Sox2 in Embryonic Stem Cells and Lung Development

María Cristina Gontán Pardo

© M. C. Gontán Pardo, 2009

This thesis was written within the departments of Pediatric Surgery and Cell Biology at the Erasmus University Medical Center, Rotterdam

The research described in this thesis was financially supported by the Sophia foundation for science research (SSWO), project number 413

The printing of this thesis was financially supported by the J.E. Jurrianse Stichting

Cover photography: http://www.istockphoto.com
Printed by PrintPartners Ipskamp, Enschede

Sox2 in Embryonic Stem Cells and Lung Development

Sox2 in embryonale stamcellen en longontwikkeling

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Wednesday 8 April 2009 at 15.45 hrs

by

María Cristina Gontán Pardo

born in Madrid, Spain

Doctoral Committee

Promotors:	Prof.dr. D. Tibboel
	Prof.dr. F.G. Grosveld
Other members:	Prof.dr. J.N.J. Philipsen
	Dr.ir. D.N. Meijer
	Dr. J. Gribnau
Copromotor:	Dr. R.J. Rottier

A mis padres y a Ria

Contents

	List of abbreviations	8
	Scope of this thesis	9
Chapter 1	Lung development	11
	I. Anatomy of the mature respiratory system	13
	II. Embryonic/fetal development 1. Phases of lung development 2. Respiratory epithelial cells types in the lung 3. Lung cell fate specification in the foregut 4. Lung bud formation 5. Branching morphogenesis 6. Molecular control of lung epithelial differentia	14
Chapter 2	Sox proteins and Sox2 functions in embryonic stem cells	45
	I. Sox proteins 1. Transcription factors 2. The HMG box superfamily 3. The Sox family	47
	II. Sox2 functions in embryonic stem cells 1. Definition of embryonic stem cells 2. Core transcription factor network in ES cells 3. Promoter co-occupancy of pluripotency transcription factors 4. Induced pluripotent stem cells	55
Chapter 3	Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation	65
Chapter 4	Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors	81
Chapter 5	The Sox2 protein interaction network in embryonic stem cells	101
Chapter 6	Discussion	125
Chapter 7	Summary/Samenvatting	143
Curriculum vitae		150
List of publications		151
PhD Portfolio		153
Acknowledgements		155
Appendix		159

List of abbreviations

AEG	anophthalmia-esophageal-genital
A-P	anterior-posterior
BMP	bone morphogenetic protein
CC10	Clara cell protein 10
CCAM	congenital cystic adenomatoid malformation
cGRP	calcitonin gene-related peptide
ChIP	chromatin Immunoprecipitation
DNA	deoxyribonucleid acid
E	embryonic day
EA/TOF	esophageal atresia with tracheoesophageal fistula
ES	embryonic stem
Exp	exportin
Fgf	fibroblast growth factor
GST	glutathione-S-transferase
H&E	hematoxiline and eosine
HMG	high mobility group
ICM	inner cell mass
Imp	importin
IP	immunoprecipitation
iPS	induced pluripotent stem cells
kD	kilo dalton
NE	nuclear envelope
NEBs	neuroendocrine bodies
NES	nuclear export signal
NLS	nuclear localization signal
NPC	nuclear pore complex
Oct	octamer-binding protein
PcG	polycomb group proteins
P-D	proximal-Distal
PN	postnatal
PNECs	pulmonary neuroendocrine cells
PRCs	polycomb repressive complex
rtTA	reverse tetracycline transactivator
Shh	sonic hedgehog
SMA	smooth muscle actin
Sox	sry-related HMG box
SPC	surfactant protein C
SRY	sex-determining region of the Y chromosome
SUMO	small-ubiquitin-related modifier
Tgf	transforming growth factor
wk	weeks

Scope of this thesis

Sox2 is a fascinating transcription factor with multiple roles during embryonic development. In early embryonic development, Sox2 is one of the key transcription factors in the maintenance of the pluripotent status of the cells of the inner cell mass (ICM). Sox2 is also in the transcription factor network that controls the self-renewal and pluripotency of embryonic stem (ES) cells. Later during development, the expression pattern of Sox2 in the developing lung epithelium may suggest a role of Sox2 in lung development.
In this thesis we focus on three aspects of Sox2, its function in lung development, its nuclear import and its interacting partners in embryonic stem cells.

Chapter 1 gives an overall introduction of the basics of lung development, which is helpful to understand the study presented in the chapter 3 of this thesis. I start with a description of the mature respiratory system, the phases of lung development and the different cell types of the respiratory epithelia. Subsequently, the molecular control underlying the specification of the foregut, lung bud formation and branching morphogenesis is discussed in more detail. The chapter finishes with a description of some of the factors involved in the molecular control of lung epithelium differentiation.

Chapter 2 describes the characteristics of the Sox family of transcription factors, their classification within the HMG superfamily, their proteins-protein interactions and their nuclear transport properties. I also describe the expression pattern of Sox2 and its function during mouse development. The second part of this chapter gives an overview of the role of Sox2 in ES cells and in the generation of iPS cells.

Chapter 3 describes the effect of Sox2 ectopic expression on lung development. Transgenic mice were generated that express Sox2 in the airway epithelium, in a doxycycline-inducible manner. The results showed that continuous expression of Sox2 disrupts the normal branching morphogenesis, suggesting that downregulation of Sox2 is necessary for the airway epithelium to respond to branch inducing signals. Furthermore, Sox2 ectopic expression does not lead to a complete abrogation of the epithelial differentiation program but prematurely drives cells into committed progenitors.

Chapter 4 reports the identification of Exportin 4 (Exp4) as the nuclear import factor for Sox2 and Sry. It also presents two alternative nuclear import pathways for Sox2, mediated by Imp9 and the Impß/7 heterodimer. These data suggest that several parallel import pathways facilitate the nuclear import of Sox proteins, which makes import regulation at the level of the import receptor unlikely.

Chapter 5 presents a biochemical approach to identify Sox2 interacting proteins in mouse ES cells. Affinity-purified FLAG-tagged Sox2 and its interacting proteins were analyzed by mass spectrometry. The list of putative Sox2 partners contains ES cell self-renewal factors, chromatin remodelling factors, lineage-specific transcriptions factors and Exportin 4. These results are discussed in relation to recent studies performed in ES cells that reported the target genes of Sox2, the promoter co-occupancy with other pluripotency factors, including Oct4 and Nanog, and the effect on gene expression of Sox2 depletion.

Chapter 6 is a general discussion of the studies carried out in this thesis and the possible future directions.

Chapter

1

Lung development

Lung development

Formation of the vertebrate lung occurred relatively late in evolution and was a significant step in the adaptation of vertebrates to breathing air[1]. The lung is a complex organ, formed by dozens of different cell types [2]. Its principal function is gas exchange, to deliver oxygen to cells for metabolic energy and to remove the respiratory product carbon dioxide from the blood.

Embryonic lung development is an attractive system to study the basic principles of organogenesis and branching morphogenesis. The advantage of studying the lung in comparison to other organ systems is the possibility of fetal organ culture, the stereotyped branching process and that the lung is not required for prenatal existence. However, it must reach a certain maturation stage to be immediately functional at birth. Finally, understanding the molecular control of lung morphogenesis may improve the treatment of congenital defects in humans, such as congenital diaphragmatic hernia (CDH), congenital cystic adenomatoid malformation of the lung (CCAML), or esophageal atresia with tracheoesophageal fistula (EA/TOF) [3].

This chapter describes the molecular basis of lung bud formation, branching morphogenesis and airway epithelial differentiation. Lung alveolar and vascular development is outside the scope of this introduction, for recent reviews on these subjects, see [4, 5].

I. Anatomy of the mature respiratory system

The respiratory system can be divided into two functional components: the conducting and the respiratory zone (**Figure 1**).

1. **The conducting zone:** this is the site where inspired air is warmed up, moistened, and filtered. It is reinforced with cartilage and smooth muscle. This part of the respiratory system consists of the following components (http://www.ana.ed.ac.uk):
   - **Nasal cavities** contain hairs that filter the inspired air and olfactory mucosa responsible for the sense of smell.
   - **Pharynx** is part of the alimentary tract and also acts as a resonant cavity for the voice.
   - **Larynx** (voice box) contains the vocal cords, which generate the voice.
   - **Trachea** (windpipe) is a flexible connection, surrounded by *cartilage* rings, between the upper respiratory tract and the lungs.
   - **Bronchi**, the trachea divides into the two main bronchi that lead into the left and right lungs. Mammalian lungs are generally asymmetrically lobulated; for example, in the mouse, four lung lobes form on the right and one on the left side (**Figure 4 D**), while in the humans there are three lobes on the right and two on the left side [1, 2, 6]. The bronchi continue to divide within the lung, and after multiple divisions, give rise to bronchioles.
   - **Bronchioles** are the final conducting part of the lungs, which undergo more branching divisions. The bronchiole wall is mainly formed by smooth muscle, without the presence of cartilage.

