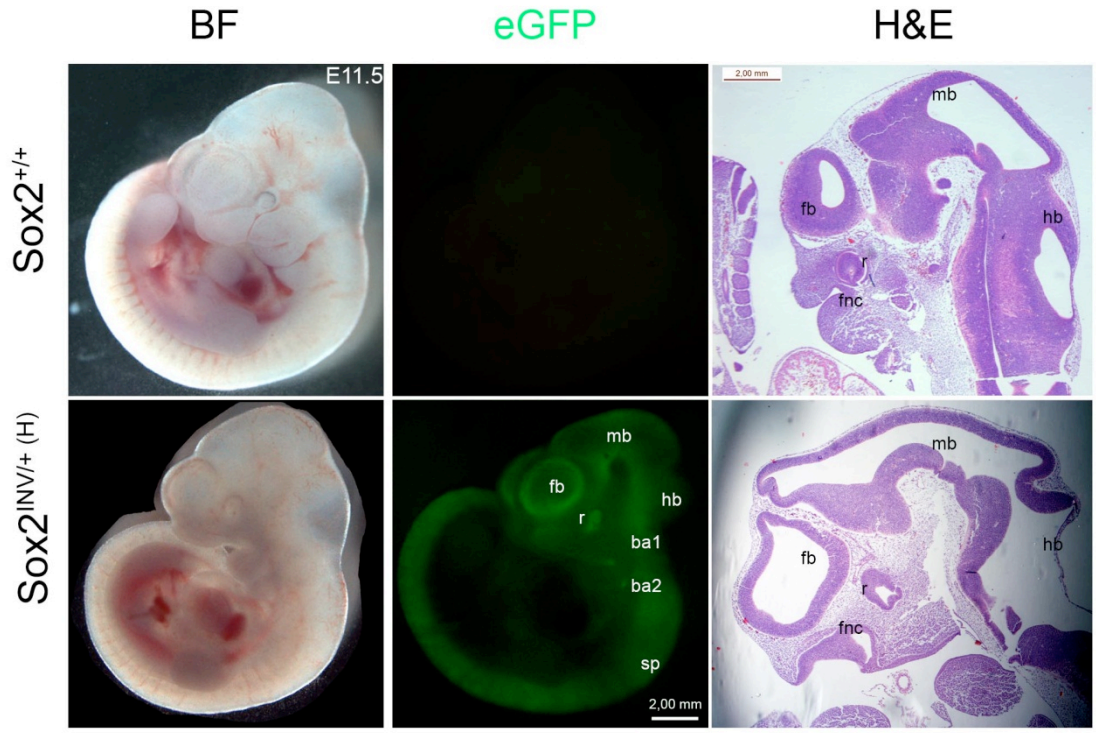


Figure 12. *Sox2*$^{INV/+ (H)}$ E11.5 embryos exhibit brain and craniofacial malformations. *Sox2* is downregulated in the forebrain (fb), midbrain (mb), hindbrain (hb), branchial arches 1 and 2 (ba1 and ba2), the spinal cord and in the retina (r). *Sox2*$^{INV/+(H)}$ embryos have enlarged forebrain and midbrain ventricles, absence of hindbrain structures, while the frontonasal cavity (fnc) is severely reduced.

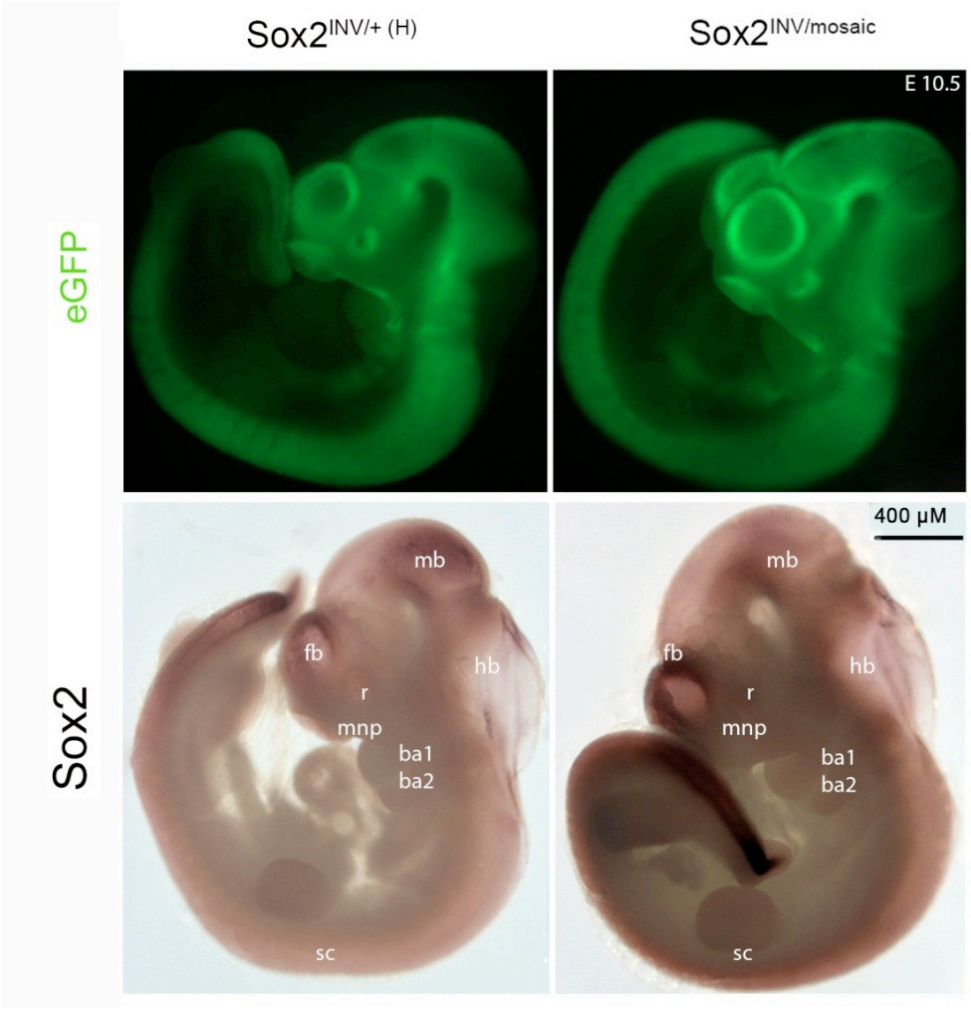


Figure 13. *Sox2* expression is dramatically reduced in *Sox2*$^{INV/+\ (H)}$ and *Sox2*$^{INV/mosaic}$ mutants. Note that eGFP expression is higher in *Sox2*$^{INV/mosaic}$ mutants than in *Sox2*$^{INV/+\ (H)}$. *Sox2* expression shows that *Sox2* is downregulated in the medial nasal process (mnp), in the retina (r), in the spinal cord (sc) and in the neuroepithilial cell layer of branchial arches 1 and 2 (ba1, ba2). *Sox2*$^{INV/+\ (H)}$ and *Sox2*$^{INV/mosaic}$ embryos have an edema in the hindbrain (hb), while forebrain (fb) and midbrain ventricles are clearly enlarged. Frontonasal development is severely affected in both mutants.

Hox genes play an essential role on early patterning events in the hindbrain and in the development of craniofacial structures. In order to study whether 0*Sox2* loss leads *Hox* gene deregulation, we analysed the expression of *Hoxa2* and *Hoxb1* at E8.5 *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ mutant embryos. For Hoxa2 no reports for potential gegentic interactions with Sox2 have been reported. *Hoxa2* has a rostral limit of expression in the rhombencephalic neural tube corresponding precisely to the boundary between rhombomeres 1 and 2 (Prince and Lumsden, 1994). Moreover, *Hoxa2* is an integral component of NCCs morphogenetic program during migration, while subpopulations of postmigratory NCCs require Hoxa2 at discrete time points to pattern distinct derivatives (Santagati et al., 2005). It has been also reported that null *Hoxa2* embryos lack craniofacial and cartilage elements derived from the first and second branchial arch which eventually causes ectopic expression of elements derived from them, such as ossification centers of the middle ear (Rijli et al., 1998; Trainor and Krumlauf, 2001; Vieille-Grosjean et al., 1997). Mutants die perinatally with cleft palate. Overexpression of *Hoxa2* in branchial arches 1 and 2 causes drastic reduction of first arch cartilages, such as Meckel's (midbrain derived) and the quadrate (hindbrain derived). *Hoxa2* is downregulated in r3 but not in r5 of E9.5 *Sox2*$^{INV/mosaic}$ mutant embryos suggesting a novel role for *Sox2* in the regulation of *Hoxa2* in vivo (Minoux, 2011).

Hoxb1 has been reported to be regulated by *Sox2*. At E9.5 it is expressed throughout rhombomere 4 and along the spinal cord. It is required for the specification of facial branchiomotor neuron progenitors that are programmed to innervate the facial muscles (Arenkiel et al., 2004). Null *Hoxb1* homozygous mice die postnatally with narrow face, runting, absent facial motor nuclei, and several facial nerve/muscle defects (Arenkiel et al., 2003; Ogura and Evans, 1995; Pata et al., 1999). Thus we analysed the expression of *Hoxb1* in our mutants. We observed that *Sox2* loss do not affect *Hoxb1* expression in rhombomere 4 in E8.5 *Sox2*$^{INV/+ (H)}$ and *Sox2*$^{INV/mosaic}$ embryos (Figure 14) and in the spinal cord (data not shown). Thus, our results do not support previous in vitro reports on *Sox2* involvement in the regulation of *Hoxb1*, at least at E8.5 in the hindbrain and spinal cord regions.

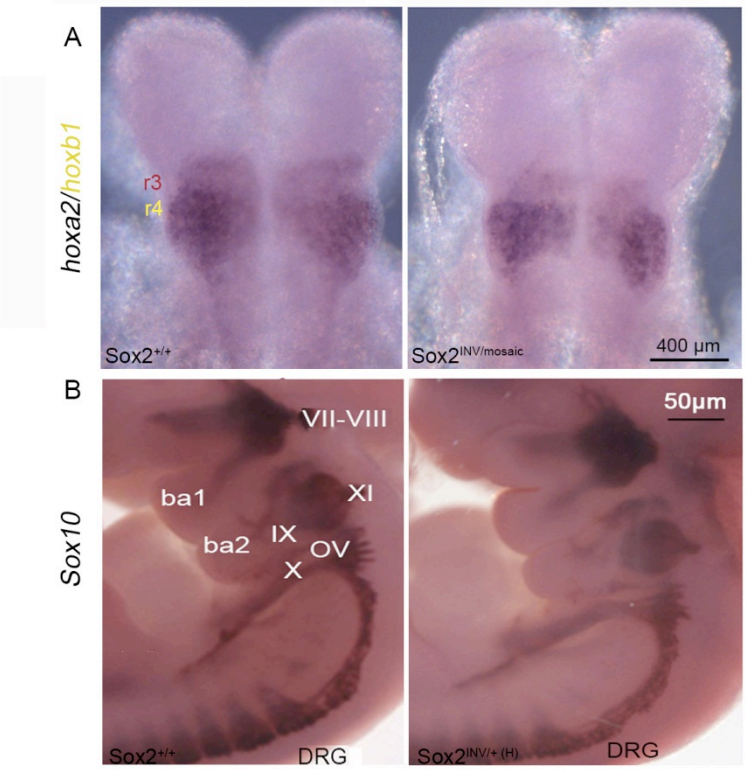


Figure 14. In situ hybridization for *Hoxa2* and *Hoxb1* at E8.5 *Sox2*$^{INV/+}$ and *Sox2*$^{INV/mosaic}$ embryos. *Hoxa2* is downregulated in rhombomere 2 of *Sox2*$^{INV/mosaic}$ embryo. *Hoxb1* is normally expressed in the spinal cord and in r4 in *Sox2*$^{INV/mosaic}$ mutants.

As Hoxb1 has been shown to be important for facial muscle innervation, we then examined cranial nerve formation at E11.5 mutant embryos. *Sox2* loss does not affect cranial nerve function in the ganglia of the spinal accessory (XI), vagus (X), glossopharyngeal (IX), and branches of facial (VII) nerves. It is believed that expression of *Sox10* is essential for the myelination of these nerves as well as for the regulation of peripheral glial development. Our results show that *Sox2* is not involded directly in the formation of the developing cranial nerves. Thus, innervation of the facial structures proceeds normally in *Sox2* conditional mutants.

*Sox*E (*Sox8*, *Sox9*, and *Sox10*) genes play a crucial role in NCC development. *Sox10* is critical for the formation of NCCs, maintenance of multipotency, specification and differentiation (Kelsh, 2006). *Sox10* is the only *Sox*E gene that maintains its expression during migration of NCCs (McKeown et al., 2005). . The understanding of *Sox10* function in NCC development is of particular significance since *Sox10* mutations lead to several craniofacial abnormalities in humans, called neurocristopathies, including Waardenburg-Hirschsprung syndrome and peripheral neuropathies (Hoke, 2012). *Sox10* overexpression arrests the neuroepithelial cells and cranial mesenchymal cells in an undifferentiated state, causing a range of cell fate specification defects (Ahlstrom and Erickson, 2009) .Neural progenitor cells, which overexpress *Sox10,* fail to form neuronal, Schwann, or melanocyte cells (Stolt et al., 2008). On the other hand, *Sox10* null embryos lack peripheral glial cells, melanocytes, and autonomic and enteric neurons, and die around birth (Betancur et al., 2011; Bremer et al., 2011; Lupski, 2010; Parthey et al., 2012; Takada et al., 2010). Thus, we performed *Sox10* mRNA in situ hybridization in E9.5 *Sox2*$^{INV/mosaic}$ mutant embryos, as at that embryonic stage migrating NCCs are apparent along the lateral surface of the neural tube (i.e. the ventromedial portion of the sclerotome). Additionally, few NCCs are present along the ventrolateral portion of the sclerotome at E9.5, as well as in derivatives from the ventral region of the neural tube while they migrate along the dorsolateral pathway (Serbedzija et al., 1990). We observed that *Sox10*+ cells expressing high levels of *Sox10* are present in the frontonasal region and the forebrain and in the branchial arches 1-2 in *Sox2*$^{INV/+ (H)}$ mutants

(Figure 14A). Downregulation of *Sox2* causes a dramatic upregulation of *Sox10* and an outflow of Sox10+ cells in the hindbrain, forebrain and frontonasal areas; in regions where normally is expressed as well as in areas where *Sox10* is normally not expressed in *Sox2*$^{+/+}$ control embryos (Figure 15). These cells ectopically express *Sox10* in *Sox2*$^{INV/mosaic}$ mutant embryos remain undifferentiated. We believe that according to previous reports and these events may be linked to the defects that we observe in these regions. Generally, in *Sox2*$^{INV/+}$ $^{(H)}$ and *Sox2*$^{INV/mosaic}$ embryos *Sox10* is overexpressed in the craniofacial areas at E9.5 embryos, suggesting that migrating NCCs from branchial arches remain undifferentiated at this area, leading to the malformation of the frontonasal elements (Figure 15). Our data suggest that *Sox2* may act as a suppressor of *Sox10* expression, a fact that reinforces the idea that *Sox2* acts antagonistically with *Sox10* in the anterior axis during NCC formation (Adameyko et al., 2012). In this way, *Sox2* acts as a master regulator of the epithelial to mesenchymal transition leading to neural crest development in mice and possibly in humans.

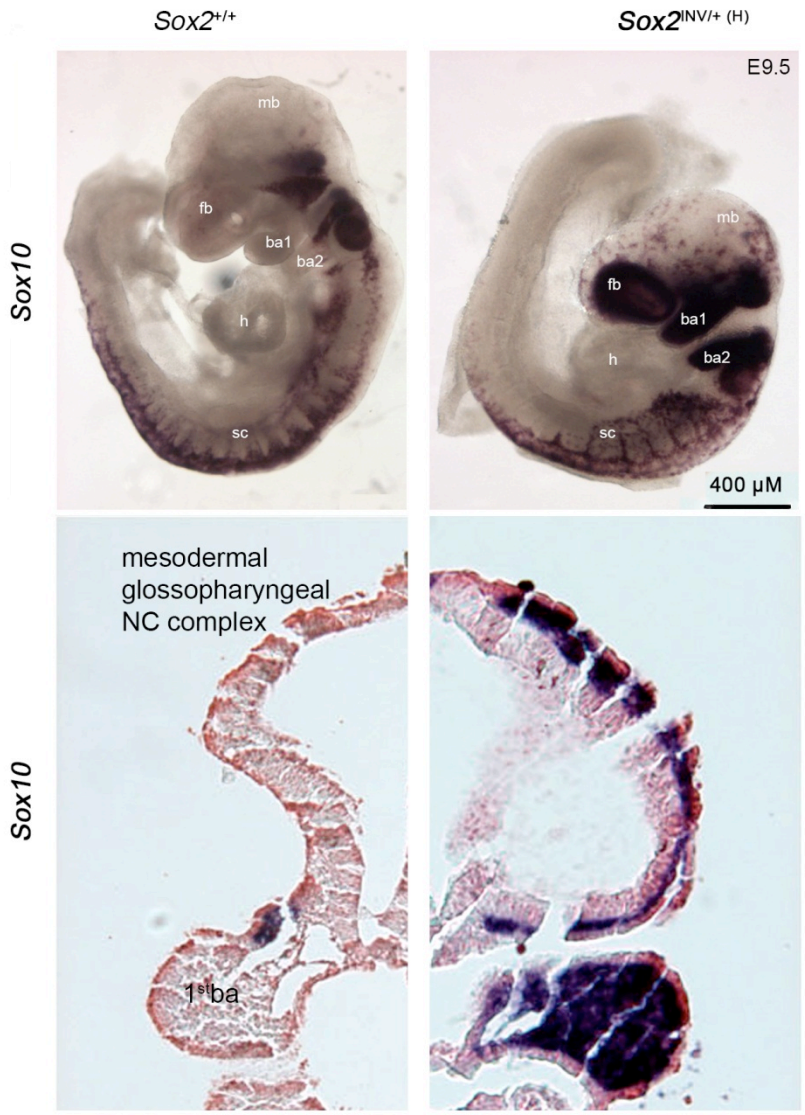


Figure 15. A. NCC formation and migration is exacerbated in *Sox2*$^{INV/+ (CH)}$ mouse embryos. *Sox10* is highly upregulated in branchial arches 1 and 2 (ba1, ba2), in the forebrain (fb), and in the frontonasal area of *Sox2*$^{INV/+ (H)}$ embryos when compared to *Sox2*$^{+/+}$ embryos.

4. Discussion

4.1 COIN technology

We show here the application of COIN technology to generate a conditional-null allele for a single exon gene, *Sox2*. The COIN method was invented at least in part to overcome the challenges and limitations of traditional site-specific recombinase-based strategies such as Cre/Lox for designing conditional alleles (Nagy, 2000). These include the placement of Lox sites as well as the distance between them (Ringrose et al., 1999), defining critical exons (i.e. the exons of the gene that need to be deleted by Cre in order to bring about the desired allelic state) (Skarnes et al., 2011; Testa et al., 2004), and the lack of unified strategy for including a reporter that can mark those cells that harbor the post-recombinase allele. To date, COIN has been successfully applied in generating conditional alleles for more than twenty-five protein-coding genes (Economides et al., 2012), but its applicability to single exon genes has not been tested.

Single exon genes present a design challenge for engineering conditional alleles by traditional, e.g. simple floxing, methods. First, the Lox sites should be placed in a position that does not affect the expression of the target locus, a design decision that can be complicated by the lack of specific knowledge of the exact position of promoters and regulatory elements. An additional design challenge is presented if a reporter that marks the conversion from 'wild-type' to null is desired. The COIN method addresses both of these challenges irrespective of gene structure by avoiding the placement of Lox and FRT sites, reporters, and other functional elements within regions upstream of the target gene's coding sequence (Economides et al., 2012). Instead, COIN employs an 'exon-splitting' artificial intron to place an optimized module – the COIN module – within a coding exon, yet in the antisense strand. Perusal of the regulatory elements mapping within single coding exon of human Sox2 suggests that the COIN intron was inserted at a position that does not result in disruption of any such elements (data not shown) As a result, the COIN module is stealth to transcription, and does not alter the expression of the modified gene. Although it is possible

that introduction of the COIN module intron alters the kinetics of transcription (Seoighe and Korir, 2011; Swinburne et al., 2008; Swinburne and Silver, 2008), we have not examined this possibility at the single cell level; at the population level and at steady state, the level of *Sox2* mRNA as expressed from the $Sox2^{COIN}$ locus does not appear different to that of wild type. Lastly, because in COINs the Lox and FRT sites are placed within the artificial intron, they do not disrupt of promoters or regulatory sequences, and are also not incorporated into mRNA.

The particular choice of *Sox2* to test the COIN method's applicability to single exon genes presented additional challenges in that the majority of *Sox2*'s single exon is contained within a CpG island, and there is also an overlapping non-coding transcript *Sox2ot*. Due to design constraints – specifically the need to place the COIN module as near the initiating ATG as possible – the COIN module intron was inserted into *Sox2*'s exon in manner that disrupts the CpG island. However, this had no apparent effect on Sox2 expression and had not apparent phenotypic consequences in the mouse. *Sox2ot* levels also remained unaltered, indicating that at least in the antisense position the presence of the COIN module has no effect on the expression of *Sox2ot*. This was evident by the normal phenotype of $Sox2^{COIN/\beta geo2}$ embryos, in which only the COIN allele can generate wild type mRNA. This genotype should sensitize the embryo to any reduction in Sox2 levels, and thus provides a stringent comparison between the wild type allele and the COIN allele prior to inversion.

Equally important is the fact that post-inversion of the COIN module, the resulting allele, $Sox2^{INV}$ phenocopies the previously generated null alleles upon homozygosis. In addition, the reporter embedded in the COIN module, is expressed in a manner representative of Sox2 expression, thereby generating a tool to visualize Sox2 expression and to follow the conversion of the COIN allele into a null by Cre recombination.