- **Terminal bronchioles**, the bronchial tree continues branching until it reaches the level of terminal bronchioles, which lead to the respiratory zone.

2. **The respiratory zone:** this is the site of gas exchange with the blood, it contains:
   - **Respiratory bronchioles**, which are similar to terminal bronchioles, but their walls are regularly interrupted by alveoli.
   - **Alveolar ducts,** which are the areas where the frequency of interspersed alveoli increases, filling up most of the airways surface.
   - **Alveoli,** which are the unit of gas exchange, providing an enormous surface area of arround 160 m^2 (80 times the area of our skin), surrounded by a network of capillaries.

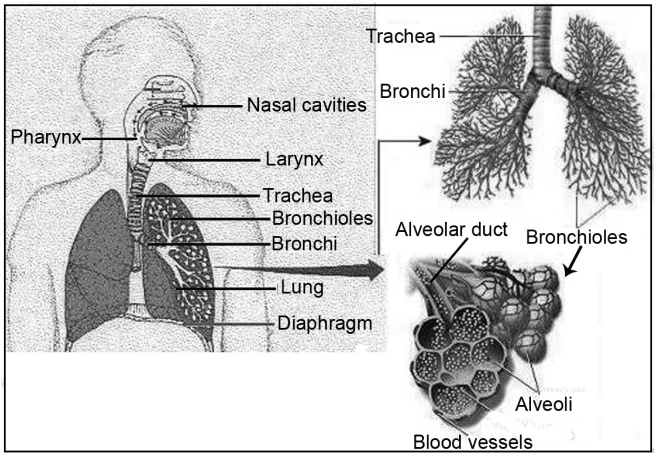


Figure 1. Anatomy of the mature respiratory system
(modified from http://universe-review.ca/I10-13-respiratory.jpg)

II. Embryonic/fetal development

1. Phases of lung development

Lung development is organized in six distinct overlapping stages based on the anatomic changes that occur in lung architecture (**Table 1 and Figure 2**): the early organogenesis period comprises the embryonic and the pseudoglandular stages. The **embryonic** stage starts with the outgrowth of the lung bud from the ventral region of the foregut into the surrounding mesenchyme, eventually forming the main stem bronchi and the future trachea. During the **pseudoglandular** stage, the main branches of the bronchial tree develop generating the conducting part of the lung until the terminal bronchioles. The differentiation of the fetal lung starts at the **canalicular** stage in which the respiratory part and surrounding vasculature of the lung begin to develop. The terminal bronchioles divide to form the respiratory bronchioles, forming the primitive alveolar ducts, which end in terminal sacs. At the same time, extensive angiogenesis within the peripheral mesenchyme

leads to vascularisation of the developing respiratory structures. Subsequently, in the **saccular** stage, a rapid expansion of the respiratory portion of the lung takes place. Terminal sacs dilate and branch to form primitive alveoli, type I cells flatten and spread increasing the surface area for gas exchange, and type II cells start to secrete surfactant. The stromal compartment reduces in size, bringing the growing capillary network into close contact with the immature alveoli. Although already functional, the lung is still relatively immature at birth. Lung development will be completed during the **alveolar** stage when fully mature alveoli appear. Finally, the **microvascular maturation** stage occurs during postnatal lung development and is characterized by the fusion of the double capillary network into a single capillary network in the inter-alveolar septa [7]. Secondary septal subdivision of existing immature alveoli generates new alveoli and this process will continue into adulthood. The decrease in the relative proportion of parenchyma to total lung volume contributes significantly to increase the gas exchange surface area after birth; thereafter all components will grow proportionately until adulthood [1, 7-10].

Table 1. Phases of lung development

Phase	Events	Mouse	Human
Embryonic	Formation of the lung buds from the primitive foregut, mayor bronchi and division of tracheal-esophageal tube	E9-11.5	3-7 wk
Pseudoglandular	Formation of bronchial tree and all terminal bronchioles, initiation of **epithelial differentiation**	E11.5-16.5	5-17 wk
Canalicular	Completion of conducting airways, formation of prospective the respiratory portions of the lung, epithelial differentiation, capillary bed	E16.5-17.5	16-26 wk
Saccular	Formation alveolar ducts, expansion of the air spaces, cell differentiation, surfactant production	E17.5-PN5	24-38 wk
Alveolar	Alveolarization, thinning of interalveolar septa, expansion of air spaces	PN5-30	36 wk to maturity
Microvascular maturation	Fusion of double alveolar capillary network into a single layer	Birth to maturity	Birth to maturity

2. Respiratory epithelial cells types in the lung

From E14.5 onwards in mouse development, the primitive undifferentiated columnar epithelium that lines the primordial lung branching tree starts to differentiate. After differentiation two functionally and structurally distinct areas can be distinguished in the lung epithelium, the conducting airways and the respiratory epithelium. The conducting airways can be further subdivided into a proximal part with pseudostratified columnar epithelium (trachea and bronchi) and a distal part with more cuboidal epithelium (terminal bronchioles). In the respiratory epithelium there is an abrupt transition between the more proximal part (respiratory bronchioles) lined by cuboidal epithelium and the distal part (alveoli) consisting of cuboidal/squamous epithelium [8].

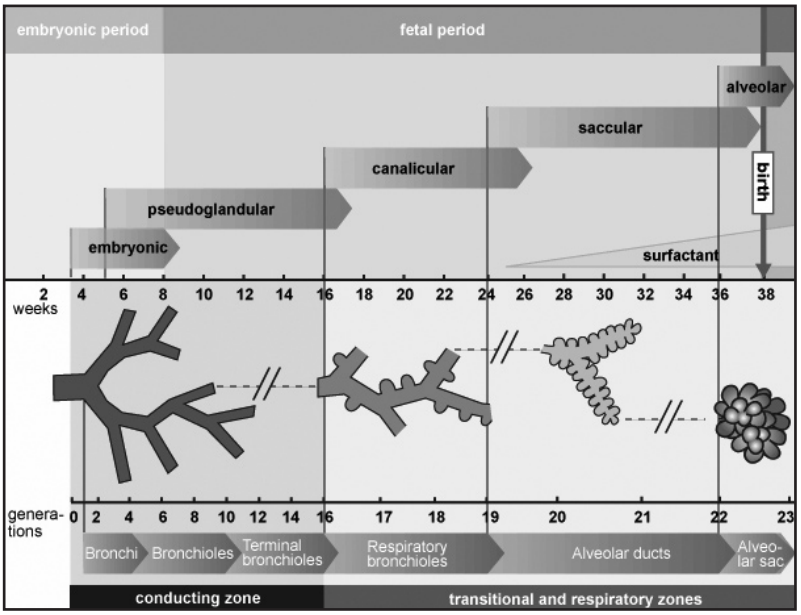


Figure 2. Phases of human lung development (from www.embryology.ch).

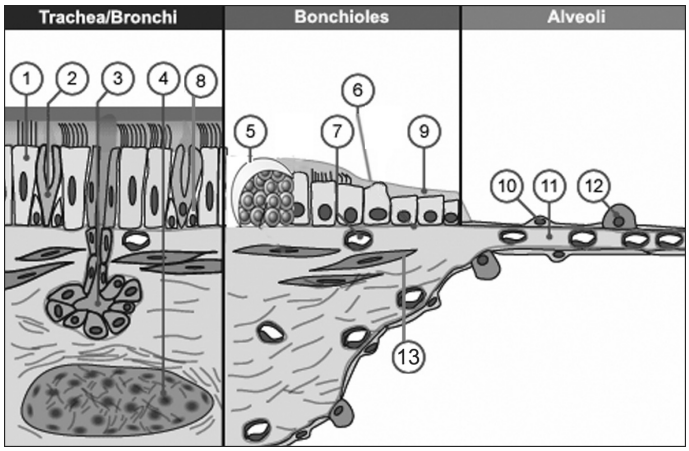
The lung epithelial cells proliferate and differentiate into multiple cell types that vary along the proximal-distal (P-D) axis of the lung. The pseudostratified columnar epithelium of the proximal conducting airways consists of five main cell types: ciliated, goblet, basal, Clara and neuroendocrine cells. The more distal cuboidal epithelium of the bronchi is mostly formed by ciliated and Clara cells, as well as, less abundant neuroendocrine cells. Ciliated cells possess motile cilia on their apical surface, and goblet cells secrete mucous covering of the airways. Together they participate in mucociliary clearance. Basal cells, as their name implies, are small cells at the basal lamina that are progenitor cells for other airway epithelial cells. The major secretory *protein* of Clara cells is the Clara cells secretory protein (CC10) that is thought to play a role in homeostasis and host defense. Finally, the relatively rare pulmonary neuroendocrine cells (PNEC) are present in the airway epithelium as isolated cells or in clusters termed neuroendocrine bodies (NEB)[11]. These cells are among the first ones to differentiate in the lung [12, 13] and express neuroendocrine peptides, such as the calcitonin gene-related peptide (cGRP) [2]. Their role is not completely understood, but they may function as oxygen sensitive chemoreceptors [11], as well as regulating further epithelial differentiation by expressing different cytokines [9, 14, 15].