In addition to the conditional-null allele presented here, four other conditional-null alleles of *Sox2* have been published (Favaro et al., 2009; Miyagi et al., 2008; Smith et al., 2009; Taranova et al., 2006). All four rely on floxing of the single exon of *Sox2*, though the

placement of the LoxP sites and selection cassettes (and their retention) varies among alleles. One of the main differences between these alleles and *Sox2*^{COIN}, is that they do not incorporate a reporter that is activated after Cre acts on the allele. There is however a paucity of published data such as expression analysis of *Sox2*, *Sox2ot*, and *miR1897* to allow further comparisons between *Sox2*^{COIN} and the previously described conditional-null alleles. Given the increasing evidence for roles that Sox2 plays in a wide range of pathological and patho-physiological conditions, assays for the normal regulation of *Sox2* expression need to be conducted in a variety of cell types. Overall, our results highlight the importance of the Conditional by Inversion technology as a method of choice in targeting intronless (single exon) genes, especially when complexity of the locus and desire for inclusion of a reporter are taken into consideration.

4.2 *Sox2* role in neural crest development

NCCs arise within the ectoderm during neuralation and give rise to most of the peripheral nervous system.

NCCs are induced by interactions between the adjacent neural and non-neural ectoderm (Dickinson et al., 1995) (Selleck and Bronner-Fraser, 1995) by a combination of signals, including Wnt proteins, bone morphogenetic proteins (BMPs) (Liem et al., 1995), fibroblast growth factors (FGFs), retinoic acid, and the Notch/Delta system (Aybar et al., 2002; Aybar and Mayor, 2002; Christiansen et al., 2002; Endo et al., 2002; Garcia-Castro et al., 2002; LaBonne and Bronner-Fraser, 1998; Villanueva et al., 2002; Wu et al., 2003). They emerge from the dorsal neural tube, migrate along defined pathways and differentiate into many cell types, including neurons and glia of the peripheral sensory and autonomic nervous ganglia, Schwann cells, and melanocytes, as well as chromaffin cells of the adrenal medulla, and much of the skeletal and connective tissue of the craniofacial complex (Ayer-Le Lievre and Le Douarin, 1982) (Figure 16). Induction of NCCs is triggered by the expression of target

genes (Gammill and Bronner-Fraser, 2002) including transcription factors such as *Sox8*, *Sox9*, and *Sox10* of the SoxE subgroup (Bowles et al., 2000). These genes are expressed in premigratory NCCs and later in distinct NCC-derived subpopulations. Among these genes, *Sox9* is the first to be expressed, at a similar time to FoxD3, overlapping with expression of BMP4, cadherin 6b, Slug, and RhoB (Liu and Jessell, 1998) (Cheung and Briscoe, 2003). *Sox10* is expressed shortly after *Slug*, while *Sox8* expression begins soon after *Sox10*, before migration from the neural tube (Cheng et al., 2000). As migration commences at trunk levels, Slug, RhoB, N-cadherin, and cadherin 6b are down-regulated (Akitaya and Bronner-Fraser, 1992) (Monier, 1995) (Jessel and Weiss, 1998; Nakagawa and Takeichi, 1995), while *Sox10* and *FoxD3* continue to be expressed in migratory NCCs (Cheng et al., 2000) (Dottori et al., 2001).

Of *Sox*B genes (Sox1, 2, 3, 14, 19, and 21), *Sox2* has been reported to play a cell-autonomous role in NCC development in mice (Pan and Schultz, 2011). *Sox2* is one of the early genes activated in the developing neural plate (Graham et al., 2003; Hutton and Pevny, 2011; Wen et al., 2008). Neural plate expression of *Sox2* is reduced as NCCs segregate and migrate from the dorsal neural tube NCC. Expression of *Sox2* in a subset of these cells that contribute to the developing peripheral nervous system is subsequently upregulated and gradually becomes restricted to NC derived glial sublineages (Aquino et al., 2006). Ectopic expression of *Sox2* in embryonic ectoderm and in neural plate explants reveals that *Sox2* expression is sufficient to inhibit NCC formation (Papanayotou et al., 2008; Remboutsika et al., 2011b) whereas later in the NCC pathway, i.e. migratory and postmigratory NCC in the peripheral nervous system, *Sox2* has a role in proliferation and differentiation (Wakamatsu, 2004). Another study implicates the Nk-1-related gene, Nbx, as having an upstream role in negative regulation of *Sox2* expression and positive regulation of the transcription factor Slug, which induces NC formation (Kurata and Ueno, 2003). A recent study has also shown that self-renewing NSCs are restricted to *Sox2* expressing cell (Remboutsika et al., 2011b). Although it is not believed to be essential in early NC development, *Sox2* has been reported

to prevent terminal differentiation of Schwann cells (Le et al., 2005; Wakamatsu et al., 2004). Other studies have also reported that *Sox2* also plays an important role in the specification of melanocytes by maintaining a NCC fate and repressing a melanocyte fate (Laga et al., 2010). Finally, a late study supported that *Sox2* is able to abort melanocyte differentiation in cells already specified to take up such a fate, and loss of *Sox2* function in mice leads to increased numbers of Mitf positive cells at the expense of *Sox10* positive NCCs (Adameyko et al., 2012).

To study the role of *Sox2* in stem cells, we developed a conditional inversion of *Sox2* in the epiblast (Mandalos, 2012). We revealed that downregulation of *Sox2* in epiblast-derived structures leads to severe craniofacial and heart defects, often described as neurocristopathies in humans (Cordero et al., 2011; Jain et al., 2010), a unifying concept of disease arising in NC maldevelopment (Butler and Willshaw, 1989), leading also to the formation of tumours along the neural axis.

Epiblast-inversion of *Sox2* affects strikingly cardiac development. Cardiac NCC is a subpopulation of the NCC and plays a crucial role in the cardiac outflow tract formation during development. Experimental studies have suggested that abnormal looping and convergence can cause misalignment of inflow and outflow segments of the hearts (Rossant, 1996). Epithelial-mesenchymal transition (EMT) is the process during which epithelial cells are transformed to mobile and developmentally plastic mesenchymal cells. All heart cells are generated from one or more EMT steps (Machado et al., 2012). In *Sox2* epiblast-inverted mouse embryos, we observed that EMT was disrupted during elongation of the heart tube by the addition of progenitor cells from adjacent pharyngeal mesoderm, leading to the failure of the heart tube to loop. As a result, all the subsequent steps following heart looping, including the formation of heart ventricles are delayed. Loss of cardiac NCC causes persistent truncus arteriosis due to the failure of the outflow tract to be separated. This defect is accompanied by a defect in myocardial contractility, with consequent ventricular dilatation, due to an intrinsic defect in the myocardium (Drysdale et al., 1997). Endocardial and epicardial EMT

create most of the noncardiomyocyte lineages of the mature heart. Endocardial EMT is responsible for the production of valve progenitor cells and is necessary for the development of cardiac valves and cardiac septation. Epicardial EMT is essential for normal myocardial growth and coronary vessel generation, forming cardiac fibroblasts, vascular smooth muscle cells and cardiomyocytes (von Gise and Pu, 2012). Although, we cannot exclude that minor cardiac neural crest defects can occur, the mutant hearts form normally. As *Sox2* is not expressed in the heart, analysis of signaling centers from the endoderm will reveal the mechanism employed by *Sox2* to control EMT in the heart.

During cardiac looping the heart tube elongates by addition of progenitor cells from adjacent pharyngeal mesoderm to the arterial and venous poles. This cell population is termed the second heart field (SHF). SHF contribute progressively to the poles of the elongating heart tube during looping morphogenesis, giving rise to myocardium, smooth muscle, and endothelial cells (Kelly, 2012). In mouse cardiac development, EMT occurs between E9.5-E10.5 and leads to the formation of rudimentary cushions in E10.5-E11.0 while it significantly diminishes by E13 (Camenisch et al., 2002). The initiation and cessation of EMT involves signals from the endocardium, adjacent myocardium and cushion mesenchyme (Person et al., 2005). Studies on mammalian systems have suggested that *TGFβ1*, *2* and *3* play important roles in the regulation of cardiac EMT (Mercado-Pimentel and Runyan, 2007). *TGFβ2* null mice exhibited a range of cardiovascular anomalies that resulted from failure of normal completion of looping and septation of the outflow tract and the AV canal, as well as abnormalities of valve differentiation and arterial growth and die around birth. *TGFβ2* null embryos present similar phenotypes with the *Sox2* epiblast-inverted mutants and that leads us to believe that *TGFβ2* could be an excellent candidate for a downstream effector of *Sox2* during heart development. Further analysis would be required to understand the role of *Sox2* in the heart looping most possibly during second heart field formation.

During development cranial neural crest (CNC) cells delaminate from the lateral ridges of the neural plate and then emigrate towards the developing branchial arches. CNC

cells migrate towards specific regions where they fuse with the existing population of mesodermal cells (Kouskoura et al., 2011). The proliferation of the CNC cells leads to the generation of tissues around the future oral cavity. CNC cell proliferation is responsible for the formation of the frontonasal process and of pairs of maxillary and mandibular processes. Later in development, these processes eventually give rise to the completed face (Bronner-Fraser, 2002). *Sox2* is important for the normal development of NCC and its derivatives (Cimadamore et al., 2012; Hutton and Pevny, 2011; Remboutsika et al., 2011b). The severity of defects in *Sox2* epiblast inverted mutants in the developing brain and facial structures underline a *Sox2* dosage-dependent mode of craniofacial development. We find that low levels of *Sox2* lead to the absence of almost the entire facial structures derived from the NCC domain facial skeletogenic (FSNC) of the posterior diencephalon, mesencephalon, hindbrain and rhombomeres 1 and 2. (Creuzet, 2009; Le Douarin et al., 2012).

Moreover, recent studies have also correlated *Sox2* expression with the clinical stage of neuroblastomas in humans (Yang et al., 2012). Further analysis of whetherif the absence of *Sox2* in NCCs is sufficient to trigger the formation of neublastoma cells needs to be done. These findings, in accordance with our studies and recent studies which show that *Sox2* levels in neural stem cells control NCC development (Remboutsika et al., 2011b), indicate that the levels of *Sox2* during the formation of neural plate directly affect the expression of downstream effector genes such as *Sox10* and *Hoxa2*. The exact role of *Sox2* in the genetic pathway which controls the induction, migration and differentiation of NCCs needs to be studied more extensively.

As mentioned above, abnormal development of NCCs leads to the generation of various human diseases, mostly described as neurocristopathies in humans (Liu and Xiao, 2011). Bolande first introduced the term neurocristopathy in 1974 to denote disorders that involve defective derivatives of NC.1–3; he divided the NC disorders into simple and complex. A simple neurocristopathy is exemplified by aganglionic megacolon (Hirschsprung's disease), in which segments of intestine lack submucosal and myenteric plexi of

parasympathetic ganglion cells. Another example could be familial dysautonomia (Riley-Day syndrome), a genetic disorder manifested by defective dorsal root and autonomic ganglia. Neurofibromatosis and neurocutaneous melanosis were used by Bolande as examples of complex neurocristopathies, but he did not extend the concept to encompass all neurocutaneous syndromes. Bolande also noted that the complex diseases tended to follow mendelian inheritance and that simple neurocristopathies were usually sporadic (Redies et al., 2012).

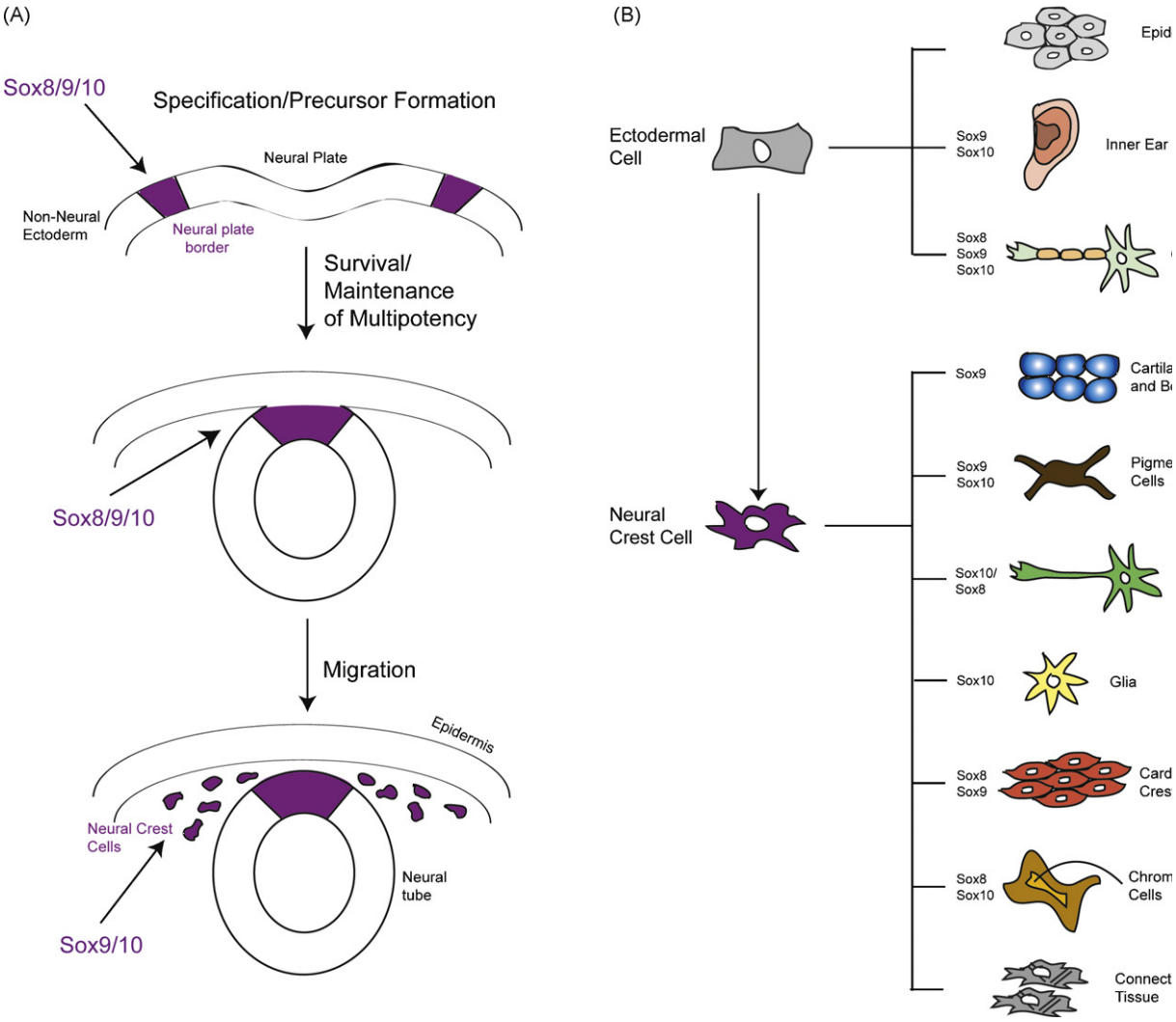


Figure 16. Schematic overview of NCC development. A) NC precursors with stem cell properties form at the border of the neural plate and non-neural ectoderm. Following neural tube closure, NCCs undergo an epithelial-mesenchymal transition (EMT) and migrate throughout the embryo to points of differentiation. SoxE proteins are required for the specification, precursor formation, survival and migration of NCCs throughout this process. B) NCCs give rise to a diverse set of derivatives. SoxE proteins are required to direct the differentiation of multiple NC derivatives including glia, melanocytes and cartilage, and are also required for the development of non-NC derivatives.

In recent years, and particularly with the fountainof new molecular genetic data that continues to emerge, the importance of NC as an inducer not only of peripheral neural structures such as ganglia but also of many tissues in craniofacial development and other peripheral mesodermal structures is becoming more and more evident. The category of what have traditionally been regarded as “neurocutaneous syndromes” can be attributed in a large part to abnormal NC migration and differentiation as well, thus expanding Bolande’s most recent classification of neurocristopathies to an entire category of abnormal neural tube induction of non-neural peripheral structures of the body that represent NC derivatives. Waardenburg syndrome is also included as a neurocutaneous syndrome for the same reason that unifies the other diseases in this category (Sarnat and Flores-Sarnat, 2005).

Sox2 has been implicated in defects observed in CHARGE syndrome. CHARGE syndrome is characterized by coloboma of the iris or retina, heart defects, atresia of the choanae, retardation of growth and/or development, genital, and ear abnormalities (Aramaki et al., 2007; Aramaki et al., 2006a; Aramaki et al., 2006b). Seventy-one percent of children clinically diagnosed with CHARGE have genetic mutations in the *Chd7* gene (Vallaster et al., 2012). Recently, Chd7 has been identified as a *Sox2* transcriptional cofactor (Engelen et al., 2011; Puc and Rosenfeld, 2011). More studies point to a possible role for *Sox2* in cardiac cell fate during embryogenesis as *Sox2* participates in a genetic *Oct4-Sox2* loop in ES cells to turn on *Sox17*, a gene essential for the specification of cardiac mesoderm (Stefanovic et al., 2009). Finally, heart development also controlled by regulation of canonical *WNT/β-catenin* signaling activity through feedback regulatory loops involving the ligands, agonists and antagonists, the availability of intracellular pools of active β-catenin and the cross-regulation of the WNT activity by β-catenin (Tanaka et al., 2011). The genetic connection of *WNT/β-catenin* and *Sox2* genes has been also reported to play a crucial role for the regulation of cell fate in various tissues during embryogenesis (Agathocleous et al., 2009; Hashimoto et al., 2012). It remains for us to identify how these pathways could merge to specify heart looping in early mouse embryos using *Sox2* epiblast inverted mouse embryos

Here, we show that ablation in the epiblast results into *Sox2* severe craniofacial abnormalities, early cardiac development and cranial facial defects. We propose that *Sox2* acts as a rheostat of EMT transitions during heart and NCC development, crucial cell fate processes indispensible for normal development and homeostasis (Han et al., 2012).

.

5. Epilogue

It has been shown that *Sox2* null embryos die around implantation. So far, various conditional alleles have been generated and have been used for the study of *Sox2* in development and homeostasis. Still deletion of the gene sequences, the lack of reporter cassette and the lack of information on the effect of these mutations on genetic and regulatory elements of the locus such as *Sox2ot* lncRNA, *miR1897* and *Sox2* CpG island in these conditional alleles prompted us to generate a new conditional allele called *Sox2*^{COIN}. To overcome the above limitations, we designed a novel cassette, which has been inserted in the antisense strand of *Sox2*, splitting the gene in two exons, in a way that the CpG island has not been removed as in previous conditional alleles. At the same time, we introduced a reporter cassette that was expressing eGFP in frame with Sox2 amino acid sequences only after inversion by Cre recombinase. COIN cassette expresses eGFP constantly once *Sox2* ablation is triggered. *A*part from the independent regulatory elements that control the its expression, *Sox2* could potentially be transcribed also as part of a *lncRNA*, called *Sox2ot,* which contains *miR1897*, both of whose function remains elusive. *Sox2ot* and *mir1897* remain unaffected in *Sox2*^{COIN} and *Sox2*^{INV} embryos. As a result, we can be assured that the phenotypes generated by the use of *Sox2*^{COIN} conditional allele derive exclusively by the ablation of *Sox2* function. Analysis of Sox2 loss of function in the epiblast revealed a role for *Sox2* in the epithelial to mesenchymal transition (EMT) during NCC development in mice, resulting in craniofacial abnormalities described also in human diseases, called neurocristopathies. Surprisingly, our results revealed a novel role for *Sox2* during heart looping in mouse embryos, even though *Sox2* is not expressed in the heart. Further analysis will reveal whether this is due to the role of *Sox2* in the endoderm influencing heart development. These findings open novel avenues for the role of *Sox2* in a number of diseases during embryogenesis and in adult life both in mice and humans.

Appendix I

CURRICULUM VITAE

Nikolaos Mandalos, BSc, MSc ,PhD

Stem Cell Biology Laboratory
Institute of Molecular Biology and Genetics
BSRC "Alexander Fleming"
34 Fleming Street,16272 Vari Athens, Greece
email: mandalos@fleming.gr,
nikosmandalos@gmail.com
Tel.: +30-2109656310
Date of birth: April 22, 1980
Place of birth: Athens, Greece
Citizenship: Greek

Education

1992-1998 **Secondary Education**

37rd High school of Athens, Athens, Greece

1999-2000 A-Levels (Cambridge University) C.A.S.E Education College, Athens, Greece

2000-2005 **Degree of Bachelor of Science** with Second Honours (Division Two) in Developmental Biology (Edinburgh University)

2003-2004 **Research assistant** (Gap year Placement) Working in developmental biology laboratory, studying the molecular mechanisms that regulate embryonic development of the forebrain

2005-2006 **MSc in Neuroscience**
Institute of Psychiatry, King's College London University

2006-2012 **Ph.D.** BSRC "Alexander Fleming", Stem Cell Biology Laboratory and National and University of Crete Medical School, Crete, Greece

2010 **International Course on Laboratory Animal Science,** BSRC "Al. Fleming", 22 Sept-3 Oct 2008, Vari Athens, Greece.

Foreign Languages

- **Greek** – Native Language
- **English** - Proficiency in English Certificate – University of Cambridge
- **French** - Diplôme approfondi de langue française

Research interests

Stem Cell and Developmental Biology, Neurobiology of neurodegenerative diseases. Neural crest development, craniofacial abnormalities

Research Experience

2003 **Research Trainee**. Centres for Integrative Physiology and Neuroscience Research, School of Biomedical Sciences, University of Edinburgh

2004 **Diploma thesis** on "Function and regulation of Gli3 in the developing mouse midbrain". School of Biomedical Sciences, University of Edinburgh (Supervisor: Dr. Mason JO.)