The alveolar epithelium in the alveoli forms the gas-exchange region of the lung and is lined by cuboidal type II cells and squamouse type I cells. Type II cells are located at the junction of the alveoli and secrete surfactant lipids and proteins, substances necessary for reduction of surface tension in the alveoli and essential for perinatal respiratory adaptation [16, 17]. In addition, they are the only dividing cells in the alveolar epithelium and function as progenitor cells that give rise to type I cells. Type I cells

are in close proximity with the capillaries and form part of the blood/air interface required for efficient gas exchange (see **Table 2** and **Figure 3**, which gives an overview of the distinct epithelial cell types present in the adult lung, their location, morphology, function and their cell-specific markers).

Table 2. Respiratory epithelial cells types in the lung

Location	Cell Type	Morphology	Functions	Specific marker
Conducting airways (trachea, bronchi, bronchioles)	Ciliated	Columnar, narrowed toward the base; motile cilia present on the apical surface	Proximal transport of the mucus stream	Foxj1 ß-tubulinIV
	Clara	Columnar non-ciliated; apical cytoplasmic protrusion containing secretory-like dense granules	Secretion of surfactant proteins; detoxification of toxins; renewal of bronchiolar epithelium	CC10
	Goblet	Columnar, narrowed toward the base; contain granular mucin droplets that discharge apically	Produce and secrete the mucus covering of the airways	Mucin 5A
	Basal	Small round to polygonal cells situated along the basal lamina	Progenitor cells in upper airways (of ciliated and goblet cells)	p63 K14
	Neuroendocrine	Triangular cells broadest at base and narrowed apically; found in clusters in neuroepithelial bodies (NEB) or solitary	Possible function as oxygen sensitive chemoreceptor with regulatory functions	cGRP
Alveoli	Type I cells	Squamous cells that cover about 95% of the alveolar surface; They are extremely thin and form part of the blood/air interface. Non-dividing	Air/blood gas exchange	T1α Aquaporin 5
	Type II cells	Cuboidal cells containing lamellar bodies (surfactant storage sites); generally located at the junctions between alveoli.	Proliferate in response to lung injury acting as progenitor for type I cells; synthesize, secrete and recycle surfactant	Sftpc



1. Ciliated cells
2. Goblet cell
3. Gland
4. Cartilage
5. Pulmonary neuroendocrine cells
6. Clara cell
7. Capillary
8. Basal cells
9. Surfactant
10. Type I cells
11. Alveolar septum
12. Type II cells
13. Smooth muscle cell

Figure 3. Epithelial cell types present in the adult lung (adapted from www.embryology.ch).

3. Lung cell fate specification in the foregut

The definitive endoderm is established as one of the three germ layers during gastrulation. The formation of the presumptive gut occurs through the invagination of endodermal cells at the anterior and posterior side of the developing embryo to form a tube-like structure [18, 19]. The foregut, the most anterior part of the primitive gut tube, will generate the lungs and its neighbouring foregut derivatives (including pancreas, liver, thyroid, esophagus and stomach) [18]. The formation of these organs involves paracrine signaling from the surrounding mesoderm structures (heart, notochord and the septum transversum) and the competence of the foregut endoderm to respond to them [20]. Additionally, organ specification and morphogenesis depend on both ubiquitous and tissue-specific transcription factors that work in interacting networks [21, 22] (discussed below).

The signals that regionalize the initial multipotential foregut endoderm, prior to lung bud formation, are poorly understood. *In vitro* embryonic foregut cultures helped to identify the tissue interactions necessary for the initial specification of pulmonary cell fate. The cardiac mesoderm influences the antero-posterior (A-P) differentiation of the foregut endoderm [19] and recombinant fibroblast growth factors (Fgfs), Fgf1 and Fgf2, can replace the requirement for cardiac mesoderm: low Fgf concentrations activate liver-specific genes, while higher concentrations activate lung-specific genes. Additional data indicate that signaling through Fgf receptor 4 (Fgfr4), expressed in the ventral foregut endoderm, is involved in lung specification [23]. Together, these findings suggest that a temporally specific gradient of Fgf signalling, produced by the cardiac mesoderm, is crucial for early determination of lung endoderm specification in the foregut [23]. Retinoic acid also plays a role in the early specification of the foregut (see section **4.3.**), other signaling pathways, such as sonic hedgehog (Shh) and bone morphogenetic protein (Bmps), are also present in the foregut but until now there is no clear evidence

for their involvement in lung endoderm specification [19]. However, Bmp signalling is involved in the morphogenesis of the trachea and esophagus (see section **5.4.**) [24, 25].

By E8-9.5, a number of transcription factors are expressed in overlapping domains along the A-P axis of the gut endoderm that determine the different prospective organ specific domains [23] (**Figure 4A**). Transcription factors such as Sox2, Foxa1, Foxa2, Gata4 and Gata6, which are expressed early in the endoderm, are also involved in morphogenesis of the foregut [19, 26]. Understanding the function and regulation of these factors will help to comprehend how the endoderm generates different organs from the different regions of the gut. Interestingly, many of these factors will be reutilized later in lung development and they will be discussed later.

To date, transcription factors that uniquely specify lung formation have not been identified. The lung forms in the anterior ventral foregut, between the thyroid and the liver. At E9 the first respiratory progenitors can be visualized in this regions as a group of cells expressing the transcription factor thyroid transcription factor-1 **(**Ttf1) [27], but this protein also marks the thyroid primordium, which starts to develop one day earlier [28, 29] (**Figure 4B**). Furthermore, mice deficient for Ttf1 still initiate lung morphogenesis (see later). The surfactant protein C (*Sftpc*) is the most specific marker of lung epithelial cell lineage, but its expression is detected only by E10-10.5, after the primary buds form [30].

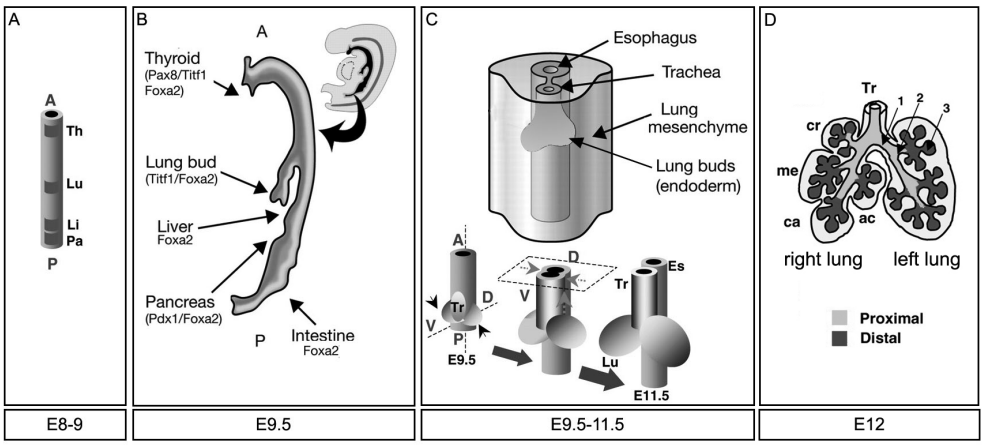


Figure 4. Patterning of the foregut endoderm into organ-specific domains along the A-P axis, which will generate the thyroid (Th), trachea (Tr), lung (Lu), liver (Li) and pancreas (Pa) (**A**). Mouse endoderm is depicted at approximately E9.5. Transcription factors (Pax8, Titf1, P-Dx1, Foxa2) identify cells that contribute to organ formation along the A-P axis. The thyroid domain is marked by co-expression of Ttf1, Foxa2 and Pax8, while the lung domain is marked by co-expression of Ttf1 and Foxa2 [31, 32] (**B**). Top: lung bud formation. Bottom: lung buds, trachea and esophagus (Es) morphogenesis at E9.5–11.5 (**C**). Branching morphogenesis, the left lung has one lobe and the right has four: the cranial (cr), medial (me), caudal (ca) and accessory (ac) lobes. Numbers indicate primary, secondary and tertiary generations of buds (**D**) (modified from [1, 2, 19]).