2005-2006 **Master Thesis, Neuroscience Master's Program**; "TCF7L2 gene polymorphisms susceptibility in Alzheimer's disease in patients with diabetes type II", Vari, Greece. **Supervisor:** Dr. Powell J.

2006-2012 **PhD** ; "Generation of conditional by inversion *Sox2* allele", Stem Cell Biology Laboratory, BSRC "Al. Fleming", Vari, Greece. Supervisor: Dr. E.Remboutsika.

Conference Participation

2010 24th International Mammalian Genome Conference (IMBG), October 17-21, 2010. Aldemar Knossos Royal Village, Heraklion Crete, Greece. Mandalos, N., Saridaki, M., Harper, J.L., Kotsosni, A., Economides, A. N., and Remboutsika, E. Epiblast-ablation of Sox2 expression leads to neurocristopathies in mouse embryos. (Abstract 142).

2012 63rd Congress of the Hellenic Society of Biochemistry and Molecular Biology. A novel role for Sox2 as a rheostat of epithelial to mesenchymal transition Nikolaos Mandalos, Maria Poulou, Theodoros Karnavas, Muriel Rhinn, Aris N. Economides, Pascal Dollé and Eumorphia Remboutsika

Publications

1. Mandalos, N., Saridaki, M., Harper, J.L., Kotsoni, A., Yang, P., Economides, A.N., and Remboutsika, E. (2012). Application of a novel strategy of engineering conditional alleles to a single exon gene, sox2. PLoS One *7*, e45768.

2. Poulou, M. Mandalos, N., Karnavas, T. and Remboutsika, E. (2012) From Stem Cells to Organs (Research Topic). Frontiers in Physiology (Craniofacial Biology).

3. Karnavas, T., Mandalos, N. and Remboutsika, E. (2012). SoxB, cell cycle and neurogenesis. (Invited Review). Neurogenesis Research: New Developments. Nova Publishers, in press.

4. Nikolaos Mandalos, Maria Poulou, Theodoros Karnavas, Muriel Rhinn, Aris N. Economides, Pascal Dollé and Eumorphia Remboutsika (2012). A novel role of Sox2 as a rheostat of EMT transition during mouse development, in preparation.

5. Saridaki, M., Mandalos, N., Harper, J. L., Karampelas, I., Economides, A. and Remboutsika, E. (2012). Radial glia ablation of *Sox2* expression leads to neurodegeneration, in preparation.

Appendix II: Publications

OPEN ACCESS Freely available online

PLOS ONE

Application of a Novel Strategy of Engineering Conditional Alleles to a Single Exon Gene, *Sox2*

Nikolaos Mandalos[1], Marannia Saridaki[1¤a], Jessica Lea Harper[1,2], Anastasia Kotsoni[1¤b], Peter Yang[2], Aris N. Economides[2]*, Eumorphia Remboutsika[1]*

1 Stem Cell Biology Laboratory, Institute of Molecular Biology and Genetics, Biomedical Sciences Research Centre "Alexander Fleming", Vari, Greece, **2** Regeneron Pharmaceuticals Inc., Tarrytown, New York, United States of America

Abstract

Background: The Conditional by Inversion (COIN) method for engineering conditional alleles relies on an invertible optimized gene trap-like element, the COIN module, for imparting conditionality. The COIN module contains an optimized 3′ splice site-polyadenylation signal pair, but is inserted antisense to the target gene and therefore does not alter transcription, until it is inverted by Cre recombinase. In order to make COIN applicable to all protein-coding genes, the COIN module has been engineered within an artificial intron, enabling insertion into an exon.

Methodology/Principal Findings: Therefore, theoretically, the COIN method should be applicable to single exon genes, and to test this idea we engineered a COIN allele of *Sox2*. This single exon gene presents additional design challenges, in that its proximal promoter and coding region are entirely contained within a CpG island, and are also spanned by an overlapping transcript, *Sox2Ot*, which contains *mmu-miR1897*. Here, we show that despite disruption of the CpG island by the COIN module intron, the COIN allele of *Sox2* ($Sox2^{COIN}$) is phenotypically wild type, and also does not interfere with expression of *Sox2Ot* and *miR1897*. Furthermore, the inverted COIN allele of *Sox2*, $Sox2^{INV}$ is functionally null, as homozygotes recapitulate the phenotype of $Sox2^{\beta geo/\beta geo}$ mice, a well-characterized *Sox2* null. Lastly, the benefit of the eGFP marker embedded in the COIN allele is demonstrated as it mirrors the expression pattern of *Sox2*.

Conclusions/Significance: Our results demonstrate the applicability of the COIN technology as a method of choice for targeting single exon genes.

Citation: Mandalos N, Saridaki M, Harper JL, Kotsoni A, Yang P, et al. (2012) Application of a Novel Strategy of Engineering Conditional Alleles to a Single Exon Gene, *Sox2*. PLoS ONE 7(9): e45768. doi:10.1371/journal.pone.0045768

Editor: Andreas Androutsellis-Theotokis, Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden, Germany

Received July 1, 2012; **Accepted** August 20, 2012; **Published** September 27, 2012

Funding: This work was financially supported by BILATERAL COOPERATION GRANT (NON-EU82) from the General Secretariat of Research and Technology and it was performed under the auspices of a collaboration agreement with REGENERON Pharmaceuticals Inc., New York, USA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Three authors (Jessica Lea Harper, Peter Yang and Aris N. Economides) have the affiliation to the commercial funders of this research REGENERON Pharmaceuticals Inc. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: remboutsika@gmail.com (ER); aris.economides@regeneron.com (ANE)

¤a Current address: Biomedical Research Foundation of the Academy of Athens, Athens, Greece
¤b Current address: Institute of Immunology, Biomedical Sciences Research Centre "Alexander Fleming", Vari, Greece

Introduction

Intronless (single exon) genes are thought to be evolutionary innovations, whose formation via reverse transcription–mediated mechanisms represents an important route of evolution for tissue-specific functions in animal cells [1,2]. Approximately 12% of the human and 13.4% of the mouse protein-coding genes are intronless [3–6], and include genes that encode for regulatory proteins and components of signal transduction pathways [7], histones [8–11], G Protein-coupled Receptors [3] and transcription factors such as the *Sox* (SRY-related HMG box) family [12–14].

Sox2 is a well-characterized and important example of a single exon gene. Sox2 pairs with tissue-specific partners [15] to impart and maintain pluripotency [16] and multipotency [14,17] during development and homeostasis [18]. *Sox2* null embryos fail to form the epiblast and die at E5.5 [19]. However, even reduction in *Sox2* levels to 25–30% relative to the wild type leads to pathological phenotypes in mice. These include neurodegeneration in the cortical region and hippocampus [20], hypoplasia of optic nerves and chiasmata and variable microphthalmia [21], failure of nasal placode induction [22], failure of taste buds to mature [23], malformation of the epithelium lining the conducting airways in the lung [24], enlargement of the lateral ventricles at E14.5 [25], and immature differentiation of cochlea hair follicles [26].

From a gene structure standpoint, *Sox2* presents a complex locus rich in genetic elements, including an overlapping transcript [27], a putative microRNA [28], and a CpG island [29–31] [32]. The combination of a well-conserved compact locus with overlapping transcripts and regulatory elements [33–39], together with the apparent need to maintain proper levels of Sox2 for organogenesis and homeostasis, underscore the difficulties associated with designing conditional alleles for *Sox2*. We hypothesized that a recently developed method for generating conditional alleles – Conditional by Inversion (COIN) – might present a better choice over simple floxing of

Sox2, and generated the corresponding conditional-null allele, $Sox2^{COIN}$. We show that this method is successful in that the $Sox2^{COIN}$ allele starts as wild type, and it is converted into a null by the action of Cre, at which point, the expression of *Sox2* is replaced by that of a marker, *eGFP*. This work indicates that the COIN method can be applied to single exon genes and provide a new design modality that can be adopted for other genes like *Sox2*.

Results

Generation of the $Sox2^{COIN}$ Allele

Sox2 (ENSMUSG00000074637) is a single exon gene encoding a 319 amino acid protein. The *Sox2* locus contains several features that render it complex from the standpoint of engineering modified alleles (**Fig. 1A**). To begin with, *Sox2's* proximal promoter and coding region comprise a CpG island [15]. Furthermore, the *Sox2* exon is contained with the intron of a long non-coding RNA (ncRNA), termed *Sox2* overlapping transcript (*Sox2ot*) or "non-protein coding RNA 43", which also contains mmu-miR1897 (miR1897) [40]. *Sox2ot* is transcribed from the same strand as *Sox2* but its molecular and biological functions remains elusive. *Sox2ot* transcript is expressed in mouse embryonic stem cells and in other tissues, including the nervous system, where *Sox2* is also highly expressed [41], while an isoform of *Sox2ot*, *Sox2dot*, located around 500 base pairs upstream of *Sox2*, was detected exclusively in adult mouse brain [27]. Because of this complexity, *Sox2* is a challenging locus to apply conditional mutagenesis, and therefore presents a stringent test for new methods of allele design, such as COIN.

The COIN method relies on an optimized gene trap-like element, referred to as the COIN module [42]. The COIN module is comprised of a *3' splice region-reporter cDNA-polyadenylation region* optimized to function as an efficient transcriptional block, and it is flanked by *Lox71* and *Lox66* sites are in a mirror image configuration to enable Cre-mediated inversion [43]. In order to generate conditional-null alleles, the COIN module is placed in a position antisense to the target gene, either within a native intron, or an exon. The latter is made possible by embedding the COIN module within an artificial intron – the COIN module intron – and using that intron to split the target exon into two operational halves [42].

To generate $Sox2^{COIN}$, the COIN module intron was inserted directly into the single exon of *Sox2* (Figure 1B), after the 30th nucleotide of *Sox2*'s open reading frame, splitting the single *Sox2* exon into two exons. The COIN module lies inertly within the antisense strand of *Sox2*, stealth to transcription. Upon Cre-mediated inversion into the sense orientation, the $Sox2^{COIN}$ allele is converted into a null allele, $Sox2^{INV}$. This is accomplished by the COIN module abrogating transcription of full length Sox2, effectively replacing it with expression of the COIN module's eGFP reporter. The expression of *eGFP* in place of *Sox2* is controlled by *Sox2*'s promoter and regulatory elements, and enables visual identification of the inversion event at the tissue and cellular level. The functionality of the allele was assessed *in vivo* in a series of experiments that assessed whether $Sox2^{COIN}$ is a truly wild type allele, and whether $Sox2^{INV/INV}$ recapitulate the null phenotype, while providing a useful marker that faithfully reproduces the expression profile of *Sox2*.

$Sox2^{COIN}$ is Wild Type in Homozygosis

Offspring of $Sox2^{COIN/+}$ intercrosses were born in Mendelian ratios and no lethality was observed in embryos, newborn pups and adults (Table 1). Homozygote mice fed normally, showed no abnormal behavior and they had normal weight in adulthood (data not shown). Macroscopic analysis of E14.5 $Sox2^{+/+}$, $Sox2^{COIN/+}$, and $Sox2^{COIN/COIN}$ mice showed that the COIN module does not affect normal embryonic mouse development (Figure 2C). These phenotypic observations are further corroborated by the result that $Sox2^{COIN/\beta geo2}$ E6.5 embryos were morphologically indistinguishable from $Sox2^{+/+}$ or $Sox2^{\beta geo2/+}$ embryos derived from a $Sox2^{\beta geo2/+}$ with $Sox2^{COIN/COIN}$ cross (Figure 3B, F), where $Sox2^{\beta geo2}$ is a null allele of *Sox2* [19] (see below). Furthermore, examination of *Sox2* mRNA (Figure 2F) and Sox2 protein (Figure 2H) expression levels show no apparent difference between the three genotypic classes, $Sox2^{+/+}$, $Sox2^{COIN/+}$, and $Sox2^{COIN/COIN}$, demonstrating that the COIN module has no effect on the expression of *Sox2*. Thus, by all of these criteria – heritability, phenotype, expression of mRNA and protein – $Sox2^{COIN}$ behaves as a wild type allele.

The Expression of *Sox2ot and miR1897* are Unaffected in $Sox2^{COIN/COIN}$ Mice

To assess whether the COIN module affects *Sox2ot* RNA expression, we isolated RNA from E14.5 mouse embryos from different intercrosses and quantified *Sox2ot* RNA levels by Taqman Real-Time PCR analysis (Figure 2G). No significant difference was detected in the expression of *Sox2ot* in $Sox2^{+/+}$, $Sox2^{COIN/+}$, and $Sox2^{COIN/COIN}$, demonstrating that the COIN module has no effect on the expression of *Sox2ot*, at least prior to inversion. Identical observations where made for *miR1897*, which is embedded in *Sox2ot* (Figure 2H).

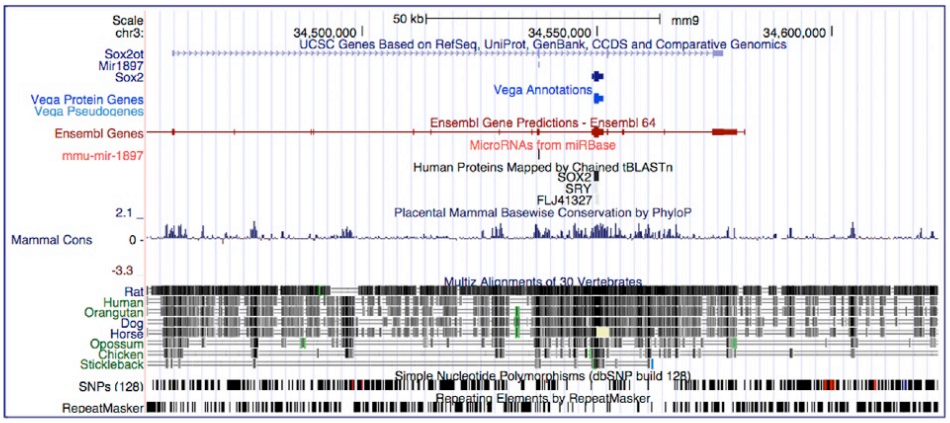
$Sox2^{COIN}$ is Efficiently Inverted by Cre to Generate $Sox2^{INV}$

To assess whether we could trigger COIN inversion upon Cre expression, $Sox2^{COIN/+}$ adult mice were intercrossed with *Sox2*Tg(-Sox2:CRE) transgenic mice to generate $Sox2^{INV/+}$ embryos and adult mice (Figure 2D, E). In contrast to the partial infertility phenotype that has been observed with $Sox2^{\beta geo2/+}$ mice [19], $Sox2^{INV/+}$ adult mice exhibited no obvious phenotypes and transmitted the inverted allele in Mendelian ratios (Table 2), irrespective of whether the $Sox2^{INV}$ allele is transmitted via the male or female germline (data not shown). More importantly, E14.5 $Sox2^{INV/+}$ embryos (Figure 2E) displayed vivid eGFP expression in the cerebral cortex, retina, olfactory bulb, hair follicles, olfactory epithelium and spinal cord (Figure 2E), mirroring what has been observed with X-gal stained E14.5 $Sox2^{\beta geo2/+}$embryos [44]. In addition, the presence of eGFP protein can be detected by Western blotting in protein extracts derived from $Sox2^{INV/+}$ embryos, and appears to be accompanied by a reduction in the levels of Sox2 protein, similar to what has been observed in the $Sox2^{\beta geo2/+}$ embryos (Figure 2I).

$Sox2^{INV}$ is a Null Allele of *Sox2*

Mice carrying a loss of function mutation in the *Sox2* locus have been generated by the insertion of a βgeo cassette into the *Sox2* locus ($Sox2^{\beta geo}$ [19], and $Sox2^{\beta geo2}$ [44]). Upon homozygosis, both alleles yield *Sox*2-null embryos that fail to form the epiblast and die around implantation. To test whether $Sox2^{INV/INV}$ phenocopy $Sox2^{\beta geo2/\beta geo2}$, we performed $Sox2^{INV/+}$ intercrosses (Table 3). No $Sox2^{INV/INV}$ offsprings were born. More specifically, $Sox2^{INV/INV}$ mutants failed to survive shortly after implantation (Figure 3I), phenocopying $Sox2^{\beta geo/\beta geo}$ and $Sox2^{\beta geo2/\beta geo2}$ embryos. Only $Sox2^{+/+}$ and $Sox2^{INV/+}$ embryos reach the embryonic stage of E6.5 (Figure 3A, G). Histological examination of whole decidual swellings harvested at 6.5 dpc revealed that 25% of deciduas carried abnormal implants, which had no egg cylinder structure and lacked the epithelial cells typical of epiblast (Figure 3I).

A. *Sox2* genomic locus



B. Targeted allele

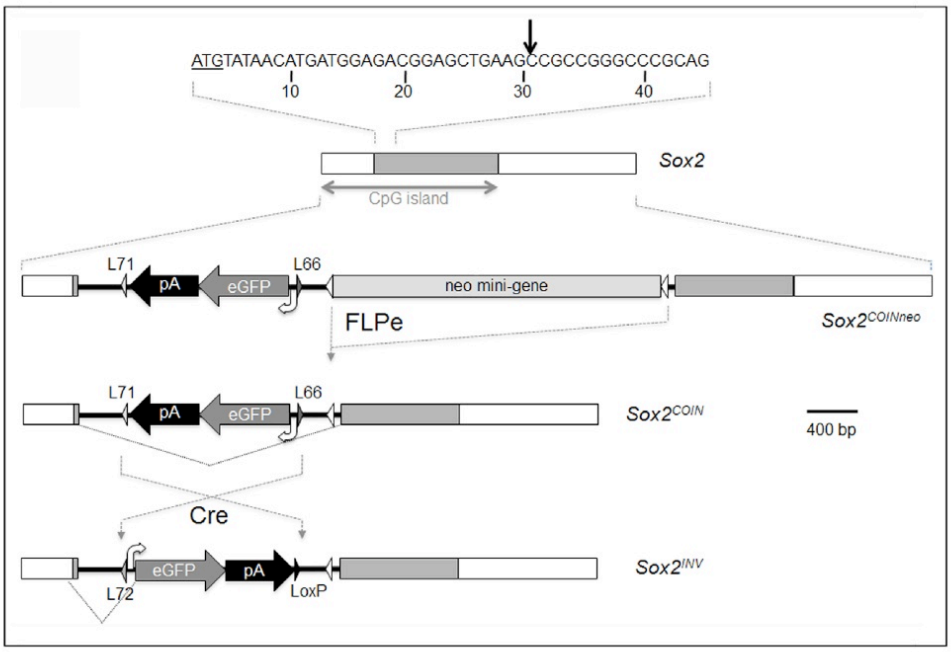


Figure 1. Targeting strategy generating a COIN allele of *Sox2*. (**A**) Schematic representation of the mouse *Sox2* locus indicating the relative location of the exon on chromosome 3, as well as that of *mir1897*, the non-coding RNA *Sox2ot*, and CpG islands in the genomic region. The degree of conservation of the locus sequence between mammalian species (ECRs) is indicated. Adapted from http://genome.ucsc.edu. (**B**) Schematic representation of the $Sox2^{COIN}$ allele. The COIN module intron is inserted after the 30th nucleotide of *Sox2*'s coding region (i.e. coordinate 34549367 on Chromosome 3) splitting the single exon of Sox2 into two exons and also dividing the CpG island. The COIN module is comprised of an optimized gene trap-like element composed of the 3′ splice region of the rabbit beta globin gene (HBB_RABIT), followed by *eGFP* (lacking an initiating ATG) and the polyadenylation region from HBB_RABIT, all placed in the antisense strand. The COIN module has been flanked with *Lox71* and *Lox66* sites in a

mirror image orientation, thereby enabling inversion by Cre. For BHR and targeting, a FRT-flanked *neo* cassette has been incorporated into the COIN intron. After targeting, *neo* is removed to give rise to the $Sox2^{COIN}$ allele. The COIN module is antisense to *Sox2*, and hence it predicted not to interfere with expression of *Sox2*. However, after inversion of the COIN module to the sense strand, transcription terminates around the polyadenylation region of the COIN module, and as a result expression of *Sox2* is replaced by *eGFP*.
doi:10.1371/journal.pone.0045768.g001

Instead, many trophoblast giant cells could be identified (Figure 3C, H, I). The same phenotype is observed in $Sox2^{INV/\beta geo2}$ embryos (Figure 3H). These results demonstrate the failure of $Sox2^{INV/INV}$ embryos to develop an epiblast similarly to the $Sox2^{\beta geo/\beta geo}$ and $Sox2^{\beta geo2/\beta geo2}$ mutants. Thus, the inverted COIN cassette generates a true *Sox2* null phenotype.

Discussion

We show here the application of COIN technology to generate a conditional-null allele for a single exon gene, *Sox2*. The COIN method was invented at least in part to overcome the challenges and limitations of traditional site-specific recombinase-based strategies such as Cre/Lox for designing conditional alleles [45]. These include the placement of Lox sites as well as the distance between them [46], defining critical exons (i.e. the exons of the gene that need to be deleted by Cre in order to bring about the desired allelic state) [47,48], and the lack of unified strategy for including a reporter that can mark those cells that harbor the post-recombinase allele. To date, COIN has been successfully applied in generating conditional alleles for more than twenty-five protein-coding genes [42], but its applicability to single exon genes has not been tested.

Single exon genes present a design challenge for engineering conditional alleles by traditional, e.g. simple floxing, methods. First, the Lox sites should be placed in a position that does not affect the expression of the target locus, a design decision that can be complicated by the lack of specific knowledge of the exact position of promoters and regulatory elements. An additional design challenge is presented if a reporter that marks the conversion from 'wild-type' to null is desired. The COIN method addresses both of these challenges irrespective of gene structure by avoiding the placement of Lox and FRT sites, reporters, and other functional elements within regions upstream of the target gene's coding sequence [42]. Instead, COIN employs an 'exon-splitting' artificial intron to place an optimized module – the COIN module – within a coding exon, yet in the antisense strand. Perusal of the regulatory elements mapping within single coding exon of human Sox2 suggests that the COIN intron was inserted at a position that does not result in disruption of any such elements (data not shown) As a result, the COIN module is stealth to transcription, and does not alter the expression of the modified gene. Although it is possible that introduction of the COIN module intron alters the kinetics of transcription [49–51], we have not examined this possibility at the single cell level; at the population level and at steady state, the level of *Sox2* mRNA as expressed from the $Sox2^{COIN}$ locus does not appear different to that of wild type. Lastly, because in COINs the Lox and FRT sites are placed within the artificial intron, they do not disrupt of promoters or regulatory sequences, and are also not incorporated into mRNA.

Table 1. Analysis of progeny from $Sox2^{COIN/+}$ intercrosses#.

Genotypic distribution in live progeny*					
	Live	**Dead**	$Sox2^{+/+}$	$Sox2^{COIN/+}$	$Sox2^{COIN/COIN}$
Age	No	No	No (%)	No (%)	No (%)
P21	95	0	27 (25%)	40 (50%)	28 (25%)

Genotyping of $Sox2^{COIN/+}$ heterozygous intercross progeny.
#Data collected from mice in C57BL6 background.
*Genotypes were assessed by PCR of genomic tail DNA.
doi:10.1371/journal.pone.0045768.t001

The particular choice of *Sox2* to test the COIN method's applicability to single exon genes presented additional challenges in that the majority of *Sox2*'s single exon is contained within a CpG island, and there is also an overlapping non-coding transcript *Sox2ot*. Due to design constraints – specifically the need to place the COIN module as near the initiating ATG as possible – the COIN module intron was inserted into *Sox2*'s exon in manner that disrupts the CpG island. However, this had no apparent effect on Sox2 expression and had not apparent phenotypic consequences in the mouse. *Sox2ot* levels also remained unaltered, indicating that at least in the antisense position the presence of the COIN module has no effect on the expression of *Sox2ot*. This was evident by the normal phenotype of $Sox2^{COIN/\beta geo2}$ embryos, in which only the COIN allele can generate wild type mRNA. This genotype should sensitize the embryo to any reduction in Sox2 levels, and thus provides a stringent comparison between the wild type allele and the COIN allele prior to inversion.

Equally important is the fact that post-inversion of the COIN module, the resulting allele, $Sox2^{INV}$ phenocopies the previously generated null alleles upon homozygosis. In addition, the reporter embedded in the COIN module, is expressed in a manner representative of Sox2 expression, thereby generating a tool to visualize Sox2 expression and to follow the conversion of the COIN allele into a null by Cre.

In addition to the conditional-null allele presented here, four other conditional-null alleles of *Sox2* have been published [21,52–54]. All four rely on floxing of the single exon of *Sox2*, though the placement of the LoxP sites and selection cassettes (and their retention) varies among alleles. One of the main differences between these alleles and $Sox2^{COIN}$, is that they do not incorporate a reporter that is activated after Cre acts on the allele. There is however a paucity of published data such as expression analysis of *Sox2*, *Sox2ot*, and *miR1897* to allow further comparisons between $Sox2^{COIN}$ and the previously described conditional-null alleles. Given the increasing evidence for roles that Sox2 plays in a wide range of pathological and patho-physiological conditions, assays for the normal regulation of *Sox2* expression need to be conducted in a variety of cell types. Overall, our results highlight the importance of the Conditional by Inversion technology as a method of choice in targeting intronless (single exon) genes, especially when complexity of the locus and desire for inclusion of a reporter are taken into consideration.

Methods

Gene Targeting

Targeted $Sox2^{COINneo/+}$ ES cells were generated using VelociGene™ methodology, essentially as described [55]. Briefly, the BAC-based targeting vector was assembled on bacterial artificial chromosome (BAC) RP23_406a6 that encompasses the single protein-coding exon of *Sox2* flanked by approximately 95 and 71 kb upstream and downstream respectively. The COIN module

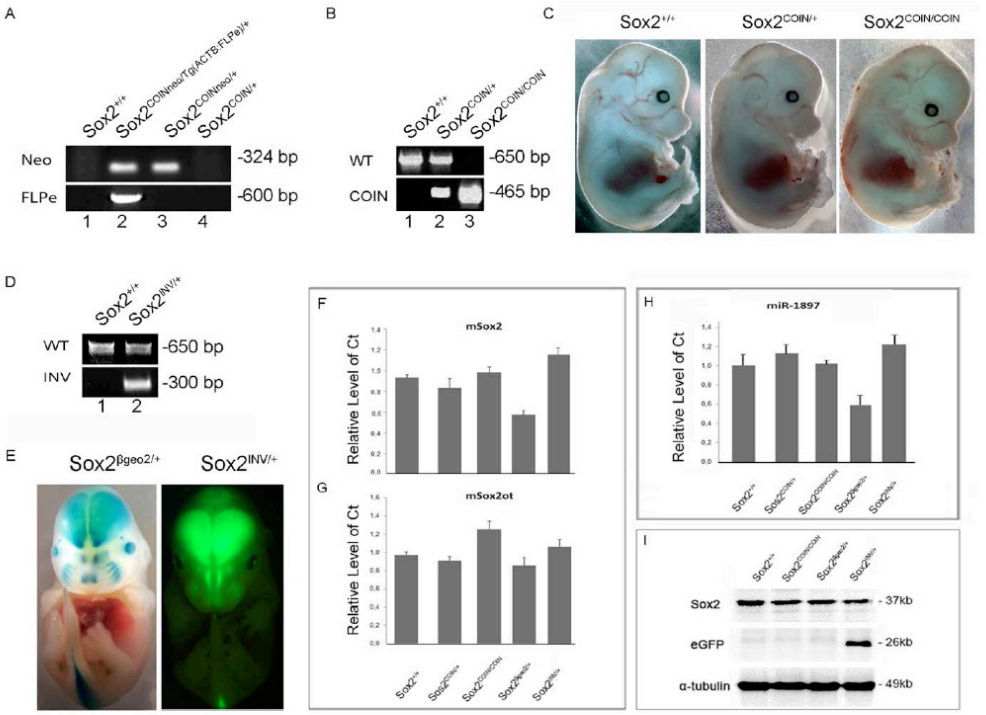


Figure 2. *Sox2*^{COIN} is a functional conditional allele. (**A**) Efficient removal of the *neo* cassette by Flpe recombinase to generate a *Sox2*$^{COIN/+}$ allele. *Sox2*$^{+/+}$ (lane 1), *Sox2*$^{COIN/+}$ (lane 2), *Sox2*$^{COIN/COIN}$ (lane 3), *Sox2*$^{INV/+}$ (lane 4), *Sox2*$^{βgeo2/+}$ (lane 5) E14.5 mice. PCR genotyping of (1) *Sox2*$^{+/+}$, (2) *Sox2*$^{COIN/+}$ *Tg(ACTB:FLPe)*, (3) *Sox2*$^{COINneo/+}$ (4) *Sox2*$^{COIN/+}$. (**B**) PCR genotyping of *Sox2*$^{+/+}$ (lane 1), *Sox2*$^{COIN/+}$ (lane 2), *Sox2*$^{COIN/COIN}$ (lane 3) (genomic DNA from tail biopsies). (**C**) E14.5 *Sox2*$^{COIN/COIN}$ embryos are morphologically indistinguishable from *Sox2*$^{+/+}$ and *Sox2*$^{COIN/+}$. (**D**) Efficient inversion of the COIN module to generate *Sox2*$^{INV/+}$ mice. PCR-based genotyping of (1) *Sox2*$^{+/+}$ and (2) *Sox2*$^{INV/+}$ (genomic DNA from tail biopsies). (**F, G**) *Sox2* and *Sox2ot* qPCR analysis in *Sox2*$^{+/+}$, *Sox2*$^{COIN/+}$, *Sox2*$^{COIN/COIN}$, *Sox2*$^{βgeo2/+}$ and *Sox2*$^{INV/+}$ E14.5 embryos. Cyclophilin was used as a control label. Data are presented as the mean±SEM (n = 3–6) for each genotype. The COIN module does not affect expression of *Sox2* prior to inversion. Additionallly, the COIN module does not appear to affect the expression level of *Sox2ot* either in *Sox2*$^{COIN/COIN}$ or in *Sox2*$^{INV/+}$ embryos. (**H**) *MiR1897* qPCR analysis in *Sox2*$^{+/+}$, *Sox2*$^{COIN/+}$, *Sox2*$^{COIN/COIN}$, *Sox2*$^{βgeo2/+}$ and *Sox2*$^{INV/+}$ E14.5 embryos. (**I**) Sox2 protein analysis. Western blot showing Sox2 and eGFP protein detected with specific antibodies in whole protein extracts from *Sox2*$^{+/+}$, *Sox2*$^{COIN/+}$, *Sox2*$^{COIN/COIN}$, *Sox2*$^{βgeo2/+}$, and *Sox2*$^{INV/+}$ E14.5 embryos. The COIN module does not affect Sox2 protein levels prior to inversion.

doi:10.1371/journal.pone.0045768.g002

intron was introduced by bacterial homologous recombination [56] after the 30th nucleotide of *Sox2*'s coding region (i.e. coordinate 34549367 on Chromosome 3, as annotated on Ensembl release 67), splitting the single coding exon of *Sox2* into two exons of 441 bp and 2016 bp respectively (Figure 1B).

Experimental Animals

Sox2$^{COINneo/+}$ mice were bred with *Tg(ACTB:FLPe)* mice (Flp-deleter mice) to excise the Neo cassette and generate *Sox2*$^{COIN/+}$ *Tg(ACTB:FLPe)* mice. These were bred with C57BL6 mice to bring the *Sox2*^{COIN} allele into the germline. *Sox2*$^{COIN/+}$ mice were in turn bred with Tg(*Sox2*:CRE) mice to generate *Sox2*$^{COIN\text{-}INV/+}$ mice. All animals were handled in strict accordance with good animal practice as defined by the Animals Act 160/03.05.1991 applicable in Greece, revised according to the 86/609/EEC/ 24.11.1986 EU directive regarding the proper care and use of laboratory animals and in accordance to the Hellenic License for Animal Experimentation at the BSRC'' Alexander Fleming'' (Prot. No. 767/28.02.07) issued after protocol approval by the Animal Research Committee of the BSRC ''Alexander Fleming'' (Prot. No. 2762/03.08.05).

Embryo Processing, Tissue Preparation and Histological Analysis

For staging of the embryos, midday of the vaginal plug was considered as embryonic day 0.5 (E0.5). E6.5 decidua and E14.5 embryos were collected and dissected in cold PBS. Tissues were fixed with 10% formalin for 24 hours at room temperature and then washed several times with 1% PBS, then placed in embedding cassettes. Paraffin sections (10 μm) were stained with Hematoxylin and Eosin (H&E) using standard procedures and mounted with xylene based mounting medium. E14.5 *Sox2*$^{βgeo2/+}$ LacZ staining was performed following standard protocol [19].

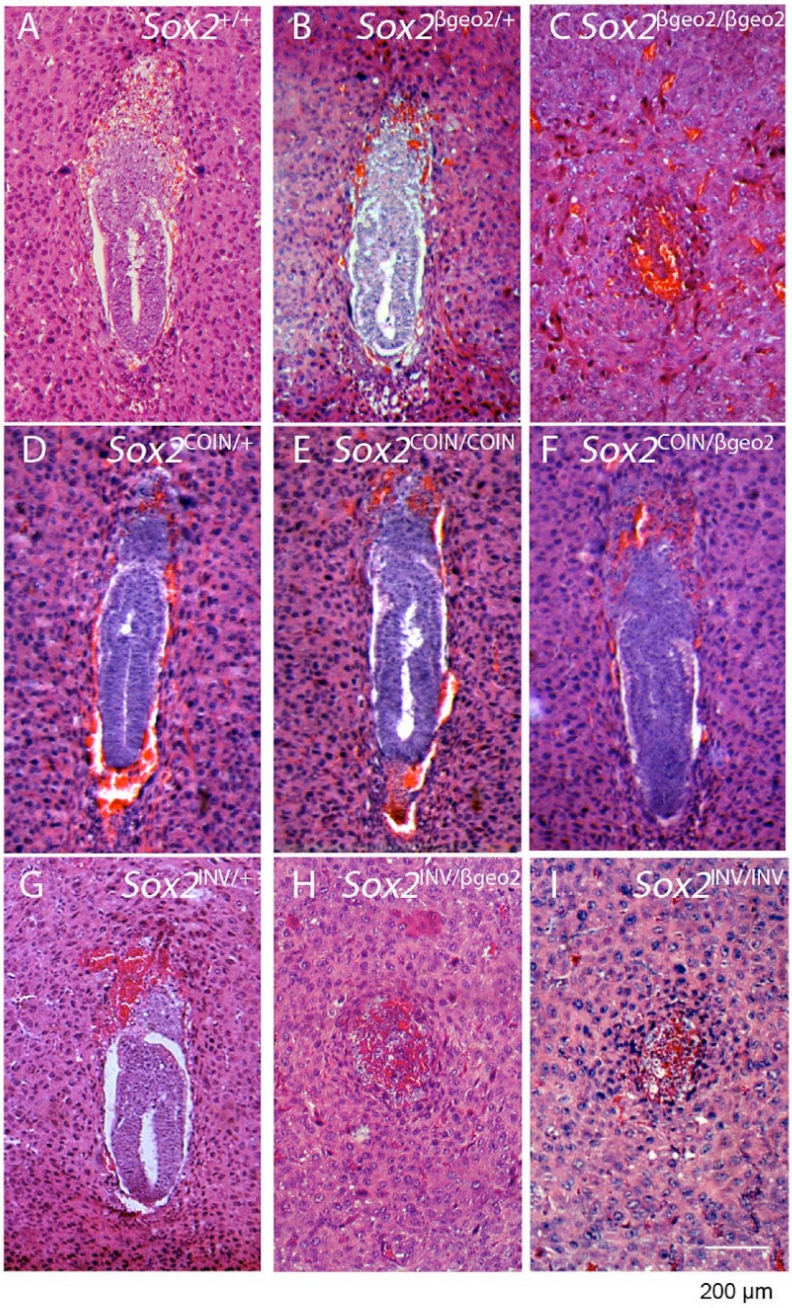

A Sox2+/+
B Sox2βgeo2/+
C Sox2βgeo2/βgeo2
D Sox2COIN/+
E Sox2COIN/COIN
F Sox2COIN/βgeo2
G Sox2INV/+
H Sox2INV/βgeo2
I Sox2INV/INV
200 μm

Figure 3. *Sox2*^{INV} is a null allele. Histological analysis of 6.5-dpc embryos from *Sox2*$^{\beta geo2/+}$, *Sox2*$^{COIN/+}$, *Sox2*$^{INV/+}$ intercrosses. *Sox2*$^{+/+}$ (**A**), *Sox2*$^{\beta geo2/+}$ (**B**), *Sox2*$^{COIN/+}$ (**D**), *Sox2*$^{COIN/COIN}$ (**E**), *Sox2*$^{COIN/\beta geo2}$ (**F**), and *Sox2*$^{INV/+}$ (**G**) are all phenotypically normal. In contrast, *Sox2*$^{INV/INV}$ (**I**) and *Sox2*$^{INV/\beta geo2}$ (**H**) embryos form disorganized extraembryonic tissues, but fail to form the epiblast, effectively phenocopying *Sox2* $^{\beta geo2/\beta geo2}$ (**C**), and displaying the phenotype previously described for *Sox2* homozygous-null embryos. Sections were stained with hematoxylin and eosin (H & E).
doi:10.1371/journal.pone.0045768.g003

Genotyping

Tail, yolk sack or embryonic tissues were isolated and processed according to previously described methodology. Deletion of the EM7p-neo-polyA FRT-flanked (FRTed) in the germline dual-purpose antibiotic/drug selection cassette by FLP recombinase was documented by genotyping PCR from adult mice tail. Neo gene was amplified with *Neo* Frw (5′-CTGAATGAACTGCAG-GACGA-3′), *Neo* Rvs (5′-ATACTTTCTCGGCAGGAGCA-3′) (172 bp); FLPe with *FLP* Frw (5′-GAGAAGAACGGCA-TAGTGCG-3′), *FLP* Rvs (5′-GACAAGCGTTAGTAGGCA-CAT-3′) (600 bp). Genomic DNA from E14.5 embryos was isolated from yolk sack. In E6.5 embryos genomic DNA was isolated by scraping carefully under the stereoscope off glass slides after staining with H&E and mounting. Detachment of cover slides was done by embedding mounted slides back in xylene. Detection of *Sox2*$^{COIN/+}$ mice that carry the assay for the COIN cassette was performed using the following set of primers: 5′ end primer combination (338 bp), INVdiaF2 (F2): 5′ CACTTTC-TACTCTGTTGAC 3′, INVdiaR2 (R2): 5′ CCTTACATGTTT-TACTAG 3′, 3′ end primer combination (470 bp), IN-Vdia(eGFP).F3 (F3): 5′CTGAAGCACTGCACGCCGTAG 3′, INVdia.R4 (R4): 5′ CTCAGAGTATTTTATCCTCATCTC 3′ (Figure 2a). PCR amplification of COIN cassette in mice and embryos E6.5 and E14.5 was also performed with the following primers: Sox2ATG407 Frw: 5′ATGTATAACATGATG-GAGA3', R2*R 5′TATGACTGGGAGTAGTCAGGAGAG-GAGGAA3' (465 bp). PCR amplification of *Sox2* in WT and heterozygous mice was performed with the following primers: Sox2ATG407 Frw: 5′ATGTATAACATGATGGAGA 3′, Sox21053 Rvs: 5′ CTGGTCATGGTGTTG 3′ (650 bp). *Sox2*-$^{INV/+}$ mice and embryos that carry inverted COIN cassette, were genotyped with INVdiaR2 (R2), INVdia.R4 (R4) (600 bp) and INVdiaF1: 5′ GTTTTCAGGGTGTTGTTTAG 3′, IN-Vdia(eGFP).F3 (F3) (300 bp) set of primers. Sequences for PCR amplification of *Tg(Sox2-cre)* and *Sox2*$^{\beta geo2}$ were found in MGI (ID: 103270 and 1915777 respectively). All PCR reactions were performed using 0.2 U/μL Taq Polymerase, standard PCR conditions, and 1 M betaine.

Imaging Analysis

Conventional bright field and fluorescence microscopy was performed under a Leica MZ16FA stereoscope, while the dissection of the embryos took place either in 1X cold PBS or in DMEM medium supplemented with 2 mM glutamine and 0.5 mM penicillin and streptomycin.

Western Blotting

E14.5 embryos were lysed by sonication and the resulting pellets were washed with PBS and dissolved in cold buffer A (20 mM Tris-HCl, 420 Mm NaCl, 0.2 Mm EDTA, 0.5 mM DTT, 25% glycerol, 0.5 mM PMSF, 1.5 mM $MgCl_2$, 0.5% NP40). Incubation at 4°C for 15 min and centrifugation for 15 min at 10,000×g followed. The supernatant was recovered and the protein concentration was determined by BCA Protein Assay Reagent (bicinchoninic acid) according to the instructions of the manufacturer (Thermo Scientific Pierce BCA Protein Assay Kit). Proteins (50 μg per lane) were separated on 10% SDS-polyacrylamide gel, transferred to nitrocellulose membrane and membrane was blocked in western blot blocking buffer (5% milk, 10 mM Tris-HCl pH 7.6, 0.15 mM NaCl, 0.05% Tween-20) for 2h at RT, incubated o/n with the primary antibody at 4°C. Sox2 rabbit Polyclonal IgG, goat anti-GFP polyclonal antibody (Santa Cruz Biotechnology, Inc, Santa Cruz, CA, USA) and goat β-actin polyclonal antibody (Cell Signaling Technology, Inc, Danvers, MA, USA) were used (1:1000). After extensive washing in TBST.1 (10 mM Tris-HCl, 0.15 mM NaCl, 0.1% Tween-20), goat anti-rabbit HRP conjugated secondary antibody was applied (1:10,000) for 2 h at RT. Proteins were visualized by chemiluminescence detection using ECL (Cell Signaling Technology, Inc., Danvers, MA, USA).

RNA Analysis

RNA was extracted from E14.5 mouse embryos and subjected to Taqman. Real-time PCR analysis typically, *Gapdh* was used as a control house-keeping gene, although analysis was also performed using *Cyclophlin* and *β-actin* with similar results. For miR1897 analysis, miR16 and Sno135 were used as internal controls. All probes are hydrolysis probes with 5' Fam Fluorophore and 3' quencher (BHQ) (Biosearch Technologies). Probes codes and sequences for each gene are: for mSox2, Applied Biosystems, Inc, TaqMan assay ID: Mm00488369_s1, for mSox2ot, Applied Biosystems, Inc, TaqMan assay ID: Mm01291217_m1, for Lac-Z, FRW: TTTCAGCCGCGCTGTACTGGA, RVS: TGTTGCCACTCGCTTTAATGATG, for eGFP: FRW:

Table 2. Analysis of progeny from *Sox2*$^{INV/+}$ × *Sox2*$^{+/+}$ intercrosses$^{\#}$.

Genotypic distribution in live progeny*				
	Live	**Dead**	***Sox2*$^{+/+}$**	***Sox2*$^{INV/+}$**
Age	No (%)	No (%)	No (%)	No (%)
E6.5	17	0	9 (53%)	8 (47%)
P21	73	0	40 (55%)	33 (45%)

Genotyping of *Sox2*$^{INV/+}$ heterozygous intercross progeny.
$^{\#}$Data collected from mice in C57BL6 background.
*Genotypes were assessed by PCR either from tail biopsies or whole embryos.
doi:10.1371/journal.pone.0045768.t002

Table 3. Analysis of progeny from *Sox2*$^{INV/+}$ × *Sox2*$^{INV/+}$ intercrosses$^{\#}$.