4. Molecular control of early lung morphogenesis

In mice, at E9-9.5 the lung starts to develop as an outgrowth of cells from the ventral foregut endoderm into the splanchnic mesenchyme[1]. The foregut is divided by a longitudinal septum in two tubes, the future trachea and lung buds on the ventral side and the esophagus on the dorsal side (**Figure 4A-C**). However, the molecular mechanisms that septate the foregut dividing the alimentary and the respiratory system are not well understood. Faulty separation of the trachea and esophagus causes the development of fistulas between the esophagus and the trachea, leading to a variety of congenital anomalies in humans, of which esophageal atresia with tracheoesophageal fistulas (EA/TOF) is the most frequent and clinically relevant [33].

The trachea and the bronchial tree are lined with endodermally derived epithelium. Similar to gut morphogenesis, epithelial-mesenchymal interactions that ultimately result in local activation of complex gene networks, are important for trachea formation, primary bud outgrow and lung branching morphogenesis [20]. *In vitro* transplantation experiments have demonstrated the importance of the lung mesenchyme in directing airway epithelium to branch. Proximal mesenchyme from the tracheal region was unable to induce branching when apposed to distal lung epithelium. In the converse experiment, isolated distal mesenchyme (from the region of lung bud formation) will induce budding of tracheal epithelium [34-36]. Recent experiments have demonstrated that autocrine-paracrine signaling cascades, such as Shh and Fgf signaling, are directly implicated in early lung development, together with Retinoic acid (RA), Hox genes, Ttf1 and T-Box transcription factors, which are also involved in these early morphogenetic events [19]. These factors will be discussed next.

4 1. Thyroid transcription factor-1

As mentioned above, the homeodomain protein Ttf1, also known as Nkx2.1 [28] is the earliest known marker of respiratory epithelium, with a reported onset of expression in the developing lung and thyroid at E9-9.5 [27, 28]. Early in morphogenesis, Ttf1 is expressed in all pulmonary epithelium, but becomes progressively restricted to the distal alveolar type II cells and proximal Clara cells [31, 37-39]. In the lung, Ttf1 plays a critical role in the separation of the trachea from the esophagus, in branching morphogenesis and in epithelial cell differentiation.

Deletion of *Ttf1* in the mouse results in a neonatal lethal lung phenotype, in association with malformations of the thyroid [29]. The Ttf1-deficient lungs fail to form the septa between trachea and esophagus, leading to a tracheoesophageal fistula. Furthermore, lung development is arrested at the pseudoglandular stage, with the main bronchi developing into abnormal sac-like lungs and failing to undergo further branching morphogenesis (**Figure 5**). The mutant epithelium fails to express any of the surfactant proteins typically found in the normal distal lung, but does, however, express cilia and goblet cells, normally found in the proximal airways [27, 40, 41]. The Ttf1 target genes that are critical for early lung morphogenesis are not known, although Bmp4 expression was shown to be directly regulated by Ttf1 [42]. These data suggest that Ttf1 is essential for

lung branching and terminal cell differentiation (see section 5.1.), but that it is not required for the initial specification of the lung.

4 2. Fibroblast growth factor signaling

The Fgf family is involved in the initiation and expansion of the lung bud, which is mediated by Fgf1, Fgf7, Fgf9 and Fgf10. Fgf1, Fgf7, and Fgf10 are expressed in the mesenchyme around the lung bud and later in lung development in the terminal and lateral branches [43]. *Fgf9* is expressed in airway epithelium at E10.5 and stimulates cell proliferation in the adjacent mesenchyme by maintaining Shh signaling [44, 45]. Fgf10 binds to and signals via the receptor Fgfr2, which is expressed in lung bud epithelium [46]. The appositional expression of *Fgf10* and *Fgfr2* is a clear example of the mesenchymal-epithelial interactions during lung morphogenesis [47].

Mice with a targeted inactivation of *Ffg10* die at birth as a result of lung agenesis, they have a normal trachea without any further pulmonary branching [48, 49]. A similar phenotype was observed in mice with different mutations in the *Fgfr2* gene. However, *Ttf1* expression was not affected, which indicates that specification of the lung remained unperturbed[47-51] (**Figure 5**). These findings suggest that there are different signaling pathways involved in the initial morphogenesis of trachea and lung. The function of Fgf10 in lung bud outgrowth was further demonstrated *in vitro* using lung explant cultures, where mesenchymal Fgf10 acts as a chemo-attractant for the adjacent endoderm and causes extensive budding [43, 52]. Additionally, Fgf10 alone is able to induce budding from the trachea endoderm denuded from proximal mesenchyme [53] (**Figure 6A**). The control of the Fgf10 expression and the positional establishment of the lung bud in the foregut, seem to be controlled, at least in part, by Retinoic acid-Tgfß [54] signaling , Wnt5a [55] and Tbx4 [56] (see later).

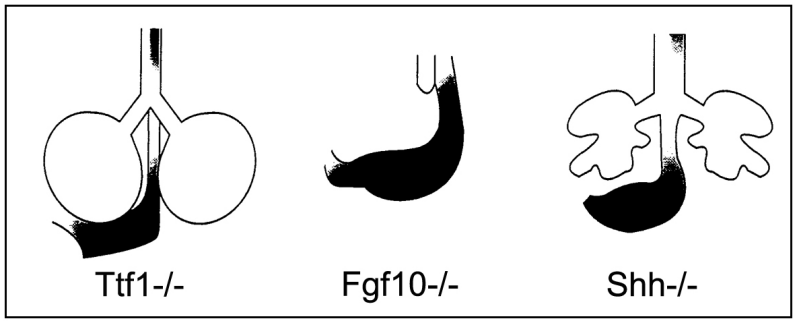


Figure 5. Ttf1 transcription factor and Shh signaling pathways influence the tracheo-esophageal separation and branching morphogenesis, while disruption of Fgf signaling results in normal trachea formation and lung agenesis (adapted from [57]).

4 3. Retinoic acid

Retinoic acid (RA) is the active form of vitamin A and is an embryonic signaling molecule involved in the patterning and differentiation of a variety of organs during development [58, 59]. RA deficiency or excess during development is associated with several congenital defects in humans [58, 60].For example, embryos deprived of vitamin A present tracheoesophageal fistula with a variety of lung defects, including lung agenesis[61]. RA signals through different isoforms of the nuclear receptors RA receptors (RARs) and retinoic X receptor (RXR) that form a heterodimer. They belong to the nuclear receptor superfamily of transcription factors, whose nuclear transport and activity are directly modulated by ligand binding [62].

RA is synthesized in the foregut during lung bud formation by the enzyme retinaldehyde dehydrogenase 2 (*Raldh2*) [63]. Mice deficient for *Raldh2*, if embryos are rescued from early lethality by maternal RA supplementation, failed to develop lungs [64].

It has been shown that RA induces *Fgf10* expression within the mesoderm adjacent to this region [65]. Disruption of RA signaling in E9 foregut cultures does not only prevent the lung bud formation, but also blocks the local expression of *Fgf10* and disrupts expression of *Ttf1* at this restricted site [66-68]. Recently, it has been shown that RA represses Tgfß signaling in the mesoderm of the primitive foregut, adjacent to the prospective lung domain, which allows the expression of Fgf10 and the initiation of the lung bud formation [54]. In RA-deficient embryos, Tgfß signaling is upregulated and the primary lung bud fails to outgrow due to defective Fgf10 signaling [54, 66]. Accordingly, compound *RARa* and *RARß* receptor gene-targeted mice present severe abnormalities in lung morphogenesis, including lung agenesis/hypoplasia and tracheal cartilage malformations [69, 70]. Mice deficient for only one of the isoforms showed a less severe phenotype due to possible redundancy among the RA receptors isoforms [58]. In vitro and in vivo data suggest that RA acts as a regulator of gene expression in the early embryonic lung region upstream of Tbx, Hox, Tgfß and Fgf10 signaling (**Figure 6A**) [54, 64].

4 4. T-Box transcription factors

Members of the T-box (Tbx) transcription factor family are key regulators in many events in vertebrate development. In the lung, Tbx1 is expressed in the epithelium, while Tbx2–5 are expressed in the mesenchyme [71]. Tbx4 and Fgf10 are co-expressed in the foregut mesoderm in the lung domain that coincides with Ttf1 expression in the epithelium. Interestingly, ectopic expression of Tbx4 activates Fgf10 and induces ectopic lung budding in the chick [56] (**Figure 6A**). It seems that Tbx4 and Fgf10 are not required to initiate Ttf1 expression, but appear to play a role in defining the posterior boundary of Ttf1 and the lung primordium [56]. Tbx genes may have redundant functions in the lung because of their overlapping expression patterns [71], which may explain the normal lung development in knockout mouse for Tbx4 or Tbx5 [72-74]. However, treatment of mouse lung explants cultures with antisense oligonucleotides to either *Tbx4* or *Tbx5* suppressed the expression of Fgf10 and lung branching *in vitro* [75]. Furthermore, *Tbx2*, *Tbx4* and *Tbx5* were downregulated in RA-deficient mice, which suggests that Tbx gene expression is RA-dependent [64].

4 5. Hox genes

Homeobox domain-containing (Hox) transcription factors are involved in axial patterning and specification of regional identity during development. At least 15 Hox genes are expressed in partially overlapping domains along the P-D axis of the developing lung, including several member of the Hoxa and Hoxb clusters In general, Hox genes are expressed in the pulmonary mesenchyme and their levels decrease with advancing gestation [76-79]. It was suggested that a specific combination of Hox genes expression is important in lung development, as they might be involved in specifying the P-D axis of the lung [77]. However, the role of Hox genes in lung patterning is still unclear, as lung development is unaffected in single Hox-null mutants, probably due to functional redundancy.

To date, only Hoxa5-null mice presented a specific lung phenotype. Hoxa5 is expressed in the foregut mesoderm at the site of lung bud formation and later in the development it is expressed in the mesenchyme of the trachea, bronchi, and distal lung [80]. Hoxa5-deficient lungs showed improper trachea formation, impaired lung branching and decreased expression of surfactant proteins, *Ttf1*, *Foxa2*, and *Nmyc* in the epithelium [79]. This suggests that *Hoxa5* expression is necessary for normal lung morphogenesis [81].

4 6. Sonic hedgehog signaling and Gli genes

Shh signaling is required for branching morphogenesis and serves as a paracrine signal that regulates smooth muscle differentiation in the lung mesenchyme [82, 83] and together with *Fgf9* controls cell proliferation and formation of the lung capillary network in the mesenchyme [45] . Shh is a secreted signaling molecule that is expressed in the foregut endoderm and later in the lung epithelium with the highest expression at the tips of the terminal buds [84]. Shh diffuses to the mesenchyme and binds to the extracellular matrix and to other proteins, such as hedgehog interacting protein-1 (Hip1) and its primary receptor patched (Ptc). Shh binding releases the basal repression that Ptc exerts on smoothened (Smo). Inside the cell, Smo activates Gli zinc finger transcription factors that regulate target gene expression [69, 84]. Three Gli genes have been described in mice: *Gli1*, *Gli2* and *Gli3*, all of which are expressed in the foregut mesoderm and later in the pulmonary mesenchyme [85]. The receptor Ptc is expressed in the lung mesenchyme until E13.5 and in the respiratory epithelium at E16.5 [86].

The importance of Shh signaling for lung development was shown in *Shh*-null mutant mice, which died at birth with tracheoesophageal fistula and sac-like lungs without any lobes or endoderm branches [87] (**Figure 5**). The mesenchyme was the primary target of the Shh deficiency, showing decreased cell proliferation and increased apoptosis [87]. In addition, Smooth muscle actin (SMA), normally surrounding the conducting epithelium, was absent in *Shh* mutant lungs [82]. P-D epithelial differentiation was normal in the *Shh* null mutant lung [82, 87]. This phenotype is similar to the *Ttf1*-deficient lungs. However, Shh was normally expressed in *Ttf1* deficient lungs [27], and *Ttf1* expression was unchanged in *Shh* null mutant lungs [82, 87], which suggest that Ttf1 and Shh regulate lung branching morphogenesis via separate pathways. The effect of Shh on the mesenchyme was

corroborated by the overexpression of Shh in distal epithelium, which resulted in smaller lungs at birth that lacked alveoli, but had an increased proportion of mesenchymal mass and increased proliferation [84].

Gli proteins have specific and overlapping functions in transducing Shh signaling during early lung development. The importance of the Gli proteins for lung bud formation was shown in compound null-mice lacking both *Gli2* and *Gli3* (*Gli2*$^{-/-}$, *Gli3*$^{-/-}$). In these mutants the trachea and esophagus were absent, while other foregut derivatives develop but are hypoplastic [88]. This phenotype was ameliorated by the presence of one *Gli3* allele (*Gli2*$^{-/-}$, *Gli3*$^{+/-}$), resulting in lungs with one hypoplastic lobe and a tracheoesophageal fistula [88]. Other single or compound mutants for the three Gli genes show a variable degree of foregut malformations, lung hypoplasia and lobulation defects [88-91].

A possible down-stream target in the Shh/Gli signaling is Foxf1 [92]. It plays a critical role in embryonic development and its deletion is lethal before gastrulation. Foxf1 haploinsufficiency caused similar malformations of the lung, trachea, and esophagus as were seen in mutants for the Shh/Gli pathway, and like in Shh deficient lungs, the P-D epithelial differentiation is normal [92, 93]. Expression of a number of genes critical for lung morphogenesis and function, including *Sftpb*, *Bmp4*, *Tbx*, *Fgf10*, and *Gli3* were reduced, implicating Foxf1 in the regulation of mesenchyme-epithelial interactions [92-94]. At the onset of lung organogenesis (E9.5), *Foxf1* expression is restricted to the foregut mesoderm around the presumptive lung bud, suggesting that it may participate in the mesenchymal-epithelial induction of lung and gut morphogenesis [95]. Foxf1 is absent from the trachea and lung mesenchyme in Shh-deficient mice, whereas exogenous Shh activated Foxf1 expression. [92].

The exchange of signals between the growing bud and the surrounding mesenchyme establishes feedback responses that control the size and shape of the bud during branching. This is illustrated by the mechanism by which the Shh pathway controls Fgf10 expression [19]. Several *in vitro* and *in vivo* studies support the idea that *Fgf10* expression in the lung is negatively regulated by Shh [43, 96]. Based on this, it has been suggested that Shh expressed in the epithelium of the lung bud gradually downregulates *Fgf10* expression in the mesenchyme as the bud grows towards the source of Fgf10, and thus limits further bud outgrowth. Interestingly, Shh may also control the shape of the bud by preventing the widespread expression of Fgf10 in the mesenchyme. In lungs from Shh-null mice, *Fgf10* transcripts are found diffusely expressed in the lung mesenchyme [82]. One possible model to explain the complicated interaction between Shh and Fgf10 involves the Shh induction of Hip1, which is a protein expressed in the distal lung mesenchyme that can bind Shh protein directly [97]. Thus, Hip1 inhibits Shh signaling by ligand sequestration, which release Shh-mediated repression of Fgf10 locally. This model is supported by the phenotype of Hip1-deficient lungs, which show defective branching. This can be rescued by the upregulation of Shh and the subsequent repression of Fgf10 [97]. All together, these data support a role for Shh/Hip1 in epithelial-mesenchymal signaling during early lung formation (**Figure 6B**).

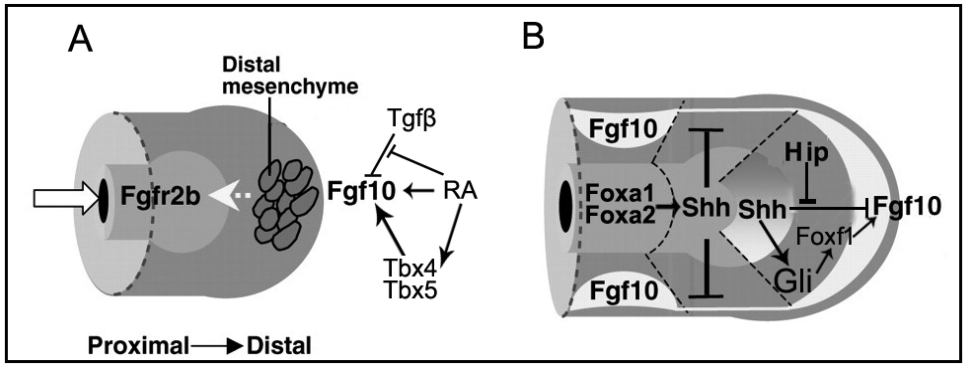


Figure 6. Model for bud initiation (A) and Shh signaling in the lung bud (B) (adapted from [19]).

5. Branching morphogenesis of the lung

Branching morphogenesis is an essential process for the development of the lung and other organs (such as the kidney and the mammary gland). Stereotypic branching occurs during lung lobation and in the first airway generations, formed between E11 to E16 in the mouse. The branching program that forms the bronchial tree in mouse consists of three geometrically distinct branching modes, which are used repeatedly during lung development and are controlled spatiotemporally by a "master" branch generator. These modes are called the domain branching (lobes scaffold), planar bifurcation (lobes edges) and orthogonal bifurcation (lobes surface and interior). These three modes occur simultaneously throughout lung morphogenesis. The branching morphogenesis is stereotypic, however there are some temporal differences and some rare and common "errors" observed during branching [98]. This stereotypic branching implies the presence of a genetic program that controls the "master" branch generator and the different branching modes. There is evidence that points to the Fgf-Fgfr-Sprouty2 pathway as the "master" branch generator [98].