Genotypic distribution in live progeny*					
	Live	**Dead**	***Sox2*$^{INV/+}$**	***Sox2*$^{INV/INV}$**	***Sox2*$^{+/+}$**
Age	No	No (%)	No (%)	No (%)	No (%)
E6.5	18	6 (33%)	5 (28%)	0	7 (29%)
P21	43	0 (0%)	21 (49%)	0	22 (51%)

Genotyping of *Sox2*$^{INV/+}$ heterozygous intercross progeny.
$^{\#}$Data collected from mice in C57BL6 background.
*Genotypes were assessed by PCR either from tail biopsies or embryo tissue.
doi:10.1371/journal.pone.0045768.t003

TCTTCA AGTCCG CCATGCCCG, RVS: CTACCCCGAC-CACATGAAGC, for miR1897: Applied Biosystems, Inc, Taq-Man assay, Probe sequence: 121199.

Ethics Statement

All animals were handled in strict accordance with good animal practice as defined by the Animals Act 160/03.05.1991 applicable in Greece, revised according to the 86/609/EEC/24.11.1986 EU directive regarding the proper care and use of laboratory animals and in accordance to the Hellenic License for Animal Experimentation at the BSRC" Alexander Fleming" (Prot. No. 767/28.02.07) issued after protocol approval by the Animal Research Committee of the BSRC "Alexander Fleming" (Prot. No. 2762/03.08.05).

Acknowledgments

We are grateful to the Head of the Animal Facility Dr. Maria Kamber for her continuous support and Dimitris Loukopoulos for excellent animal care; Dr. Wolf Oehrl for assistance on the biochemical experiments; Dr. Liz Robertson (University of Oxford, Oxford, United Kingdom) for the *Sox2CRE* mouse line.

Author Contributions

Conceived and designed the experiments: AE ER. Performed the experiments: NM MS JLH AK PY. Analyzed the data: NM MS JLH AK PY AE ER. Wrote the paper: AE ER.

References

1. Kamachi Y, Iwafuchi M, Okuda Y, Takemoto T, Uchikawa M, et al. (2009) Evolution of non-coding regulatory sequences involved in the developmental process: reflection of differential employment of paralogous genes as highlighted by Sox2 and group B1 Sox genes. Proceedings of the Japan Academy Series B, Physical and biological sciences 85: 55–68.
2. Brosius J, Gould SJ (1992) On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proceedings of the National Academy of Sciences of the United States of America 89: 10706–10710.
3. Gentles AJ, Karlin S (1999) Why are human G-protein-coupled receptors predominantly intronless? Trends in genetics : TIG 15: 47–49.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304–1351.
5. Sakharkar KR, Sakharkar MK, Culiat CT, Chow VT, Pervaiz S (2006) Functional and evolutionary analyses on expressed intronless genes in the mouse genome. FEBS letters 580: 1472–1478.
6. Sakharkar MK, Perumal BS, Lim YP, Chern LP, Yu Y, et al. (2005) Alternatively spliced human genes by exon skipping–a database (ASHESdb). In silico biology 5: 221–225.
7. Hill AE, Sorscher EJ (2006) The non-random distribution of intronless human genes across molecular function categories. FEBS letters 580: 4303–4305.
8. Friend K, Lovejoy AF, Steitz JA (2007) U2 snRNP binds intronless histone pre-mRNAs to facilitate U7-snRNP-dependent 3' end formation. Molecular cell 28: 240–252.
9. Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, et al. (2005) P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. The EMBO journal 24: 4154–4165.
10. Maquat LE, Li X (2001) Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. RNA 7: 445–456.
11. Huang Y, Carmichael GG (1997) The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation of intronless gene transcripts and of unspliced HIV-1-related mRNAs. Proceedings of the National Academy of Sciences of the United States of America 94: 10104–10109.
12. Schilham MW, van Eijk M, van de Wetering M, Clevers HC (1993) The murine Sox-4 protein is encoded on a single exon. Nucleic acids research 21: 2009.
13. Elkouris M, Balaskas N, Poulou M, Politis PK, Panayiotou E, et al. (2011) Sox1 maintains the undifferentiated state of cortical neural progenitor cells via the suppression of Prox1-mediated cell cycle exit and neurogenesis. Stem Cells 29: 89–98.
14. Remboutsika E, Elkouris M, Iulianella A, Andoniadou CL, Poulou M, et al. (2011) Flexibility of neural stem cells. Front Physiol 2: 16.
15. Kirby PJ, Waters PD, Delbridge M, Svartman M, Stewart AN, et al. (2002) Cloning and mapping of platypus SOX2 and SOX14: insights into SOX group B evolution. Cytogenetic and genome research 98: 96–100.
16. Behbahaninia M, Ramey WL, Sindhwani MK, Kalani MY (2011) Differential expression of pluripotency factors Sox2 and Oct4 regulate neuronal and mesenchymal lineages. Neurosurgery 69: N19.
17. Yoon DS, Kim YH, Jung HS, Paik S, Lee JW (2011) Importance of Sox2 in maintenance of cell proliferation and multipotency of mesenchymal stem cells in low-density culture. Cell proliferation 44: 428–440.
18. Parrinello S, Napoli I, Ribeiro S, Digby PW, Fedorova M, et al. (2010) EphB signaling directs peripheral nerve regeneration through Sox2-dependent Schwann cell sorting. Cell 143: 145–155.
19. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, et al. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development 17: 126–140.
20. Cavallaro M, Mariani J, Lancini C, Latorre E, Caccia R, et al. (2008) Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. Development 135: 541–557.
21. Taranova OV, Magness ST, Fagan BM, Wu Y, Surzenko N, et al. (2006) SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes & development 20: 1187–1202.
22. Donner AL, Episkopou V, Maas RL (2007) Sox2 and Pou2f1 interact to control lens and olfactory placode development. Developmental biology 303: 784–799.
23. Okubo T, Pevny LH, Hogan BL (2006) Sox2 is required for development of taste bud sensory cells. Genes & development 20: 2654–2659.
24. Que J, Okubo T, Goldenring JR, Nam KT, Kurotani R, et al. (2007) Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 134: 2521–2531.
25. Ferri AL, Cavallaro M, Braida D, Di Cristofano A, Canta A, et al. (2004) Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development 131: 3805–3819.
26. Dabdoub A, Puligilla C, Jones JM, Fritzsch B, Cheah KS, et al. (2008) Sox2 signaling in prosensory domain specification and subsequent hair cell differentiation in the developing cochlea. Proceedings of the National Academy of Sciences of the United States of America 105: 18396–18401.
27. Amaral PP, Neyt C, Wilkins SJ, Askarian-Amiri ME, Sunkin SM, et al. (2009) Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA 15: 2013–2027.
28. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic acids research 34: D140–144.
29. Barrand S, Andersen IS, Collas P (2010) Promoter-exon relationship of H3 lysine 9, 27, 36 and 79 methylation on pluripotency-associated genes. Biochem Biophys Res Commun 401: 611–617.
30. Barrand S, Collas P Chromatin states of core pluripotency-associated genes in pluripotent, multipotent and differentiated cells. Biochem Biophys Res Commun 391: 762–767.
31. Alonso MM, Diez-Valle R, Manterola L, Rubio A, Liu D, et al. Genetic and epigenetic modifications of Sox2 contribute to the invasive phenotype of malignant gliomas. PLoS One 6: e26740.
32. Farthing CR, Ficz G, Ng RK, Chan CF, Andrews S, et al. (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet 4: e1000116.
33. Wiebe MS, Wilder PJ, Kelly D, Rizzino A (2000) Isolation, characterization, and differential expression of the murine Sox-2 promoter. Gene 246: 383–393.
34. Funabashi H, Takatsu M, Saito M, Matsuoka H (2010) Sox2 regulatory region 2 sequence works as a DNA nuclear targeting sequence enhancing the efficiency of an exogenous gene expression in ES cells. Biochem Biophys Res Commun 400: 554–558.
35. Iwafuchi-Doi M, Yoshida Y, Onichtchouk D, Leichsenring M, Driever W, et al. (2011) The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates. Dev Biol 352: 354–366.
36. Miyagi S, Nishimoto M, Saito T, Ninomiya M, Sawamoto K, et al. (2006) The Sox2 regulatory region 2 functions as a neural stem cell-specific enhancer in the telencephalon. J Biol Chem 281: 13374–13381.
37. Miyagi S, Saito T, Mizutani K, Masuyama N, Gotoh Y, et al. (2004) The Sox-2 regulatory regions display their activities in two distinct types of multipotent stem cells. Mol Cell Biol 24: 4207–4220.
38. Sikorska M, Sandhu JK, Deb-Rinker P, Jezierski A, Leblanc J, et al. (2008) Epigenetic modifications of SOX2 enhancers, SRR1 and SRR2, correlate with in vitro neural differentiation. J Neurosci Res 86: 1680–1693.
39. Tomioka M, Nishimoto M, Miyagi S, Katayanagi T, Fukui N, et al. (2002) Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res 30: 3202–3213.
40. Griffiths-Jones S (2006) miRBase: the microRNA sequence database. Methods in molecular biology 342: 129–138.
41. Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, et al. (2003) Mutations in SOX2 cause anophthalmia. Nature genetics 33: 461–463.

42. Economides AN, Frendewey D, Yang P, Dominguez MG, Dore AT, et al. (2012) Conditionals-by-Inversion: a novel universal method for the generation of conditional alleles. in preparation.
43. Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7: 649–659.
44. Ekonomou A, Kazanis I, Malas S, Wood H, Alifragis P, et al. (2005) Neuronal migration and ventral subtype identity in the telencephalon depend on SOX1. PLoS biology 3: e186.
45. Nagy A (2000) Cre recombinase: the universal reagent for genome tailoring. Genesis 26: 99–109.
46. Ringrose L, Chabanis S, Angrand PO, Woodroofe C, Stewart AF (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. Embo J 18: 6630–6641.
47. Testa G, Schaft J, van der Hoeven F, Glaser S, Anastassiadis K, et al. (2004) A reliable lacZ expression reporter cassette for multipurpose, knockout-first alleles. Genesis 38: 151–158.
48. Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, et al. (2011) A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474: 337–342.
49. Seoighe C, Korir PK (2011) Evidence for intron length conservation in a set of mammalian genes associated with embryonic development. BMC Bioinformatics 12 Suppl 9: S16.
50. Swinburne IA, Miguez DG, Landgraf D, Silver PA (2008) Intron length increases oscillatory periods of gene expression in animal cells. Genes Dev 22: 2342–2346.
51. Swinburne IA, Silver PA (2008) Intron delays and transcriptional timing during development. Dev Cell 14: 324–330.
52. Miyagi S, Masui S, Niwa H, Saito T, Shimazaki T, et al. (2008) Consequence of the loss of Sox2 in the developing brain of the mouse. FEBS letters 582: 2811–2815.
53. Smith AN, Miller LA, Radice G, Ashery-Padan R, Lang RA (2009) Stage-dependent modes of Pax6-Sox2 epistasis regulate lens development and eye morphogenesis. Development 136: 2977–2985.
54. Favaro R, Valotta M, Ferri AL, Latorre E, Mariani J, et al. (2009) Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nature neuroscience 12: 1248–1256.
55. Valenzuela DM, Murphy AJ, Frendewey D, Gale NW, Economides AN, et al. (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nature biotechnology 21: 652–659.
56. Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet 20: 123–128.

Appendix III

List of Abbreviations

SRY, Sex-determining region Y;

HMG, High Mobility Group box

ncRNA, long noncoding RNA

***Sox2ot*,** overlapping transcript

***Sox2dot*,** distal overlapping transcript

ExEn, extraembryonic endoderm

NSCs, Neural stem cells

ICM, inner cell mass

EB, epiblast

HB, hypoblast

EC, ectoderm

PS, primitive streak

NE, neuroectoderm

VZ, ventricular zone

SVZ, subventricular zone

KO, knock-out

eGFP, enhanced green fluorescent protein

CNS, central nervous system

EA, esophageal atresia

NCC, neural crest cells

BMP, bone morphogenetic protein

EMT, epithelial-mesenchymal transition

COIN, Conditional by Invertion

BAC, bacterial artificial chromosome

ORF, open reading frame

FRT, Flp Recognition Target

FLP, flipase

PE, primitive ectoderm

ECR, evolutionary conserved region

NP, neural progenitor

Appendix IV

The role of Sox2 in radial glia cells

Sox2 is an HMG-box containing transcription factor required for pluripotency of embryonic stem cells and formation of the epiblast. During early development, Sox2 is expressed in the morula, inner cell mass of the blastocyst and later on in the sensory placodes, foregut and the nervous system. Sox2 null embryos die at implantation stage hampering the analysis of Sox2 function during development. We have demonstrated that Sox2 is important for the maintenance of cortical radial glia identity *ex vivo*. Here, we describe a novel Sox2 Conditional by Inversion (COIN) allele in radial glial cells *in vivo*.

Appendix V

Sox2 COIN sequence

```
LOCUS       SOX2 (129)   7099 BP DS-DNA             SYN
13-MAR-1996
DEFINITION  M.musculus SOX2 gene.
ACCESSION   X94127
VERSION     X94127.1  GI:1209429
KEYWORDS    sox-2 gene; SOX2 protein.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi;
Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Collignon,J., Sockanathan,S., Hacker,A., Cohen-
Tannoudji,M.,
            Norris,D., Rattan's., Stevanovic,M.,
Goodfellow,P.N. and
            Lovell-Badge,R.
  TITLE     A comparison of the properties of Sox-3 with Sry
and two related
            genes, Sox-1 and Sox-2
  JOURNAL   Development 122 (2), 509-520 (1996)
  MEDLINE   96189340
REFERENCE   2  (bases 1 to 2418)
  AUTHORS   Sockanathan,T.E.L.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1995) TEL.S. Sockanathan,
National Institute For
            Medical Research, The Ridgeway, Mill Hill, London
NW7 1AA, UK
COMMENT     Overlaps with X55491, D50603 and U31967.
FEATURES             Location/Qualifiers
     source          join(1..>436,<5118..7099)
                     /organism="Mus musculus"
                     /note="mol_type=genomic DNA"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /dev_stage="8.5 dpc [Split]"
     gene            join(407..>436,<5118..6047)
                     /gene="sox-2 [Split]"
     CDS             join(407..>436,<5118..6047)
                     /gene="sox-2"
                     /function="transcription factor"
                     /codon_start="1"
```

```
                     /product="SOX2 protein"
                     /protein_id="CAA63847.1"
                     /db_xref="GI:1209430"
                     /db_xref="GOA:P48432"
                     /db_xref="Swiss-Prot:P48432"

/translation="MYNMMETELKPPGPQQASGGGGGGGNATAAATGGNQKNSPDRVK
                     RPMNAFMVWSRGQRRK... [Split]"
     misc_feature    5208..5444
                     /gene="sox-2"
                     /note="HMG box"
     misc_feature    436..436
                     /note="insert IMeC after this base? …ƒ0
cassette!"
     primer_bind     402..422
                     /note="MaF"
     primer_bind     join(425..>436,<5118..5120)
                     /note="MaP [Split]"
     primer_bind     complement(5122..5135)
                     /note="MaR"
     primer_bind     401..421
                     /note="MbF"
     primer_bind     join(425..>436,<5118..5120)
                     /note="MbP [Split]"
     primer_bind     complement(5124..5139)
                     /note="MbR"
     primer_bind     411..433
                     /note="McF"
     primer_bind     5128..5139
                     /note="McP"
     primer_bind     complement(5150..5164)
                     /note="McR"
     misc_recomb     361..436
                     /note="HB5' for Sox@ COIN"
     misc_recomb     5118..5318
                     /note="HB3' for Sox2 COIN"
     frag            437..5117
                     /note="1 to 4681 of
pINV.(GFP@ƒ0.ƒUNƒ)2.Sox2 DON"
     primer_bind     5064..5083
                     /note="INVpA(3' end).dia1-5'"
     primer_bind     5001..5021
                     /note="INVpA(3' end).dia2-5'"
     primer_bind     1814..1834
                     /note="EGFP.MPErc"
     intron          <437..>5117
                     /note="-1 globin intron B w/
(L71).(SA/eGFP(ATG-)@ƒ20-
                     glpA)/L66.(ƒUNƒ)"
     misc_feature    2357..>2359
                     /note="PspOMI site"
     misc_recomb     <2149..2178
```

```
                     /note="Lox66"
     misc_feature    complement(2166..2178)
                     /note="irL (∂LoxP)"
     misc_feature    <2149..2157
                     /note="irR (∂Lox66)"
     misc_recomb     complement(801..<833)
                     /note="Lox71"
     misc_feature    <801..812
                     /note="irL of Lox71"
     misc_feature    complement(821..833)
                     /note="irL (∂LoxP)"
     polyA_signal    complement(>834..1371)
                     /note="human ß-globin pA region"
     frag            complement(>4160..5061)
                     /note="1 to 2719 of ƒUBNƒ PmeI (∂1872)"
     misc_recomb     complement(5009..5056)
                     /note="FRT"
     protein_bind    complement(4897..4902)
                     /note="Sp 1 binding site"
     promoter        complement(>3789..5001)
                     /note="hUBp (human Ubiquitin promoter)"
     primer          complement(4754..4772)
                     /note="hUBifw1"
     primer          complement(4526..4546)
                     /note="hUBi.pr.rc2"
     primer          complement(4305..4329)
                     /note="hUBiprfw2"
     TATA_signal     complement(4692..4699)
                     /note="TATA box"
     protein_bind    complement(4713..4718)
                     /note="Sp 1 binding site"
     misc_feature    complement(4667..4667)
                     /note="transcription start site"
     exon            complement(4604..4667)
                     /note="exon 1"
     intron          complement(>3792..4603)
                     /note="intron 1"
     primer          4606..4623
                     /note="hUB.seq1-3' (Tm = 48°C)"
     primer          complement(4608..4626)
                     /note="hUB.seq2-5' (Tm = 51°C)"
     primer          4303..4320
                     /note="hUB.seq3-3' (Tm = 51°C)"
     primer          complement(4255..4275)
                     /note="hUB.seq4-5' (Tm = 50°C)"
     primer          4849..4869
                     /note="hUB.seq7-3' (Tm = 52°C)"
     primer          complement(4849..4869)
                     /note="hUB.seq6-5' (Tm = 52°C)"
     frag            5062..>5117
                     /note="1 to 70 of 3' splice
site+(BsmBI).ClaI"
```

```
     misc_RNA        5116..5117
                     /note="SA"
     misc_RNA        5102..>5115
                     /note="12Py N C tract of SA region"
     misc_RNA        5081..5087
                     /note="Lariat region (predicted)"
     misc_binding    5072..5077
                     /note="PspOMI"
     frag            <437..800
                     /note="1 to 386 of NotI-MluI  PCR for ƒ2
[Split]"
     misc_RNA        <437..442
                     /note="5' splice consensus"
     primer_bind     <437..459
                     /note="iZ RED PCR5' primer"
     primer_bind     complement(514..537)
                     /note="iZ RED.5' dia.2-3'"
     primer_bind     4997..5015
                     /note="UBP.seq1-3'"
     primer_bind     4922..4938
                     /note="UBP.seq3-3'"
     frag            1372..2148
                     /note="1 to 777 of pCRT.(L66)SA-
GFP(ƒ0).L71"
     frag            complement(1378..<2089)
                     /note="679 to 1398 of pEGFP-N1 (361)
[Split]"
     CDS             complement(1378..<2094)
                     /note="eGFP (Met-) ƒ0"
     misc_feature    complement(>2095..<2140)
                     /note="SA region from ß-1 globin intron
B"
     misc_RNA        complement(2095..2096)
                     /note="SA"
     misc_RNA        complement(2097..2110)
                     /note="12Py N C tract of SA region"
     misc_RNA        complement(2125..2131)
                     /note="Lariat region (predicted)"
     frag            complement(>2371..4159)
                     /note="6590 to 8382 of pZEN-UB2 FRT(1872)
[Split]"
     DNA             <2434..3062
                     /note="PGK-neo [Split]"
     frag            complement(2434..2917)
                     /note="poly A "
     CDS             complement(2918..<3721)
                     /note="neo"
     polyA_site      complement(2434..2917)
                     /note="poly A"
     misc_recomb     complement(>2381..2427)
                     /note="FRT"
     promoter        complement(3722..3788)
```

```
                     /note="EM7 promoter"
     primer          complement(3809..3829)
                     /note="hUbi.pr.rc1"
     exon            complement(>3789..3791)
                     /note="5' end of exon 2"
     primer          3942..3959
                     /note="hUB.seq5-3' (Tm = 48°C)"
     primer          complement(850..874)
                     /note="2154 MUR"
     primer_bind     834..847
                     /note="2154 MUP"
     primer          774..798
                     /note="2154 MUF"
     primer_bind     2184..2211
                     /note="2154 del MDP"
     primer          complement(2216..2239)
                     /note="2154 MDR"
     primer          2100..2122
                     /note="2154 MDF"
     primer          244..261
                     /note="Sox2C.dia1-5'"
     primer          211..227
                     /note="Sox2C.dia2-5'"
     primer          complement(5326..5343)
                     /note="Sox2C.dia3-3'"
     primer          complement(5368..5385)
                     /note="Sox2C.dia4-3'"
BASE COUNT      1810 A      1866 C      1795 G      1628 T
0 OTHER
ORIGIN      -
        1 CTTGTTCAAA AAAGTATCAG GAGTTGTCAA GGCAGAGAAG
AGAGTGTTTG CAAAAAGGGA
       61 AAAGTACTTT GCTGCCTCTT TAAGACTAGG GCTGGGAGAA
AGAAGAGGAG AGAGAAAGAA
      121 AGGAGAGAAG TTTGGAGCCC GAGGCTTAAG CCTTTCCAAA
AACTAATCAC AACAATCGCG
      181 GCGGCCCGAG GAGGAGAGCG CCTGTTTTTT CATCCCAATT
GCACTTCGCC CGTCTCGAGC
      241 TCCGCTTCCC CCCAACTATT CTCCGCCAGA TCTCCGCGCA
GGGCCGTGCA CGCCGAGGCC
      301 CCCGCCCGCG GCCCCTGCAT CCCGGCCCCC GAGCGCGGCC
CCCACAGTCC TGGCCGGGCC
      361 GAGGGTTGGC GGCCGCCGGC GGGCCGCGCC CCGCCCCAGC
GCGCGCATGT ATAACATGAT
      421 GGAGACGGAG CTGAAGGTGA GTTTGGGGAC CCTTGATTGT
TCTTTCTTTT TCGCTATTGT
      481 AAAATTCATG TTATATGGAG GGGGCAAAGT TTTCAGGGTG
TTGTTTAGAA TGGGAAGATG
      541 TCCCTTGTAT CACCATGGAC CCTCATGATA ATTTTGTTTC
TTTCACTTTC TACTCTGTTG
      601 ACAACCATTG TCTCCTCTTA TTTTCTTTTC ATTTTCTGTA
ACTTTTTCGT TAAACTTTAG
```