The principles that control lung branching morphogenesis are still not fully understood. In the mouse, an increasing array of signaling molecules, receptors and transcription factors that play critical roles during this process have been identified. Their expression patterns, levels, and identity change dynamically during lung development. Several *in vitro* and *in vivo* studies suggest that signaling via Fgfs, Shh, Bmps, Wnt7/ß-catenin and transforming growth factor ß (Tgfß) forms a regulatory circuit that regulates lung development [2, 96]. These signaling pathways regulate, in a reciprocal way, the expression and activity of members of multiple transcription factors families that function in complex networks to regulate cell fate and functions [2]. Some of these factors have been already described above, as they also participate in lung bud formation, other transcription factors and signaling pathways involved in the regulation of lung branching morphogenesis will be discussed next.

5 1. Thyroid transcription factor-1

Beside its role in early lung development, Ttf1 is also known to directly bind and regulate the expression of a number of genes expressed in respiratory epithelial cells that are important for lung formation and function, such as genes regulating surfactant homeostasis (CC10 and surfactant proteins), vasculogenesis and host defense [31, 99-101]. Study of the Ttf1 promoter reveals the presence of multiple enhancers, with binding sites for both ubiquitous and tissue-specific transcription factors, including Gata6, Foxa2 and Ttf1 itself. Supporting these findings, cotransfection assays verified that the Ttf1 promoter activity is stimulated by Gata6 [102] and Foxa2 [103]. The binding of Ttf1 itself suggests a positive feedback mechanism that sustains expression of *Ttf1*. Thus, initiation of *Ttf1* expression may require initial activation by other factors, but may then be sustained by autoregulation [9].

Ttf1 also regulates postnatal lung development and homeostasis. Postnatally, Ttf1 expression decreases proximally, but remains detectable in adult bronchiolar and alveolar type II cells [37, 39, 104]. Overexpression of Ttf1 in the distal lung epithelial cells, using the SPC promoter, did not affect prenatal lung development, but it perturbed postnatal alveolarization and led to emphysema, severe inflammation and fibrosis [105]. Altogether it seems that Ttf1 plays a central role in lung morphogenesis.

5 2. Foxa family

Foxa1 and Foxa2 (formally designated Hnf-3α and Hnf-3β) belong to the Forkhead box (Fox) family of transcription factors and are known to influence gene expression in endodermally derived tissues including lung [95]. Foxa1 and Foxa2 genes display overlapping expression patterns during lung development [95]. They are expressed in the ventral foregut endoderm before lung bud formation, during lung morphogenesis and they overlap with the expression pattern of Ttf1 in the epithelium. In the mature lung they are detected in the proximal epithelium and in the distal alveolar type II cells [37, 106-108].

The role of Foxa1 and Foxa2 in lung morphogenesis was difficult to study, due to their functional redundancy. Foxa1-deficient mice displayed normal lung development [109-111], whereas inactivation of Foxa2 was early embryonic lethal, before lung formation, due to defects in foregut morphogenesis [112, 113]. Selective deletion of Foxa2 in respiratory epithelium did not affect branching but delayed the maturation of the peripheral lung resulting in death after birth [114, 115]. However, deletion of both Foxa1/Foxa2 genes resulted in lung hypoplasia, inhibited branching, and lack of epithelial and smooth muscle differentiation [116]. Deletion of Foxa1 and Foxa2 inhibited Shh signaling, resulting in decreased formation of bronchial and vascular smooth muscle [116]. Thus, Foxa proteins regulate paracrine signaling via Shh, which is required for normal lung branching morphogenesis (**Figure 6B**). Additionally, *in vitro* studies showed that *Foxa2* activated the expression of surfactant proteins, CC10, Wnt7b and Ttf1, supporting its function in differentiation of the respiratory epithelium [2, 95].

5 3. Tgfß signaling

Tgfß1, Tgfß2 and Tgfß3 are members of the Tgfß superfamily that consists of highly conserved secreted signaling molecules, including activins, Bmps, and Tgf-ßs [117]. The biological activities of Tgfßs are mediated by their binding to heteromeric complexes of serine/threonine kinase receptors, known as Tgfß receptors type I and type II (TgfßrI and II). Upon Tgfß -induced aggregation of the receptors, the constitutively active type II receptor phosphorylates and activates the type I receptor, which subsequently phosphorylates specific Smad proteins, Smad2 and 3. Phospho-Smad2-3 form a complex with Smad4 and translocate into the nucleus to act as transcriptional co-modulators to induce or repress target gene expression [118]. Tgfß ligands, their receptors (TgfßrI and II) and their signal transducing proteins (Smads), are expressed in the developing mouse lung and have been implicated in regulation of branching morphogenesis [119].

Tgfß1 mRNA is present in the mesenchyme during lung development and the protein accumulates in regions in between buds (mesenchyme and epithelium), where extracellular matrix (ECM) components are also present [120-122]. Tgfß1 plays an inhibitory role on lung branching, as transgenic mice overexpressing Tgfß1 in the distal epithelium presented decreased lung branching morphogenesis [123-127]. Supporting these data, in vitro inhibition of TgfßrII signaling stimulated lung morphogenesis in whole lung explants [128]. Interestingly, Tgfß1-null mice presented normal lung morphogenesis, although its phenotype may have been rescued by maternal transfer of Tgfß1 [129, 130]. In vitro studies indicate that activation of Tgfß1 signaling in mesenchymal cells inhibits Fgf10 expression [54, 96, 131, 132]. This suggests a mechanism where Tgfß1 blocks Fgf10 expression in the mesenchyme between buds, preventing budding locally (**Figure 6A** and **7**). Tgfß2 is expressed in the distal lung epithelium, while Tgfß3 is present in the proximal lung epithelium, mesenchyme and pleura [120, 121]. Tgfß2 or Tgfß3-deficient mice die postnatally due to abnormalities in lung function at birth [133, 134].

Smad proteins are localized mainly in distal lung epithelium, where they mediate effects of Tgfß signaling during lung morphogenesis [135, 136]. Different studies showed that Smad2 and Smad3 negatively regulate branching morphogenesis, while Smad1 and Smad7 positively regulate branching morphogenesis of the lung [2]. In vitro downregulation of Smad2, Smad3 or the downstream Smad4 increased branching morphogenesis in lung explants cultures, similar to in vitro TgfßrII signaling inhibition [128]. Addition of exogenous Tgfß1 to these explant cultures did not reverse the phenotype, a finding consistent with Smad proteins being downstream of Tgfßs [136].

5 4. Bmp signaling

Bmp signaling plays an important role in regulating lung development although the exact regulatory effect is still controversial [137]. As mentioned above, Bmps are members of the Tgfß superfamily, but Bmp signaling acts through a particular type I receptors, the *activin receptor-like kinase 2, 3 and 6 (*Alk2, 3 and 6) and Smads (Smad1, 5, and 8) [137]. Several Bmp ligands are present in the developing lung, including Bmp3, 4, 5 and 7 [24, 138, 139]. In addition, most of the components of the Bmp signaling are expressed in early embryonic lung, like

the Bmp-specific Smads, which are predominantly expressed in the mesenchyme [140].

Bmp4 expression is first detected in the ventral foregut mesenchyme at the prospective lung region and in the primordial lung buds and is involved in the morphogenesis of the trachea and esophagus [24, 141]. Around 70% of mouse embryos deficient for *noggin*, a Bmp4 antagonist, presented EA/TOF together with defects in lung branching [24]. The reduction of Bmp4 expression by 50% rescued the phenotype of these mice, which suggests that normal morphogenesis of the foregut requires a tight control of the expression of Bmp4 by antagonists like noggin [24] Later in the development, Bmp4 is mainly detected in the distal epithelium of the developing lung buds. Bmp4 is not induced in the epithelium of the lungs during bud initiation but appears later, once the bud is elongating [139, 141, 142]. This spatial distribution of Bmp4 suggests its involvement in lung branching morphogenesis. Unfortunately, the conventional murine knockout for Bmp4 is early embryonic lethal, precluding the study of its *in vivo* function [143]. Bmp4 antagonizes the effect of Fgf10 in inducing epithelial bud outgrowth. Thus, Bmp4 acts as an inhibitor of budding. In this model, Bmp4 is induced at the tip of the growing lung buds in response to Fgf10 in the mesenchyme. Bmp4 will only start to inhibit Fgf10 at a specific concentration threshold, thus inhibiting further budding from the growing lung bud and ensuring a single bud. As Fgf10 expression at sites of future bud formation precedes Bmp4 induction, lung bud outgrowth will occur before Bmp4 reaches a inhibitory concentration [53, 97].This hypothesis was based on the following observations. First, overexpression of *Bmp4* in the distal epithelium of transgenic mice, using the *SPC* promoter, causes abnormal lung morphogenesis, with cystic terminal sacs and decreased epithelial proliferation [142]. Second, in mesenchyme-free lung epithelial cultures, recombinant Fgf10 induces budding and Bmp4 expression, while recombinant Bmp4 inhibits the Fgf10-mediated budding [53, 96].