661 CTTGCATTTG TAACGAATTT TTAAATTCAC TTTTGTTTAT
TTGTCAGATT GTAAGTACTT
721 TCTCTAATCA CTTTTTTTTC AAGGCAATCA GGGTATATTA
TATTGTACTT CAGCACAGTT
781 TTAGAGAACA ATTGACGCGT ACCGTTCGTA TAGCATACAT
TATACGAAGT TATCTCGAGG
841 GATCTCCATA AGAGAAGAGG GACAGCTATG ACTGGGAGTA
GTCAGGAGAG GAGGAAAAAT
901 CTGGCTAGTA AAACATGTAA GGAAAATTTT AGGGATGTTA
AAGAAAAAAA TAACACAAAA
961 CAAAATATAA AAAAAATCTA ACCTCAAGTC AAGGCTTTTC
TATGGAATAA GGAATGGACA
1021 GCAGGGGGCT GTTTCATATA CTGATGACCT CTTTATAGCC
AACCTTTGTT CATGGCAGCC
1081 AGCATATGGG CATATGTTGC CAAACTCTAA ACCAAATACT
CATTCTGATG TTTTAAATGA
1141 TTTGCCCTCC CATATGTCCT TCCGAGTGAG AGACACAAAA
AATTCCAACA CACTATTGCA
1201 ATGAAAATAA ATTTCCTTTA TTAGCCAGAA GTCAGATGCT
CAAGGGGCTT CATGATGTCC
1261 CCATAATTTT TGGCAGAGGG AAAAAGATCT CAGTGGTATT
TGTGAGCCAG GGCATTGGCC
1321 ACACCAGCCA CCACCTTCTG ATAGGCAGCC TGCACCTGAG
GTACGTACTT TGCTAGCTTA
1381 CTTGTACAGC TCGTCCATGC CGAGAGTGAT CCCGGCGGCG
GTCACGAACT CCAGCAGGAC
1441 CATGTGATCG CGCTTCTCGT TGGGGTCTTT GCTCAGGGCG
GACTGGGTGC TCAGGTAGTG
1501 GTTGTCGGGC AGCAGCACGG GGCCGTCGCC GATGGGGGTG
TTCTGCTGGT AGTGGTCGGC
1561 GAGCTGCACG CTGCCGTCCT CGATGTTGTG GCGGATCTTG
AAGTTCACCT TGATGCCGTT
1621 CTTCTGCTTG TCGGCCATGA TATAGACGTT GTGGCTGTTG
TAGTTGTACT CCAGCTTGTG
1681 CCCCAGGATG TTGCCGTCCT CCTTGAAGTC GATGCCCTTC
AGCTCGATGC GGTTCACCAG
1741 GGTGTCGCCC TCGAACTTCA CCTCGGCGCG GGTCTTGTAG
TTGCCGTCGT CCTTGAAGAA
1801 GATGGTGCGC TCCTGGACGT AGCCTTCGGG CATGGCGGAC
TTGAAGAAGT CGTGCTGCTT
1861 CATGTGGTCG GGGTAGCGGC TGAAGCACTG CACGCCGTAG
GTCAGGGTGG TCACGAGGGT
1921 GGGCCAGGGC ACGGGCAGCT TGCCGGTGGT GCAGATGAAC
TTCAGGGTCA GCTTGCCGTA
1981 GGTGGCATCG CCCTCGCCCT CGCCGGACAC GCTGAACTTG
TGGCCGTTTA CGTCGCCGTC
2041 CAGCTCGACC AGGATGGGCA CCACCCCGGT GAACAGCTCC
TCGCCCTTGC TCACCTGTAG
2101 GAAAAAGAAG AAGGCATGAA CATGGTTAGC AGAGGGGCCC
GCGGTACCGT TCGTATAATG
2161 TATGCTATAC GAAGTTATCC GGACAATTGT TATAATTAAA
TGATAAGGTA GAATATTTCT

2221 GCATATAAAT TCTGGCTGGC GTGGAAATAT TCTTATTGGT
AGAAACAACT ACACCCTGGT
2281 CATCATCCTG CCTTTCTCTT TATGGTTACA ATGATATACA
CTGTTTGAGA TGAGGATAAA
2341 ATACTCTGAG TCCAAACCGG GCCGCGACTG AAACCCTAGG
AAGTTCCTAT ACTTTCTAGA
2401 GAATAGGAAC TTCGGAATAG GAACTTCTCT AGGGCTGCAG
GTCGAGGTCT GATGGAATTA
2461 GAACTTGGCA AAACAATACT GAGAATGAAG TGTATGTGGA
ACAGAGGCTG CTGATCTCGT
2521 TCTTCAGGCT ATGAAACTGA CACATTTGGA AACCACAGTA
CTTAGAACCA CAAAGTGGGA
2581 ATCAAGAGAA AAACAATGAT CCCACGAGAG ATCTATAGAT
CTATAGATCA TGAGTGGGAG
2641 GAATGAGCTG GCCCTTAATT TGGTTTTGCT TGTTTAAATT
ATGATATCCA ACTATGAAAC
2701 ATTATCATAA AGCAATAGTA AAGAGCCTTC AGTAAAGAGC
AGGCATTTAT CTAATCCCAC
2761 CCCACCCCCA CCCCCGTAGC TCCAATCCTT CCATTCAAAA
TGTAGGTACT CTGTTCTCAC
2821 CCTTCTTAAC AAAGTATGAC AGGAAAAACT TCCATTTTAG
TGGACATCTT TATTGTTTAA
2881 TAGATCATCA ATTTCTGCAG ACTTACAGCG GATCCCCTCA
GAAGAACTCG TCAAGAAGGC
2941 GATAGAAGGC GATGCGCTGC GAATCGGGAG CGGCGATACC
GTAAAGCACG AGGAAGCGGT
3001 CAGCCCATTC GCCGCCAAGC TCTTCAGCAA TATCACGGGT
AGCCAACGCT ATGTCCTGAT
3061 AGCGGTCCGC CACACCCAGC CGGCCACAGT CGATGAATCC
AGAAAAGCGG CCATTTTCCA
3121 CCATGATATT CGGCAAGCAG GCATCGCCAT GGGTCACGAC
GAGATCATCG CCGTCGGGCA
3181 TGCGCGCCTT GAGCCTGGCG AACAGTTCGG CTGGCGCGAG
CCCCTGATGC TCTTCGTCCA
3241 GATCATCCTG ATCGACAAGA CCGGCTTCCA TCCGAGTACG
TGCTCGCTCG ATGCGATGTT
3301 TCGCTTGGTG GTCGAATGGG CAGGTAGCCG GATCAAGCGT
ATGCAGCCGC CGCATTGCAT
3361 CAGCCATGAT GGATACTTTC TCGGCAGGAG CAAGGTGAGA
TGACAGGAGA TCCTGCCCCG
3421 GCACTTCGCC CAATAGCAGC CAGTCCCTTC CCGCTTCAGT
GACAACGTCG AGCACAGCTG
3481 CGCAAGGAAC GCCCGTCGTG GCCAGCCACG ATAGCCGCGC
TGCCTCGTCC TGCAGTTCAT
3541 TCAGGGCACC GGACAGGTCG GTCTTGACAA AAAGAACCGG
GCGCCCCTGC GCTGACAGCC
3601 GGAACACGGC GGCATCAGAG CAGCCGATTG TCTGTTGTGC
CCAGTCATAG CCGAATAGCC
3661 TCTCCACCCA AGCGGCCGGA GAACCTGCGT GCAATCCATC
TTGTTCAATG GCCGATCCCA
3721 TGGTTTAGTT CCTCACCTTG TCGTATTATA CTATGCCGAT
ATACTATGCC GATGATTAAT

3781 TGTCAACACG TCTAACAAAA AAGCCAAAAA CGGCCAGAAT
TTAGCGGACA ATTTACTAGT
3841 CTAACACTGA AAATTACATA TTGACCCAAA TGATTACATT
TCAAAAGGTG CCTAAAAAAC
3901 TTCACAAAAC ACACTCGCCA ACCCCGAGCG CATAGTTCAA
AACCGGAGCT TCAGCTACTT
3961 AAGAAGATAG GTACATAAAA CCGACCAAAG AAACTGACGC
CTCACTTATC CCTCCCCTCA
4021 CCAGAGGTCC GGCGCCTGTC GATTCAGGAG AGCCTACCCT
AGGCCCGAAC CCTGCGTCCT
4081 GCGACGGAGA AAAGCCTACC GCACACCTAC CGGCAGGTGG
CCCCACCCTG CATTATAAGC
4141 CAACAGAACG GGTGACGTCA CGACACGACG AGGGCGCGCG
CTCCCAAAGG TACGGGTGCA
4201 CTGCCCAACG GCACCGCCAT AACTGCCGCC CCCGCAACAG
ACGACAAACC GAGTTCTCCA
4261 GTCAGTGACA AACTTCACGT CAGGGTCCCC AGATGGTGCC
CCAGCCCATC TCACCCGAAT
4321 AAGAGCTTTC CCGCATTAGC GAAGGCCTCA AGACCTTGGG
TTCTTGCCGC CCACCATGCC
4381 CCCCACCTTG TTTCAACGAC CTCACAGCCC GCCTCACAAG
CGTCTTCCAT TCAAGACTCG
4441 GGAACAGCCG CCATTTTGCT GCGCTCCCCC CAACCCCCAG
TTCAGGGCAA CCTTGCTCGC
4501 GGACCCAGAC TACAGCCCTT GGCGGTCTCT CCACACGCTT
CCGTCCCACC GAGCGGCCCG
4561 GCGGCCACGA AAGCCCCGGC CAGCCCAGCA GCCCGCTACT
CACCAAGTGA CGATCACAGC
4621 GATCCACAAA CAAGAACCGC GACCCAAATC CCGGCTGCGA
CGGAACTAGC TGTGCCACAC
4681 CCGGCGCGTC CTTATATAAT CATCGGCGTT CACCGCCCCA
CGGAGATCCC TCCGCAGAAT
4741 CGCCGAGAAG GGACTACTTT TCCTCGCCTG TTCCGCTCTC
TGGAAAGAAA ACCAGTGCCC
4801 TAGAGTCACC CAAGTCCCGT CCTAAAATGT CCTTCTGCTG
ATACTGGGGT TCTAAGGCCG
4861 AGTCTTATGA GCAGCGGGCC GCTGTCCTGA GCGTCCGGGC
GGAAGGATCA GGACGCTCGC
4921 TGCGCCCTTC GTCTGACGTG GCAGCGCTCG CCGTGAGGAG
GGGGGCGCCC GCGGGAGGCG
4981 CCAAAACCCG GCGCGGAGGC CATGCATAGA AGTTCCTATA
CTTTCTAGAG AATAGGAACT
5041 TCGGAATAGG AACTTCTCGA GCCCGGGAAA TGGGCCCCTC
TGCTAACCAT GTTCATGCCT
5101 TCTTCTCTTT CCTACAGCCG CCGGGCCCGC AGCAAGCTTC
GGGGGGCGGC GGCGGAGGAG
5161 GCAACGCCAC GGCGGCGGCG ACCGGCGGCA ACCAGAAGAA
CAGCCCGGAC CGCGTCAAGA
5221 GGCCCATGAA CGCCTTCATG GTATGGTCCC GGGGGCAGCG
GCGTAAGATG GCCCAGGAGA
5281 ACCCCAAGAT GCACAACTCG GAGATCAGCA AGCGCCTGGG
CGCGGAGTGG AAACTTTTGT

5341 CCGAGACCGA GAAGCGGCCG TTCATCGACG AGGCCAAGCG
GCTGCGCGCT CTGCACATGA
5401 AGGAGCACCC GGATTATAAA TACCGGCCGC GGCGCAAAAC
CAAGACGCTC ATGAAGAAGG
5461 ATAAGTACAC GCTTCCCGGA GGCTTGCTGG CCCCCGGCGG
GAACAGCATG GCGAGCGGGG
5521 TTGGGGTGGG CGCCGGCCTG GGTGCGGGCG TGAACCAGCG
CATGGACAGC TACGCGCACA
5581 TGAACGGCTG GAGCAACGGC AGCTACAGCA TGATGCAGGA
GCAGCTGGGC TACCCGCAGC
5641 ACCCGGGCCT CAACGCTCAC GGCGCGGCAC AGATGCAACC
GATGCACCGC TACGACGTCA
5701 GCGCCCTGCA GTACAACTCC ATGACCAGCT CGCAGACCTA
CATGAACGGC TCGCCCACCT
5761 ACAGCATGTC CTACTCGCAG CAGGGCACCC CCGGTATGGC
GCTGGGCTCC ATGGGCTCTG
5821 TGGTCAAGTC CGAGGCCAGC TCCAGCCCCC CCGTGGTTAC
CTCTTCCTCC CACTCCAGGG
5881 CGCCCTGCCA GGCCGGGGAC CTCCGGGACA TGATCAGCAT
GTACCTCCCC GGCGCCGAGG
5941 TGCCGGAGCC CGCTGCGCCC AGTAGACTGC ACATGGCCCA
GCACTACCAG AGCGGCCCGG
6001 TGCCCGGCAC GGCCATTAAC GGCACACTGC CCCTGTCGCA
CATGTGAGGG CTGGACTGCG
6061 AACTGGAGAA GGGGAGAGAT TTTCAAAGAG ATACAAGGGA
ATTGGGAGGG GTGCAAAAAG
6121 AGGAGAGTAG GAAAAATCTG ATAATGCTCA AAAAGAAAAA
AACCACCAAT CCCATCCAAA
6181 TTAACGCAAA AACCGTGATG CCGACTAGAA AACTTTTATG
AGAGATCTTG GGACTTCTTT
6241 TTGGGGGACT ATTTTTGTAC AGAGAAAACC TGAGGGCGGC
GGGGAGGGCG GGGGAATCGG
6301 ACCATGTATA GATCTGGAGG AAAAAAACTA CGCAAAACTT
TTTTTTAAAG TTCTAGTGGT
6361 ACGTTAGGCG CTTCGCAGGG AGTTCGCAAA AGTCTTTACC
AGTAATATTT AGAGCTAGAC
6421 TCCGGGCGAT GAAAAAAAAA GTTTTAATAT TTGCAAGCAA
CTTTTGTACA GTATTTATCG
6481 AGATAAACAT GGCAATCAAA TGTCCATTGT TTATAAGCTG
AGAATTTGCC AATATTTTTC
6541 GAGGAAAGGG TTCTTGCTGG GTTTGATCTG CAGCTTAAAT
TTAGGACCGT TACAAACAAG
6601 GAAGGAGTTT ATTCGGATTT GAACATTTTA GTTTTAAAAT
TGTACAAAAG GAAAACATGA
6661 GAGCAAGTAC TGGCAAGACC GTTTTCGTGG TCTTGTTTAA
GGCAAACGTT CTAGATTGTA
6721 CTAAATTTTT AACTTACTGT TAAAGGCAAA AAAAAAATGT
CCATGCAGGT TGATATCGTT
6781 GGTAATTTAT AATAGCTTTT GTTCAATCCT ACCCTTTCAT
TTTGTTCACA TAAAAAATAT
6841 GGAATTACTG TGTTTGAAAT ATTTTCTTAT GGTTTGTAAT
ATTTCTGTAA ATTGTGATAT

```
     6901 TTTAAGGTTT TTCCCCCCTT TTATTTTCCG TAGTTGTATT
TTAAAAGATT CGGCTCTGTT
     6961 ATTGGAATCA GGCTCCGAGA ATCCATGTAT ATATTTGAAC
TAATACCATC CTTATAACAG
     7021 CTACATTTTC AACTTAAGTT TTTACTCCAT TATGCACAGT
TTGAGATAAA TAAATTTTTG
     7081 AAAAAAAAAA AAAAAAAAA
//
```

REFERENCES

Abu-Issa, R., and Kirby, M.L. (2007). Heart field: from mesoderm to heart tube. Annual review of cell and developmental biology *23*, 45-68.

Adameyko, I., Lallemend, F., Furlan, A., Zinin, N., Aranda, S., Kitambi, S.S., Blanchart, A., Favaro, R., Nicolis, S., Lubke, M.*, et al.* (2012). Sox2 and Mitf cross-regulatory interactions consolidate progenitor and melanocyte lineages in the cranial neural crest. Development *139*, 397-410.

Agathocleous, M., Iordanova, I., Willardsen, M.I., Xue, X.Y., Vetter, M.L., Harris, W.A., and Moore, K.B. (2009). A directional Wnt/beta-catenin-Sox2-proneural pathway regulates the transition from proliferation to differentiation in the Xenopus retina. Development *136*, 3289-3299.

Ahlstrom, J.D., and Erickson, C.A. (2009). New views on the neural crest epithelial-mesenchymal transition and neuroepithelial interkinetic nuclear migration. Commun Integr Biol *2*, 489-493.

Akitaya, T., and Bronner-Fraser, M. (1992). Expression of cell adhesion molecules during initiation and cessation of neural crest cell migration. Developmental dynamics : an official publication of the American Association of Anatomists *194*, 12-20.

Alatzoglou, K.S., Kelberman, D., and Dattani, M.T. (2009). The role of SOX proteins in normal pituitary development. J Endocrinol *200*, 245-258.

Albert, H., Dale, E.C., Lee, E., and Ow, D.W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J *7*, 649-659.

Alcock, J., Lowe, J., England, T., Bath, P., and Sottile, V. (2009). Expression of Sox1, Sox2 and Sox9 is maintained in adult human cerebellar cortex. Neurosci Lett *450*, 114-116.

Alonso, M.M., Diez-Valle, R., Manterola, L., Rubio, A., Liu, D., Cortes-Santiago, N., Urquiza, L., Jauregi, P., Lopez de Munain, A., Sampron, N.*, et al.* Genetic and epigenetic modifications of Sox2 contribute to the invasive phenotype of malignant gliomas. PLoS One *6*, e26740.

Amaral, P.P., Neyt, C., Wilkins, S.J., Askarian-Amiri, M.E., Sunkin, S.M., Perkins, A.C., and Mattick, J.S. (2009). Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA *15*, 2013-2027.

Aquino, J.B., Hjerling-Leffler, J., Koltzenburg, M., Edlund, T., Villar, M.J., and Ernfors, P. (2006). In vitro and in vivo differentiation of boundary cap neural crest stem cells into mature Schwann cells. Exp Neurol *198*, 438-449.

Aramaki, M., Kimura, T., Udaka, T., Kosaki, R., Mitsuhashi, T., Okada, Y., Takahashi, T., and Kosaki, K. (2007). Embryonic expression profile of chicken CHD7, the ortholog of the causative gene for CHARGE syndrome. Birth Defects Res A Clin Mol Teratol *79*, 50-57.

Aramaki, M., Udaka, T., Kosaki, R., Makita, Y., Okamoto, N., Yoshihashi, H., Oki, H., Nanao, K., Moriyama, N., Oku, S.*, et al.* (2006a). Phenotypic spectrum of CHARGE syndrome with CHD7 mutations. J Pediatr *148*, 410-414.

Aramaki, M., Udaka, T., Torii, C., Samejima, H., Kosaki, R., Takahashi, T., and Kosaki, K. (2006b). Screening for CHARGE syndrome mutations in the CHD7 gene using denaturing high-performance liquid chromatography. Genet Test *10*, 244-251.

Archer, T.C., Jin, J., and Casey, E.S. (2011). Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis. Dev Biol *350*, 429-440.
Arenkiel, B.R., Gaufo, G.O., and Capecchi, M.R. (2003). Hoxb1 neural crest preferentially form glia of the PNS. Dev Dyn *227*, 379-386.
Arenkiel, B.R., Tvrdik, P., Gaufo, G.O., and Capecchi, M.R. (2004). Hoxb1 functions in both motoneurons and in tissues of the periphery to establish and maintain the proper neuronal circuitry. Genes Dev *18*, 1539-1552.
Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003a). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.
Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003b). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.
Aybar, M.J., Glavic, A., and Mayor, R. (2002). Extracellular signals, cell interactions and transcription factors involved in the induction of the neural crest cells. Biol Res *35*, 267-275.
Aybar, M.J., and Mayor, R. (2002). Early induction of neural crest cells: lessons learned from frog, fish and chick. Current opinion in genetics & development *12*, 452-458.
Ayer-Le Lievre, C.S., and Le Douarin, N.M. (1982). The early development of cranial sensory ganglia and the potentialities of their component cells studied in quail-chick chimeras. Developmental biology *94*, 291-310.
Barrand, S., Andersen, I.S., and Collas, P. (2010). Promoter-exon relationship of H3 lysine 9, 27, 36 and 79 methylation on pluripotency-associated genes. Biochem Biophys Res Commun *401*, 611-617.
Barrand, S., and Collas, P. Chromatin states of core pluripotency-associated genes in pluripotent, multipotent and differentiated cells. Biochem Biophys Res Commun *391*, 762-767.
Behbahaninia, M., Ramey, W.L., Sindhwani, M.K., and Kalani, M.Y. (2011). Differential expression of pluripotency factors Sox2 and Oct4 regulate neuronal and mesenchymal lineages. Neurosurgery *69*, N19.
Benarroch, E.E. (2009). Brain iron homeostasis and neurodegenerative disease. Neurology *72*, 1436-1440.
Betancur, P., Sauka-Spengler, T., and Bronner, M. (2011). A Sox10 enhancer element common to the otic placode and neural crest is activated by tissue-specific paralogs. Development *138*, 3689-3698.
Bigalke, J.T., and Burrill, S.J. (2007). Time for a second look at SOX compliance. Healthc Financ Manage *61*, 56-62.
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.
Bremer, M., Frob, F., Kichko, T., Reeh, P., Tamm, E.R., Suter, U., and Wegner, M. (2011). Sox10 is required for Schwann-cell homeostasis and myelin maintenance in the adult peripheral nerve. Glia *59*, 1022-1032.
Bronner-Fraser, M. (2002). Molecular analysis of neural crest formation. Journal of physiology, Paris *96*, 3-8.
Brosius, J., and Gould, S.J. (1992). On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc Natl Acad Sci U S A *89*, 10706-10710.

Butler, L., and Willshaw, H.E. (1989). Ocular and facial maldevelopment: the role of neural crest. Eye *3 (Pt 1)*, 64-68.
Camenisch, T.D., Molin, D.G., Person, A., Runyan, R.B., Gittenberger-de Groot, A.C., McDonald, J.A., and Klewer, S.E. (2002). Temporal and distinct TGFbeta ligand requirements during mouse and avian endocardial cushion morphogenesis. Dev Biol *248*, 170-181.
Cao, Q.P., Gaudette, M.F., Robinson, D.H., and Crain, W.R. (1995). Expression of the mouse testis-determining gene Sry in male preimplantation embryos. Mol Reprod Dev *40*, 196-204.
Capel, B. (1995). New bedfellows in the mammalian sex-determination affair. Trends Genet *11*, 161-163.
Castillo, S.D., and Sanchez-Cespedes, M. (2012). The SOX family of genes in cancer development: biological relevance and opportunities for therapy. Expert Opin Ther Targets *16*, 903-919.
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.
Chang, Z.J., Du, Q.Y., and Shao, H.W. (2002). [New advances in Sox gene family]. Yi Chuan *24*, 470-476.
Cheng, Y., Cheung, M., Abu-Elmagd, M.M., Orme, A., and Scotting, P.J. (2000). Chick sox10, a transcription factor expressed in both early neural crest cells and central nervous system. Brain Res Dev Brain Res *121*, 233-241.
Cheung, M., and Briscoe, J. (2003). Neural crest development is regulated by the transcription factor Sox9. Development *130*, 5681-5693.
Christiansen, A.E., Keisman, E.L., Ahmad, S.M., and Baker, B.S. (2002). Sex comes in from the cold: the integration of sex and pattern. Trends in genetics : TIG *18*, 510-516.
Cimadamore, F., Shah, M., Amador-Arjona, A., Peran, E.M., Chen, C., and Terskikh, A.V. (2012). SOX2 Modulates Levels of MITF in Human Embryonic Stem Cell - Derived Neural Crest, Normal Human Melanocytes, and Melanoma Lines In Vitro. Pigment Cell Melanoma Res.
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.
Cook, A.L., Smith, A.G., Smit, D.J., Leonard, J.H., and Sturm, R.A. (2005). Co-expression of SOX9 and SOX10 during melanocytic differentiation in vitro. Exp Cell Res *308*, 222-235.
Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., and Helms, J.A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. Am J Med Genet A *155A*, 270-279.
Coriat, A.M., Muller, U., Harry, J.L., Uwanogho, D., and Sharpe, P.T. (1993). PCR amplification of SRY-related gene sequences reveals evolutionary conservation of the SRY-box motif. PCR Methods Appl *2*, 218-222.
Cox, J.L., Wilder, P.J., Desler, M., and Rizzino, A. (2012). Elevating SOX2 Levels Deleteriously Affects the Growth of Medulloblastoma and Glioblastoma Cells. PLoS One *7*, e44087.
Creuzet, S.E. (2009). Regulation of pre-otic brain development by the cephalic neural crest. Proc Natl Acad Sci U S A *106*, 15774-15779.

Cunningham, D.D., Meng, Z., Fritzsch, B., and Casey, E.S. (2008). Cloning and developmental expression of the soxB2 genes, sox14 and sox21, during Xenopus laevis embryogenesis. Int J Dev Biol *52*, 999-1004.
Dabdoub, A., Puligilla, C., Jones, J.M., Fritzsch, B., Cheah, K.S., Pevny, L.H., and Kelley, M.W. (2008). Sox2 signaling in prosensory domain specification and subsequent hair cell differentiation in the developing cochlea. Proc Natl Acad Sci U S A *105*, 18396-18401.
DeFina, P. (2008). Meeting the SOX challenge. Occup Health Saf *77*, 52-53.
Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P., and Ashworth, A. (1992). A conserved family of genes related to the testis determining gene, SRY. Nucleic Acids Res *20*, 2887.
Dickinson, M.E., Selleck, M.A., McMahon, A.P., and Bronner-Fraser, M. (1995). Dorsalization of the neural tube by the non-neural ectoderm. Development *121*, 2099-2106.
Dong, C., Wilhelm, D., and Koopman, P. (2004). Sox genes and cancer. Cytogenet Genome Res *105*, 442-447.
Donner, A.L., Episkopou, V., and Maas, R.L. (2007). Sox2 and Pou2f1 interact to control lens and olfactory placode development. Dev Biol *303*, 784-799.
Dottori, M., Gross, M.K., Labosky, P., and Goulding, M. (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. Development *128*, 4127-4138.
Drysdale, T.A., Patterson, K.D., Saha, M., and Krieg, P.A. (1997). Retinoic acid can block differentiation of the myocardium after heart specification. Dev Biol *188*, 205-215.
Durocher, D., Charron, F., Warren, R., Schwartz, R.J., and Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. Embo J *16*, 5687-5696.
Economides, A.N., Frendewey, D., Yang, P., Dominguez, M.G., Dore, A.T., Lobov, I.B., Persaud, T., Rojas, J., McClain, J., Lengyel, P.*, et al.* (2012). Conditionals-by-Inversion: a novel universal method for the generation of conditional alleles. in preparation.
Ekonomou, A., Kazanis, I., Malas, S., Wood, H., Alifragis, P., Denaxa, M., Karagogeos, D., Constanti, A., Lovell-Badge, R., and Episkopou, V. (2005). Neuronal migration and ventral subtype identity in the telencephalon depend on SOX1. PLoS biology *3*, e186.
Elkouris, M., Balaskas, N., Poulou, M., Politis, P.K., Panayiotou, E., Malas, S., Thomaidou, D., and Remboutsika, E. (2011). Sox1 maintains the undifferentiated state of cortical neural progenitor cells via the suppression of Prox1-mediated cell cycle exit and neurogenesis. Stem Cells *29*, 89-98.
Ellis, P., Fagan, B.M., Magness, S.T., Hutton, S., Taranova, O., Hayashi, S., McMahon, A., Rao, M., and Pevny, L. (2004). SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. Dev Neurosci *26*, 148-165.
Endo, T., Nakagawa, T., Lee, J.E., Dong, Y., Kim, T.S., Iguchi, F., Taniguchi, Z., Naito, Y., and Ito, J. (2002). Alteration in expression of p27 in auditory epithelia and neurons of mice during degeneration. Neuroscience letters *334*, 173-176.
Engelen, E., Akinci, U., Bryne, J.C., Hou, J., Gontan, C., Moen, M., Szumska, D., Kockx, C., van Ijcken, W., Dekkers, D.H.*, et al.* (2011). Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat Genet *43*, 607-611.
Fantes, J., Ragge, N.K., Lynch, S.A., McGill, N.I., Collin, J.R., Howard-Peebles, P.N., Hayward, C., Vivian, A.J., Williamson, K., van Heyningen, V.*, et al.* (2003). Mutations in SOX2 cause anophthalmia. Nature genetics *33*, 461-463.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet *4*, e1000116.
Favaro, R., Valotta, M., Ferri, A.L., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V.*, et al.* (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nat Neurosci *12*, 1248-1256.
Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R., and Bianchi, M.E. (1992). SRY, like HMG1, recognizes sharp angles in DNA. EMBO J *11*, 4497-4506.
Ferri, A.L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S.*, et al.* (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development *131*, 3805-3819.
Finzsch, M., Stolt, C.C., Lommes, P., and Wegner, M. (2008). Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor alpha expression. Development *135*, 637-646.
Foster, J.W., Dominguez-Steglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I.D., Goodfellow, P.N.*, et al.* (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature *372*, 525-530.
Friend, K., Lovejoy, A.F., and Steitz, J.A. (2007). U2 snRNP binds intronless histone pre-mRNAs to facilitate U7-snRNP-dependent 3' end formation. Mol Cell *28*, 240-252.
Funabashi, H., Takatsu, M., Saito, M., and Matsuoka, H. (2010). Sox2 regulatory region 2 sequence works as a DNA nuclear targeting sequence enhancing the efficiency of an exogenous gene expression in ES cells. Biochem Biophys Res Commun *400*, 554-558.
Gammill, L.S., and Bronner-Fraser, M. (2002). Genomic analysis of neural crest induction. Development *129*, 5731-5741.
Garcia-Castro, M.I., Marcelle, C., and Bronner-Fraser, M. (2002). Ectodermal Wnt function as a neural crest inducer. Science *297*, 848-851.
Gentles, A.J., and Karlin, S. (1999). Why are human G-protein-coupled receptors predominantly intronless? Trends in genetics : TIG *15*, 47-49.
Geraerts, M., and Verfaillie, C.M. (2009). Adult stem and progenitor cells. Adv Biochem Eng Biotechnol *114*, 1-21.
Germana, A., Montalbano, G., Guerrera, M.C., Laura, R., Levanti, M., Abbate, F., de Carlos, F., Vega, J.A., and Ciriaco, E. (2009). Sox-2 in taste bud and lateral line system of zebrafish during development. Neurosci Lett *467*, 36-39.
Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell *69*, 185-195.
Gnjatic, S., Altorki, N.K., Tang, D.N., Tu, S.M., Kundra, V., Ritter, G., Old, L.J., Logothetis, C.J., and Sharma, P. (2009). NY-ESO-1 DNA vaccine induces T-cell responses that are suppressed by regulatory T cells. Clinical cancer research : an official journal of the American Association for Cancer Research *15*, 2130-2139.
Goodfellow, P.N., and Lovell-Badge, R. (1993). SRY and sex determination in mammals. Annu Rev Genet *27*, 71-92.
Gould, J.S. (2003). What it's like to be a Red Sox fan. Am J Orthop (Belle Mead NJ) *32*, 528.
Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. Neuron *39*, 749-765.

Griffiths-Jones, S. (2006). miRBase: the microRNA sequence database. Methods in molecular biology *342*, 129-138.
Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res *34*, D140-144.
Gu, G.M., and Wang, J.K. (2012). [DNA-binding profiles of mammalian transcription factors]. Yi Chuan *34*, 950-968.
Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature *346*, 245-250.
Guth, S., and Wegner, M. (2008). Having it both ways: Sox protein function between conservation and innovation. Cellular and Molecular Life Sciences *65*, 3000-3018.
Hacker, A., Capel, B., Goodfellow, P., and Lovell-Badge, R. (1995). Expression of Sry, the mouse sex determining gene. Development *121*, 1603-1614.
Han, X., Fang, X., Lou, X., Hua, D., Ding, W., Foltz, G., Hood, L., Yuan, Y., and Lin, B. (2012). Silencing SOX2 induced mesenchymal-epithelial transition and its expression predicts liver and lymph node metastasis of CRC patients. PLoS One *7*, e41335.
Harley, V., and Lefebvre, V. (2010). Twenty Sox, twenty years. Int J Biochem Cell Biol *42*, 376-377.
Harris, M.L., Baxter, L.L., Loftus, S.K., and Pavan, W.J. (2010). Sox proteins in melanocyte development and melanoma. Pigment Cell Melanoma Res *23*, 496-513.
Harvey, R.P. (1996). NK-2 homeobox genes and heart development. Dev Biol *178*, 203-216.
Hashimoto, S., Chen, H., Que, J., Brockway, B.L., Drake, J.A., Snyder, J.C., Randell, S.H., and Stripp, B.R. (2012). beta-Catenin-SOX2 signaling regulates the fate of developing airway epithelium. J Cell Sci *125*, 932-942.
Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002a). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mech Dev *119 Suppl 1*, S97-S101.
Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002b). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Gene Expr Patterns *2*, 93-97.
Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002c). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mech Dev *119 Suppl 1*, S97-S101.
Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002d). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Gene Expr Patterns *2*, 93-97.
Hayashi, S., Tenzen, T., and McMahon, A.P. (2003). Maternal inheritance of Cre activity in a Sox2Cre deleter strain. Genesis *37*, 51-53.
Hever, A.M., Williamson, K.A., and van Heyningen, V. (2006). Developmental malformations of the eye: the role of PAX6, SOX2 and OTX2. Clin Genet *69*, 459-470.
Hill, A.E., and Sorscher, E.J. (2006). The non-random distribution of intronless human genes across molecular function categories. FEBS Lett *580*, 4303-4305.
Hoke, A. (2012). Animal models of peripheral neuropathies. Neurotherapeutics *9*, 262-269.
Huang, Y., and Carmichael, G.G. (1997). The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation of intronless gene transcripts and of unspliced HIV-1-related mRNAs. Proc Natl Acad Sci U S A *94*, 10104-10109.
Hutton, S.R., and Pevny, L.H. (2011). SOX2 expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. Dev Biol *352*, 40-47.

Iwafuchi-Doi, M., Yoshida, Y., Onichtchouk, D., Leichsenring, M., Driever, W., Takemoto, T., Uchikawa, M., Kamachi, Y., and Kondoh, H. (2011). The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates. Dev Biol *352*, 354-366.
Jain, S., Knoten, A., Hoshi, M., Wang, H., Vohra, B., Heuckeroth, R.O., and Milbrandt, J. (2010). Organotypic specificity of key RET adaptor-docking sites in the pathogenesis of neurocristopathies and renal malformations in mice. J Clin Invest *120*, 778-790.
Jantzen, H.M., Admon, A., Bell, S.P., and Tjian, R. (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature *344*, 830-836.
Jessel, A.S., and Weiss, P.L. (1998). Orthodox practices clarified. Am J Occup Ther *52*, 71-72.
Kamachi, Y., Iwafuchi, M., Okuda, Y., Takemoto, T., Uchikawa, M., and Kondoh, H. (2009). Evolution of non-coding regulatory sequences involved in the developmental process: reflection of differential employment of paralogous genes as highlighted by Sox2 and group B1 Sox genes. Proceedings of the Japan Academy Series B, Physical and biological sciences *85*, 55-68.
Kamachi, Y., Uchikawa, M., and Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet *16*, 182-187.
Kang, P., and Svoboda, K.K. (2005). Epithelial-mesenchymal transformation during craniofacial development. J Dent Res *84*, 678-690.
Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. EMBO J *7*, 1537-1547.
Kelly, R.G. (2012). The second heart field. Current topics in developmental biology *100*, 33-65.
Kelsh, R.N. (2006). Sorting out Sox10 functions in neural crest development. Bioessays *28*, 788-798.
Kiefer, J.C. (2007). Back to basics: Sox genes. Dev Dyn *236*, 2356-2366.
Kim, E.Y., Chen, L., Ma, Y., Yu, W., Chang, J., Moskowitz, I.P., and Wang, J. (2011). Expression of sumoylation deficient Nkx2.5 mutant in Nkx2.5 haploinsufficient mice leads to congenital heart defects. PLoS One *6*, e20803.
Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.
Kirby, P.J., Waters, P.D., Delbridge, M., Svartman, M., Stewart, A.N., Nagai, K., and Graves, J.A. (2002). Cloning and mapping of platypus SOX2 and SOX14: insights into SOX group B evolution. Cytogenetic and genome research *98*, 96-100.
Koopman, P. (2005). Sex determination: a tale of two Sox genes. Trends Genet *21*, 367-370.
Koopman, P., Schepers, G., Brenner, S., and Venkatesh, B. (2004). Origin and diversity of the SOX transcription factor gene family: genome-wide analysis in Fugu rubripes. Gene *328*, 177-186.
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.
Kouskoura, T., Fragou, N., Alexiou, M., John, N., Sommer, L., Graf, D., Katsaros, C., and Mitsiadis, T.A. (2011). The genetic basis of craniofacial and dental abnormalities. Schweizer Monatsschrift fur Zahnmedizin = Revue mensuelle suisse d'odonto-stomatologie = Rivista mensile svizzera di odontologia e stomatologia / SSO *121*, 636-646.

Koyanagi, M., Iwasaki, M., Rupp, S., Tedesco, F.S., Yoon, C.H., Boeckel, J.N., Trauth, J., Schutz, C., Ohtani, K., Goetz, R.*, et al.* (2010). Sox2 transduction enhances cardiovascular repair capacity of blood-derived mesoangioblasts. Circ Res *106*, 1290-1302.
Kurata, T., and Ueno, N. (2003). Xenopus Nbx, a novel NK-1 related gene essential for neural crest formation. Dev Biol *257*, 30-40.
Kuzmichev, A.N., Kim, S.K., D'Alessio, A.C., Chenoweth, J.G., Wittko, I.M., Campanati, L., and McKay, R.D. (2012). Sox2 Acts through Sox21 to Regulate Transcription in Pluripotent and Differentiated Cells. Curr Biol *22*, 1705-1710.
Kwok, C., Weller, P.A., Guioli, S., Foster, J.W., Mansour, S., Zuffardi, O., Punnett, H.H., Dominguez-Steglich, M.A., Brook, J.D., Young, I.D.*, et al.* (1995). Mutations in SOX9, the gene responsible for Campomelic dysplasia and autosomal sex reversal. Am J Hum Genet *57*, 1028-1036.
LaBonne, C., and Bronner-Fraser, M. (1998). Induction and patterning of the neural crest, a stem cell-like precursor population. J Neurobiol *36*, 175-189.
Laga, A.C., Lai, C.Y., Zhan, Q., Huang, S.J., Velazquez, E.F., Yang, Q., Hsu, M.Y., and Murphy, G.F. (2010). Expression of the embryonic stem cell transcription factor SOX2 in human skin: relevance to melanocyte and merkel cell biology. Am J Pathol *176*, 903-913.
Laudet, V., Stehelin, D., and Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res *21*, 2493-2501.
Le Douarin, N.M., Couly, G., and Creuzet, S.E. (2012). The neural crest is a powerful regulator of pre-otic brain development. Dev Biol *366*, 74-82.
Le, N., Nagarajan, R., Wang, J.Y., Araki, T., Schmidt, R.E., and Milbrandt, J. (2005). Analysis of congenital hypomyelinating Egr2Lo/Lo nerves identifies Sox2 as an inhibitor of Schwann cell differentiation and myelination. Proc Natl Acad Sci U S A *102*, 2596-2601.
Lee, Y.H., and Saint-Jeannet, J.P. (2011). Sox9 function in craniofacial development and disease. Genesis *49*, 200-208.
Liem, K.F., Jr., Tremml, G., Roelink, H., and Jessell, T.M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell *82*, 969-979.
Liu, J.P., and Jessell, T.M. (1998). A role for rhoB in the delamination of neural crest cells from the dorsal neural tube. Development *125*, 5055-5067.
Liu, Y., and Xiao, A. (2011). Epigenetic regulation in neural crest development. Birth Defects Res A Clin Mol Teratol *91*, 788-796.
Lovell-Badge, R. (2010). The early history of the Sox genes. Int J Biochem Cell Biol *42*, 378-380.
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.
Lupski, J.R. (2010). Interruption of SOX10 function in myelinopathies. Annals of neurology *68*, 121-123.
Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev *9*, 1654-1666.
Machado, I., Lopez-Guerrero, J.A., Navarro, S., Alberghini, M., Scotlandi, K., Picci, P., and Llombart-Bosch, A. (2012). Epithelial cell adhesion molecules and epithelial mesenchymal transition (EMT) markers in Ewing's sarcoma family of tumors (ESFTs). Do they offer any prognostic significance? Virchows Arch *461*, 333-337.