In contrast to the previous results, it has been recently shown that Bmp4 is a strong inducer of branching in the presence of mesenchyme. The addition of exogenous Bmp4, at physiological concentrations, to intact embryonic lung explant cultures stimulates lung branching morphogenesis [124, 144]. Therefore, an alternative model has been proposed in which the negative or positive effect of Bmp4 on branching would depend on whether the signaling is activated via an autocrine or a paracrine mechanism. This means that the ability of the epithelium to respond to Bmp4 depends on the mesenchyme influence. When Bmp4 signaling is activated in the epithelium in an autocrine fashion, proliferation is inhibited. However, in the intact lung, Bmp4 present in the distal epithelium may also activate Bmp signaling in the adjacent mesenchyme (paracrine fashion). Bmp4 would then induce an unknown distal mesenchymal signal that increase proliferation of distal epithelial buds [124]. Bmp4 effects are, in turn, negatively regulated by the Bmp antagonists Noggin and Gremlin, which are expressed in distal lung mesenchyme[97] (**Figure 7**). In line with these data, Gremlin-null mutant mice present abnormal lung morphogenesis [145]. A recent study proposed an additional level in the regulation of Bmp4 by proteolysis. During branching morphogenesis, Fgf10 specifically induces the expression of the protease *Ctsh* (cathepsin H) in the lung epithelium, in an overlapping fashion to Bmp4 expression. It was shown that Ctsh inactivation produces an accumulation of Bmp4 in the epithelium, which disrupts lung branching [146].

Bmp4 has also been reported to promote distal lung differentiation [141]. A block of Bmp signaling at the lung bud, by *SPC* promoter-driven overexpression in the distal epithelium of Noggin (Bmp4 antagonist) or a dominant negative form of the *Alk6* Bmp type I receptor, results in reduction of distal epithelial cell phenotypes and an increased proximal cell phenotype [137, 141]. Regulation of Bmp4 expression in the epithelium is complex and dependent on signals such as Wnt, Fgf10, Pod1 (a transcriptional factor present in the mesenchyme) and Ttf1 [42, 147]. Both Bmp and Shh signaling regulate Fgf10 expression although they seem to function independently. The expression of Shh is maintained in Bmp4-overexpressing transgenic lungs [142], whereas, in Shh mutant lungs epithelial expression of Bmp4 is unaltered [82, 142]. One possibility is that they regulate different steps in the Fgf10 signaling. Bmp4 might be involved in modulating the activity of Fgf10 whereas Shh might be involved in restricting the Fgf10 expression domain. In addition, Fgf10 could possibly induce its own inhibitor, *Sprouty* (*Spry*), during lung branching [97].

5 5. Sprouty genes

Together with Bmp and Shh signaling, the Sprouty (Spry) pathway is also involved in regulating *Fgf10* expression. Spry protein family members function as negative regulators of Fgf signaling in lung development. Thus, Spry would inhibit the formation of specific signaling complexes downstream the Fgfr tyrosine-kinase cascade (Rtk-Ras-Erk/Mapk) by interacting with crucial factors (Frs2, Raf and Grb2). This results in the reduction of the Mapk activity, and the subsequent modulation of cell growth and differentiation during organogenesis [148]. Four mammalian Spry genes have been identified, and their expression patterns either overlap with, or are adjacent to, the known expression domains of one or more Fgf genes [149, 150].

Spry1 and 2 are localized at the distal epithelial tips of the growing lung and are down-regulated at sites of new bud formation, while Spry4 is predominantly present in the distal mesenchyme [151, 152]. There is a dynamic co-regulatory relationship between Fgf signaling and Spry during lung development. Expression of Fgf10 in the mesenchyme induces epithelial expression of *Spry2*, which acts as an inhibitor of Fgfr2b activity [152] (**Figure 7**). Consistent with this, reducing Spry2 activity results in increased lung branching morphogenesis and increased expression of specific lung epithelial maturation/differentiation markers [152]. Conversely, overexpression of *Spry2* in the distal lung epithelium causes lung hypoplasia and decreases branching morphogenesis[152, 153]. Furthermore, expression of *Spry* 2 and 4 is lost in *Fgf8* mutant mouse embryos [150]. All these data suggest that *Spry* genes are targets of Fgf signaling playing a vital role in regulating lung branching morphogenesis by modulating several signaling pathways to limit the effects of excessive Fgfr tyrosine kinase signaling [81]. Regulation of Fgf10 signaling by multiple overlapping pathways, such as Shh, Bmp and Spry signaling and Fgf inhibitors, might guarantee a tight, but still dynamic control of Fgf10 signaling [97].

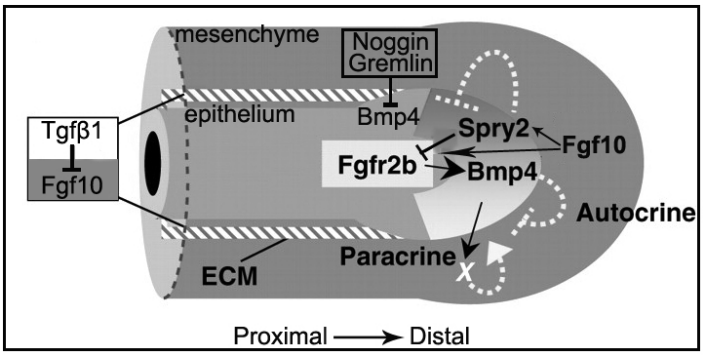


Figure 7. Control of Fgf10 and Fgfr2 expression by Tgfß and Bmp signaling, and Spry (adapted from [19]).

5 6. Wnt signaling

Several studies have suggested a role for the canonical Wnt signaling in regulating branching morphogenesis and P-D differentiation of airway epithelium [55, 154-156]. The Wnt family of secreted glycoproteins are involved in organogenesis, including cell proliferation, differentiation, and lineage specification [157]. Wnt proteins can signal through different pathways, of which the best understood is the canonical Wnt/ß-catenin pathway. This pathway involves the binding of secreted Wnt ligands to cell surface receptors of the frizzled (Fzd) family and the lipoprotein-receptor-related proteins LRP5/6. This binding will inactivate the protein glycogen synthase kinase-3ß (GSK-3ß), which phosphorylates ß-catenin, marking it for degradation. Hypophosphorylated ß-catenin translocates to the nucleus, where it binds to members of the Tcf/Lef1 transcription factor family to activate downstream target genes [158]. Several Wnt ligands, Fzd receptors and other components of the Wnt canonical pathway, such as ß-catenin and Tcf/Lef1 transcription factors, are expressed in embryonic and adult mouse lung. Their expression patterns imply that Wnt signaling can be initiated from the mesenchyme or the epithelium and can target both tissues in an autocrine and/or paracrine mode [159, 160].

The inactivation of two Wnt ligands, *Wnt5a and Wnt7b,* resulted in lung defects. Wnt7b is expressed in the distal epithelium of the lung bud tips [160-162]. *Wnt7b deficient lungs are* hypoplastic, but have a normal patterning and cell differentiation. The only apparent defect is a reduction in epithelial and mesenchymal proliferation[163]. This suggests that Wnt7b specifically controls lung growth by activating both autocrine (epithelium) and paracrine (mesenchyme) canonical Wnt signaling, which stimulates cell proliferation [163].The activation of the canonical Wnt signaling can be monitored by detection of nuclear ß-catenin. In the E11-13 lung, ß-catenin is expressed throughout the entire lung epithelium, but is nuclear only in the distal lung epithelium, i.e. at the sites that are actively branching [154, 155, 159, 164]. Inhibition of canonical Wnt signaling at these sites, either by tissue-specific deletion of ß-catenin or by targeted expression of the Wnt antagonist dickkopf1 (Dkk1), results in disruption of branching morphogenesis with no distal airway differentiation and expansion of proximal airway [154, 155, 165]. It has

been suggested that Wnt/beta-catenin signaling is a critical upstream regulator of P-D patterning in the lung, in part through regulation of Bmp4, Nmyc, and Fgf signaling. Moreover, it was shown that ß-catenin and Tcf/Lef1 activate the promoters of *Bmp4* and *Nmyc* [155]. Strikingly, hyperactive Wnt signaling, by expressing a constitutive active form of ß-catenin in proximal airway epithelium, changes the developmental potential of embryonic lung endoderm to express markers of gut endoderm differentiation [164].