Malki, S., Boizet-Bonhoure, B., and Poulat, F. (2010). Shuttling of SOX proteins. Int J Biochem Cell Biol *42*, 411-416.
Mallanna, S.K., Ormsbee, B.D., Iacovino, M., Gilmore, J.M., Cox, J.L., Kyba, M., Washburn, M.P., and Rizzino, A. (2010). Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. Stem Cells *28*, 1715-1727.
Mandalos, N., Saridaki, M., Harper, J.L., Kotsoni, A., Yang, P., Economides, A.N., and Remboutsika, E. (2012). Application of a novel strategy of engineering conditional alleles to a single exon gene, sox2. PLoS One *7*, e45768.
Maquat, L.E., and Li, X. (2001). Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. Rna *7*, 445-456.
Mason, I. (1996). Neural induction: do fibroblast growth factors strike a cord? Curr Biol *6*, 672-675.
McKeown, S.J., Lee, V.M., Bronner-Fraser, M., Newgreen, D.F., and Farlie, P.G. (2005). Sox10 overexpression induces neural crest-like cells from all dorsoventral levels of the neural tube but inhibits differentiation. Dev Dyn *233*, 430-444.
Medlin, J., Scurry, A., Taylor, A., Zhang, F., Peterlin, B.M., and Murphy, S. (2005). P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. The EMBO journal *24*, 4154-4165.
Mercado-Pimentel, M.E., and Runyan, R.B. (2007). Multiple transforming growth factor-beta isoforms and receptors function during epithelial-mesenchymal cell transformation in the embryonic heart. Cells Tissues Organs *185*, 146-156.
Minoux, M. (2011). Role of hoxa2 in mouse external ear morphogenesis: a model to decipher human craniofacial genetic disorders. Bull Group Int Rech Sci Stomatol Odontol *50*, 25-26.
Miyagi, S., Masui, S., Niwa, H., Saito, T., Shimazaki, T., Okano, H., Nishimoto, M., Muramatsu, M., Iwama, A., and Okuda, A. (2008). Consequence of the loss of Sox2 in the developing brain of the mouse. FEBS Lett *582*, 2811-2815.
Miyagi, S., Nishimoto, M., Saito, T., Ninomiya, M., Sawamoto, K., Okano, H., Muramatsu, M., Oguro, H., Iwama, A., and Okuda, A. (2006). The Sox2 regulatory region 2 functions as a neural stem cell-specific enhancer in the telencephalon. J Biol Chem *281*, 13374-13381.
Miyagi, S., Saito, T., Mizutani, K., Masuyama, N., Gotoh, Y., Iwama, A., Nakauchi, H., Masui, S., Niwa, H., Nishimoto, M.*, et al.* (2004). The Sox-2 regulatory regions display their activities in two distinct types of multipotent stem cells. Mol Cell Biol *24*, 4207-4220.
Mollaaghababa, R., and Pavan, W.J. (2003). The importance of having your SOX on: role of SOX10 in the development of neural crest-derived melanocytes and glia. Oncogene *22*, 3024-3034.
Monier, R. (1995). [Cell transformation]. Rev Prat *45*, 1867-1872.
Nagai, K. (2001). Molecular evolution of Sry and Sox gene. Gene *270*, 161-169.
Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. Genesis *26*, 99-109.
Nakagawa, S., and Takeichi, M. (1995). Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific expression of novel cadherins. Development *121*, 1321-1332.

Nakayama, T., and Inoue, N. (2006). Neural stem sphere method: induction of neural stem cells and neurons by astrocyte-derived factors in embryonic stem cells in vitro. Methods Mol Biol *330*, 1-13.
Nambu, P.A., and Nambu, J.R. (1996). The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. Development *122*, 3467-3475.
Ner, S.S. (1992). HMGs everywhere. Curr Biol *2*, 208-210.
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.
Niwa, H. (2000). [Molecular mechanism for cell-fate determination in ES cells]. Tanpakushitsu Kakusan Koso *45*, 2047-2055.
O'Brien, E.K., and Degnan, B.M. (2000). Expression of POU, Sox, and Pax genes in the brain ganglia of the tropical abalone Haliotis asinina. Mar Biotechnol (NY) *2*, 545-557.
Ogura, T., and Evans, R.M. (1995). A retinoic acid-triggered cascade of HOXB1 gene activation. Proc Natl Acad Sci U S A *92*, 387-391.
Okubo, T., Pevny, L.H., and Hogan, B.L. (2006). Sox2 is required for development of taste bud sensory cells. Genes Dev *20*, 2654-2659.
Pan, H., and Schultz, R.M. (2011). Sox2 modulates reprogramming of gene expression in two-cell mouse embryos. Biol Reprod *85*, 409-416.
Panetta, N.J., Gupta, D.M., Slater, B.J., Kwan, M.D., Liu, K.J., and Longaker, M.T. (2008). Tissue engineering in cleft palate and other congenital malformations. Pediatric research *63*, 545-551.
Papanayotou, C., Mey, A., Birot, A.M., Saka, Y., Boast, S., Smith, J.C., Samarut, J., and Stern, C.D. (2008). A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. PLoS Biol *6*, e2.
Parrinello, S., Napoli, I., Ribeiro, S., Digby, P.W., Fedorova, M., Parkinson, D.B., Doddrell, R.D., Nakayama, M., Adams, R.H., and Lloyd, A.C. (2010). EphB signaling directs peripheral nerve regeneration through Sox2-dependent Schwann cell sorting. Cell *143*, 145-155.
Parthey, K., Kornhuber, M., Kunze, C., Wand, D., Nolte, K.W., Nikolin, S., Weis, J., and Schroder, J.M. (2012). SOX10 mutation with peripheral amyelination and developmental disturbance of axons. Muscle & nerve *45*, 284-290.
Pata, I., Studer, M., van Doorninck, J.H., Briscoe, J., Kuuse, S., Engel, J.D., Grosveld, F., and Karis, A. (1999). The transcription factor GATA3 is a downstream effector of Hoxb1 specification in rhombomere 4. Development *126*, 5523-5531.
Person, A.D., Klewer, S.E., and Runyan, R.B. (2005). Cell biology of cardiac cushion development. International review of cytology *243*, 287-335.
Pevny, L., and Placzek, M. (2005). SOX genes and neural progenitor identity. Curr Opin Neurobiol *15*, 7-13.
Pevny, L.H., and Lovell-Badge, R. (1997). Sox genes find their feet. Curr Opin Genet Dev *7*, 338-344.
Pozniak, C.D., Langseth, A.J., Dijkgraaf, G.J., Choe, Y., Werb, Z., and Pleasure, S.J. (2010). Sox10 directs neural stem cells toward the oligodendrocyte lineage by decreasing Suppressor of Fused expression. Proc Natl Acad Sci U S A *107*, 21795-21800.

Prince, V., and Lumsden, A. (1994). Hoxa-2 expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest. Development *120*, 911-923.
Puc, J., and Rosenfeld, M.G. (2011). SOX2 and CHD7 cooperatively regulate human disease genes. Nat Genet *43*, 505-506.
Puligilla, C., Dabdoub, A., Brenowitz, S.D., and Kelley, M.W. (2010). Sox2 induces neuronal formation in the developing mammalian cochlea. J Neurosci *30*, 714-722.
Que, J., Luo, X., Schwartz, R.J., and Hogan, B.L. (2009). Multiple roles for Sox2 in the developing and adult mouse trachea. Development *136*, 1899-1907.
Que, J., Okubo, T., Goldenring, J.R., Nam, K.T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development *134*, 2521-2531.
Redies, C., Hertel, N., and Hubner, C.A. (2012). Cadherins and neuropsychiatric disorders. Brain Res *1470*, 130-144.
Remboutsika, E., Elkouris, M., Iulianella, A., Andoniadou, C.L., Poulou, M., Mitsiadis, T.A., Trainor, P.A., and Lovell-Badge, R. (2011a). Flexibility of neural stem cells. Front Physiol *2*, 16.
Remboutsika, E., Elkouris, M., Iulianella, A., Andoniadou, C.L., Poulou, M., Mitsiadis, T.A., Trainor, P.A., and Lovell-Badge, R. (2011b). Flexibility of neural stem cells. Front Physiol *2*, 16.
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.
Rijli, F.M., Gavalas, A., and Chambon, P. (1998). Segmentation and specification in the branchial region of the head: the role of the Hox selector genes. Int J Dev Biol *42*, 393-401.
Ringrose, L., Chabanis, S., Angrand, P.O., Woodroofe, C., and Stewart, A.F. (1999). Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. Embo J *18*, 6630-6641.
Risebro, C.A., Searles, R.G., Melville, A.A., Ehler, E., Jina, N., Shah, S., Pallas, J., Hubank, M., Dillard, M., Harvey, N.L.*, et al.* (2009). Prox1 maintains muscle structure and growth in the developing heart. Development *136*, 495-505.
Rogers, C.D., Moody, S.A., and Casey, E.S. (2009). Neural induction and factors that stabilize a neural fate. Birth Defects Res C Embryo Today *87*, 249-262.
Rossant, J. (1996). Mouse mutants and cardiac development: new molecular insights into cardiogenesis. Circ Res *78*, 349-353.
Russell, S.R., Sanchez-Soriano, N., Wright, C.R., and Ashburner, M. (1996). The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. Development *122*, 3669-3676.
Sakharkar, K.R., Sakharkar, M.K., Culiat, C.T., Chow, V.T., and Pervaiz, S. (2006). Functional and evolutionary analyses on expressed intronless genes in the mouse genome. FEBS Lett *580*, 1472-1478.
Sakharkar, M.K., Perumal, B.S., Lim, Y.P., Chern, L.P., Yu, Y., and Kangueane, P. (2005). Alternatively spliced human genes by exon skipping--a database (ASHESdb). In silico biology *5*, 221-225.
Sandberg, M., Kallstrom, M., and Muhr, J. (2005). Sox21 promotes the progression of vertebrate neurogenesis. Nat Neurosci *8*, 995-1001.
Santagati, F., Minoux, M., Ren, S.Y., and Rijli, F.M. (2005). Temporal requirement of Hoxa2 in cranial neural crest skeletal morphogenesis. Development *132*, 4927-4936.

Sarnat, H.B., and Flores-Sarnat, L. (2005). Embryology of the neural crest: its inductive role in the neurocutaneous syndromes. J Child Neurol *20*, 637-643.
Sasaki, H., Yokota, K., Hikosaka, Y., Moriyama, S., Yano, M., and Fujii, Y. (2012). Increased Sox2 copy number in lung squamous cell carcinomas. Exp Ther Med *3*, 44-48.
Schepers, G.E., Teasdale, R.D., and Koopman, P. (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell *3*, 167-170.
Schilham, M.W., Oosterwegel, M.A., Moerer, P., Ya, J., de Boer, P.A., van de Wetering, M., Verbeek, S., Lamers, W.H., Kruisbeek, A.M., Cumano, A.*, et al.* (1996). Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature *380*, 711-714.
Schilham, M.W., van Eijk, M., van de Wetering, M., and Clevers, H.C. (1993). The murine Sox-4 protein is encoded on a single exon. Nucleic acids research *21*, 2009.
Scott, C.E., Wynn, S.L., Sesay, A., Cruz, C., Cheung, M., Gomez Gaviro, M.V., Booth, S., Gao, B., Cheah, K.S., Lovell-Badge, R.*, et al.* (2010). SOX9 induces and maintains neural stem cells. Nat Neurosci *13*, 1181-1189.
Scotting, P.J., and Rex, M. (1996). Transcription factors in early development of the central nervous system. Neuropathol Appl Neurobiol *22*, 469-481.
Selleck, M.A., and Bronner-Fraser, M. (1995). Origins of the avian neural crest: the role of neural plate-epidermal interactions. Development *121*, 525-538.
Seoighe, C., and Korir, P.K. (2011). Evidence for intron length conservation in a set of mammalian genes associated with embryonic development. BMC Bioinformatics *12 Suppl 9*, S16.
Serbedzija, G.N., Fraser, S.E., and Bronner-Fraser, M. (1990). Pathways of trunk neural crest cell migration in the mouse embryo as revealed by vital dye labelling. Development *108*, 605-612.
Shevaga, V.N. (1981). [Mineral homeostasis in brain edema]. Vrachebnoe delo, 95-99.
Sikorska, M., Sandhu, J.K., Deb-Rinker, P., Jezierski, A., Leblanc, J., Charlebois, C., Ribecco-Lutkiewicz, M., Bani-Yaghoub, M., and Walker, P.R. (2008). Epigenetic modifications of SOX2 enhancers, SRR1 and SRR2, correlate with in vitro neural differentiation. J Neurosci Res *86*, 1680-1693.
Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature *346*, 240-244.
Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T.*, et al.* (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337-342.
Smith, A.N., Miller, L.A., Radice, G., Ashery-Padan, R., and Lang, R.A. (2009). Stage-dependent modes of Pax6-Sox2 epistasis regulate lens development and eye morphogenesis. Development *136*, 2977-2985.
Srinivasan, R.S., Dillard, M.E., Lagutin, O.V., Lin, F.J., Tsai, S., Tsai, M.J., Samokhvalov, I.M., and Oliver, G. (2007). Lineage tracing demonstrates the venous origin of the mammalian lymphatic vasculature. Genes Dev *21*, 2422-2432.

Stallmeyer, B., Fenge, H., Nowak-Gottl, U., and Schulze-Bahr, E. (2010). Mutational spectrum in the cardiac transcription factor gene NKX2.5 (CSX) associated with congenital heart disease. Clin Genet *78*, 533-540.
Stefanovic, S., Abboud, N., Desilets, S., Nury, D., Cowan, C., and Puceat, M. (2009). Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. J Cell Biol *186*, 665-673.
Stevanovic, M., Lovell-Badge, R., Collignon, J., and Goodfellow, P.N. (1993). SOX3 is an X-linked gene related to SRY. Hum Mol Genet *2*, 2013-2018.
Stevanovic, M., Zuffardi, O., Collignon, J., Lovell-Badge, R., and Goodfellow, P. (1994). The cDNA sequence and chromosomal location of the human SOX2 gene. Mamm Genome *5*, 640-642.
Stolt, C.C., Lommes, P., Hillgartner, S., and Wegner, M. (2008). The transcription factor Sox5 modulates Sox10 function during melanocyte development. Nucleic Acids Res *36*, 5427-5440.
Streit, A., Sockanathan, S., Perez, L., Rex, M., Scotting, P.J., Sharpe, P.T., Lovell-Badge, R., and Stern, C.D. (1997). Preventing the loss of competence for neural induction: HGF/SF, L5 and Sox-2. Development *124*, 1191-1202.
Swinburne, I.A., Miguez, D.G., Landgraf, D., and Silver, P.A. (2008). Intron length increases oscillatory periods of gene expression in animal cells. Genes Dev *22*, 2342-2346.
Swinburne, I.A., and Silver, P.A. (2008). Intron delays and transcriptional timing during development. Dev Cell *14*, 324-330.
Takada, N., Kucenas, S., and Appel, B. (2010). Sox10 is necessary for oligodendrocyte survival following axon wrapping. Glia *58*, 996-1006.
Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N., and Izumo, S. (1999). The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. Development *126*, 1269-1280.
Tanaka, S.S., Kojima, Y., Yamaguchi, Y.L., Nishinakamura, R., and Tam, P.P. (2011). Impact of WNT signaling on tissue lineage differentiation in the early mouse embryo. Dev Growth Differ *53*, 843-856.
Taranova, O.V., Magness, S.T., Fagan, B.M., Wu, Y., Surzenko, N., Hutton, S.R., and Pevny, L.H. (2006). SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev *20*, 1187-1202.
Targoff, K.L., Schell, T., and Yelon, D. (2008). Nkx genes regulate heart tube extension and exert differential effects on ventricular and atrial cell number. Dev Biol *322*, 314-321.
Testa, G., Schaft, J., van der Hoeven, F., Glaser, S., Anastassiadis, K., Zhang, Y., Hermann, T., Stremmel, W., and Stewart, A.F. (2004). A reliable lacZ expression reporter cassette for multipurpose, knockout-first alleles. Genesis *38*, 151-158.
Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res *30*, 3202-3213.
Trainor, P.A., and Krumlauf, R. (2001). Hox genes, neural crest cells and branchial arch patterning. Curr Opin Cell Biol *13*, 698-705.
Tucker, P.K., and Lundrigan, B.L. (1993). Rapid evolution of the sex determining locus in Old World mice and rats. Nature *364*, 715-717.
Uchikawa, M., Yoshida, M., Iwafuchi-Doi, M., Matsuda, K., Ishida, Y., Takemoto, T., and Kondoh, H. (2011). B1 and B2 Sox gene expression during neural plate development in

chicken and mouse embryos: universal versus species-dependent features. Dev Growth Differ *53*, 761-771.
Uwanogho, D., Rex, M., Cartwright, E.J., Pearl, G., Healy, C., Scotting, P.J., and Sharpe, P.T. (1995a). Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. Mech Dev *49*, 23-36.
Uwanogho, D., Rex, M., Cartwright, E.J., Pearl, G., Healy, C., Scotting, P.J., and Sharpe, P.T. (1995b). Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. Mech Dev *49*, 23-36.
Uy, B.R., Simoes-Costa, M., Sauka-Spengler, T., and Bronner, M.E. (2012). Expression of Sox family genes in early lamprey development. Int J Dev Biol *56*, 377-383.
Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J.*, et al.* (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nature biotechnology *21*, 652-659.
Vallaster, M., Vallaster, C.D., and Wu, S.M. (2012). Epigenetic mechanisms in cardiac development and disease. Acta Biochim Biophys Sin (Shanghai) *44*, 92-102.
van Bakel, H. (2011). Interactions of transcription factors with chromatin. Subcell Biochem *52*, 223-259.
van den Berg, G., Abu-Issa, R., de Boer, B.A., Hutson, M.R., de Boer, P.A., Soufan, A.T., Ruijter, J.M., Kirby, M.L., van den Hoff, M.J., and Moorman, A.F. (2009). A caudal proliferating growth center contributes to both poles of the forming heart tube. Circ Res *104*, 179-188.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.
Vieille-Grosjean, I., Hunt, P., Gulisano, M., Boncinelli, E., and Thorogood, P. (1997). Branchial HOX gene expression and human craniofacial development. Dev Biol *183*, 49-60.
Villanueva, S., Glavic, A., Ruiz, P., and Mayor, R. (2002). Posteriorization by FGF, Wnt, and retinoic acid is required for neural crest induction. Developmental biology *241*, 289-301.
Vincent, S.D., and Robertson, E.J. (2003). Highly efficient transgene-independent recombination directed by a maternally derived SOX2CRE transgene. Genesis *37*, 54-56.
Vitelli, F., Morishima, M., Taddei, I., Lindsay, E.A., and Baldini, A. (2002). Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. Hum Mol Genet *11*, 915-922.
von Gise, A., and Pu, W.T. (2012). Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. Circ Res *110*, 1628-1645.
Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E.*, et al.* (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell *79*, 1111-1120.
Wakamatsu, Y. (2004). Understanding glial differentiation in vertebrate nervous system development. The Tohoku journal of experimental medicine *203*, 233-240.
Wakamatsu, Y., Endo, Y., Osumi, N., and Weston, J.A. (2004). Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. Dev Dyn *229*, 74-86.
Wang, J., Liu, X.Y., and Yang, Y.Q. (2011). Novel NKX2-5 mutations responsible for congenital heart disease. Genetics and molecular research : GMR *10*, 2905-2915.
Wang, R., Cheng, H.H., Guo, Y.Q., and Zhou, R.J. (2002). [Phylogenic analysis of the Sox gene family of vertebrate]. Yi Chuan Xue Bao *29*, 990-994.

Warren, S.A., Terada, R., Briggs, L.E., Cole-Jeffrey, C.T., Chien, W.M., Seki, T., Weinberg, E.O., Yang, T.P., Chin, M.T., Bungert, J.*, et al.* (2011). Differential role of Nkx2-5 in activation of the atrial natriuretic factor gene in the developing versus failing heart. Mol Cell Biol *31*, 4633-4645.
Waterman, M.L., and Jones, K.A. (1990). Purification of TCF-1 alpha, a T-cell-specific transcription factor that activates the T-cell receptor C alpha gene enhancer in a context-dependent manner. New Biol *2*, 621-636.
Wegner, M. (2011). SOX after SOX: SOXession regulates neurogenesis. Genes Dev *25*, 2423-2428.
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.
Weiss, M.A. (2001). Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition. Mol Endocrinol *15*, 353-362.
Wen, J., Hu, Q., Li, M., Wang, S., Zhang, L., Chen, Y., and Li, L. (2008). Pax6 directly modulate Sox2 expression in the neural progenitor cells. Neuroreport *19*, 413-417.
Whitfield, L.S., Lovell-Badge, R., and Goodfellow, P.N. (1993). Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature *364*, 713-715.
Wiebe, M.S., Wilder, P.J., Kelly, D., and Rizzino, A. (2000). Isolation, characterization, and differential expression of the murine Sox-2 promoter. Gene *246*, 383-393.
Wilbertz, T., Wagner, P., Petersen, K., Stiedl, A.C., Scheble, V.J., Maier, S., Reischl, M., Mikut, R., Altorki, N.K., Moch, H.*, et al.* (2011). SOX2 gene amplification and protein overexpression are associated with better outcome in squamous cell lung cancer. Mod Pathol *24*, 944-953.
Wilkinson, D.G. (1995). RNA detection using non-radioactive in situ hybridization. Curr Opin Biotechnol *6*, 20-23.
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, M.J., and Dearden, P.K. (2008). Evolution of the insect Sox genes. BMC Evol Biol *8*, 120.
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.
Wright, E., Hargrave, M.R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A., and Koopman, P. (1995). The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. Nat Genet *9*, 15-20.
Wright, E.M., Snopek, B., and Koopman, P. (1993). Seven new members of the Sox gene family expressed during mouse development. Nucleic Acids Res *21*, 744.
Wu, J., Saint-Jeannet, J.P., and Klein, P.S. (2003). Wnt-frizzled signaling in neural crest formation. Trends Neurosci *26*, 40-45.
Wunderle, V.M., Critcher, R., Ashworth, A., and Goodfellow, P.N. (1996). Cloning and characterization of SOX5, a new member of the human SOX gene family. Genomics *36*, 354-358.
Yamada, T., Pfaff, S.L., Edlund, T., and Jessell, T.M. (1993). Control of cell pattern in the neural tube: motor neuron induction by diffusible factors from notochord and floor plate. Cell *73*, 673-686.
Yamaguchi, S., Hirano, K., Nagata, S., and Tada, T. (2011). Sox2 expression effects on direct reprogramming efficiency as determined by alternative somatic cell fate. Stem Cell Res *6*, 177-186.

Yang, S., Zheng, J., Ma, Y., Zhu, H., Xu, T., Dong, K., and Xiao, X. (2012). Oct4 and Sox2 are overexpressed in human neuroblastoma and inhibited by chemotherapy. Oncol Rep *28*, 186-192.
Yoon, D.S., Kim, Y.H., Jung, H.S., Paik, S., and Lee, J.W. (2011). Importance of Sox2 in maintenance of cell proliferation and multipotency of mesenchymal stem cells in low-density culture. Cell Prolif *44*, 428-440.
Zaffran, S., and Kelly, R.G. (2012). New developments in the second heart field. Differentiation.
Zappone, M.V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., Tiveron, C., Vescovi, A.L., Lovell-Badge, R., Ottolenghi, S.*, et al.* (2000). Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. Development *127*, 2367-2382.
Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet *20*, 123-128.