The results above conflict with that of mice lacking the *Wnt5a* gene [55]. *Wnt5a* is expressed in both developing pulmonary epithelium and mesenchyme, with highest expression in distal and branching tips. Transgenic mice overexpressing Wnt5a in the lung epithelium, under the controls of the SP-C promoter, presented lungs with reduced branching and dilated distal airways. The expression of Fgf10 was not properly regulated and the expression of Shh was downregulated. *In vitro* experiments showed that the Wnt5a overexpressing epithelium was unable to respond to Fgf10 signaling [166]. Loss of Wnt5a leads to perinatal death from lung defects including truncation of the trachea, increased cell proliferation with a resulting expansion of the distal airways, and delayed lung maturation [55]. In contrast, mice deficient for ß-catenin showed inhibition of distal airway differentiation. Absence of Wnt5a activity in the lungs also leads to an up-regulation of the expression of Fgf10, Bmp4 and Shh, which suggests that Wnt5a may interact with this signaling pathways [55]. It is possible that the discrepancy in results is due to the fact that different Wnt family members may have distinct functions, either acting as positive or negative regulators of branching and cell proliferation [19]. In addition, recent studies have shown that Wnt-ß-Catenin signaling has an essential role in regulating cell proliferation and differentiation in the lung mesenchyme [167, 168] and in regulating lung stem cell population by stimulating the proliferation of undifferentiated progenitors (see next section)[156].

5 7. Gata6

Gata6 is a member of the Gata family of zinc-finger transcription factors. Gata6 is expressed in epithelial cells of the developing lung bud and in respiratory epithelium throughout lung morphogenesis. It presents a spatial pattern similar to that of Foxa2 and Ttf1 [169, 170]. Targeted deletion of the *Gata6* gene is lethal in early embryonic development, prior to lung bud formation, due to failure in visceral endoderm formation [171, 172]. A recent study in mice with a specific deletion of *Gata6* in the airway epithelium, established that Gata6 regulates the expansion and differentiation of bronchoalveolar stem cells (BASCs) in the lung [156]. Several studies have focused on the identification of the lung stem cell niches and understanding the transcriptional network that controls the balance between their expansion and differentiation, which is essential during lung development and regeneration after injury [173] . The BASCs represent one of these progenitor cell populations, are localized at the alveolar–airway junction and express both CC10 and Sftpc epithelial cell markers, in addition to other characteristic stem cells markers like Sca1 [174]. In the absence of Gata6, the expression of Fzd2 (the noncanonical Wnt receptor) was downregulated, which activates the canonical Wnt-ß-catenin signaling in

the airway epithelium. This leads to an ectopic and premature increase in the number of BASCs and the subsequent loss of epithelial differentiation [156]. In addition, the specific deletion of Gata6 in the airway epithelium resulted in mutant lungs with increased airway dilation and decreased branching morphogenesis [156]. Finally, overexpression of Gata6, using an SPC promoter, also disrupted branching and caused a lack of distal epithelial cell differentiation[175], probably due to the inhibition of the canonical Wnt-ß-catenin signaling .

6. Molecular control of lung epithelial differentiation

The different cell types present in the lung epithelium vary during the development along the P-D axis of the lung. These cell types must be present in appropriate numbers and sites to support postnatal respiration. The current knowledge of the mechanisms controlling differentiation and proliferation of the multiple cell types derived from the pulmonary epithelium during lung morphogenesis is limited. There are autocrine-paracrine and cell-cell interactions that regulate transcription factors controlling cell fate [2]. The regional differentiation of the respiratory epithelium is also specified by the mesenchyme [176] and is dependent on the P-D polarity exhibited by the airways. Many of the already described signaling pathways have been implicated in the differentiation of lung airway epithelium. These pathways act individually and in concert with each other to regulate epithelial and mesenchymal differentiation. Gene targeting studies in mice have identified some of the genes involved in this process. In many cases the differentiation defect is associated with earlier defects in branching morphogenesis, making the analysis of their precise function more difficult.

Concerted actions of various transcription factors like Ttf1, Gata6, p63, Nmyc, Lef/Tcf1 (together with ß-catenin) and transcription factors members of the Hox, Sox, and Fox families [2] play roles in cell-specific differentiation of the airways (see [2, 57, 95, 177, 178] for review). They are expressed at different levels in distinct cell types along the P-D axis of the lung epithelium, where they influence gene expression and epithelial cell differentiation [2] and they may also be controlled by epithelial-mesenchymal interactions. The function of some of these transcription factors in the different cell types will be described briefly below.

6 1. Foxj1 in ciliated cells

Foxj1 is a transcription factor of the winged helix/forkhead family and is expressed in different tissues during development. In the lung, it is associated with the proximal differentiation of the conducting airways and is specifically expressed in ciliated cells [179-181]. Foxj1 is required for ciliogenesis and regulates a number of genes necessary for the cilia formation [182, 183]. Transgenic mice with ectopic overexpression of Foxj1 in the distal epithelium present defects in both branching morphogenesis and epithelium differentiation. They have ciliated cells in the alveolar epithelium, which normally are only present in the proximal conducting airways[179]. In contract, mice mutant for Foxj1 lack cilia formation and have random left-right lung lobulation, as Foxj1 is also involved is the control of left-right asymmetry [184, 185].

6 2. p63 in basal cells

The p63 transcription factor is a homolog of the tumor suppressor protein p53 [186]. The p63 gene encodes at least six different isoforms [186, 187]. The transactivating (TA) isoforms contain a transactivation domain (TAD), whereas the truncated isoforms (ΔNp63) lack the TAD. The TA isoforms are efficient activators of transcription [186, 187] and the ΔN isoforms generally function as repressors of transcription and can act as dominant-negative factors by inhibiting the activity of p53 in cell cycle arrest and apoptosis [188]. p63 is particularly highly expressed in the progenitor cell populations of a variety of epithelial tissues, including basal cells of the upper airways [189]. Deletion of p63 in mice results in formation of a simple columnar epithelium, deficient in basal cells. In contrast, ectopic expression of a p63 isoform that contains only the TAD domain caused squamous metaplasia and induced keratin-14 expression, a marker of stratified epithelia [190]. Thus p63 plays a critical role in the regulation of progenitor basal cells and it is necessary for the normal development of the pseudostratified epithelium of trachea and bronchi [189, 191] (the role of p63 in airway epithelial differentiation will be further discussed in the **chapter 3** of this thesis).

6 3. Nmyc

Nmyc is a member of the Myc family of proto-oncogenes**.** *Nmyc* expression in the embryonic lung is restricted to a distal population of undifferentiated epithelial cells [192]. *Nmyc*-null mice die at midgestation from cardiac failure, before the formation of the lung . However, the importance of Nmyc for lung development was confirmed in embryos with a partial deletion of Nmyc, as they die at birth from respiratory distress and present lung hypoplasia, disrupted branching and enlarged air spaces [193-196].

During lung development there is a balance between the maintenance and expansion of a pool of uncommitted progenitor cells, and their terminal differentiation. Overexpression of Nmyc in the epithelium, using the SPC promoter, inhibits differentiation and increases proliferation. By contrast, conditional deletion of Nmyc expression in the distal epithelium induces premature differentiation and inhibits proliferation. These data suggested that Nmyc is not required for lung bud formation but is essential for maintaining proliferation of distal undifferentiated progenitor cells during lung morphogenesis [192]. As previously mentioned, inhibition of Wnt canonical signaling in mutant mice caused a similar lung phenotype and it was shown that Wnt controls the levels of Nmyc and Bmp4 expression in the distal lung [155].

6 4. Notch signaling in pulmonary neuroepithelial cells

The Notch signaling pathway is evolutionarily conserved and regulates local cell-cell interaction and thereby cell fate. Notch signaling is essential for the maintenance of dividing cells and subsequent generation of cell type diversity throughout mammalian development [197]. A number of transcription factors downstream of the Notch signaling, such as the basic helix-loop-helix (bHLH) proteins Mash1 and Hes1, are known to affect pulmonary neuroendocrine development [198, 199]. However the mechanisms regulating

neuroendocrine differentiation of the lung epithelia remains to be elucidated. Mash1, the mammalian homolog of achaete-scute complex is a proneural factor, which is expressed in clusters of pulmonary neural epithelial cells (PNECs) or single PNECs and their precursors[198, 200, 201]. Hes1 (Hairy and enhancer of split1) is expressed in pulmonary epithelial cells other than PNECs and represses neural differentiation through suppression of Mash1 [198, 202]. *Mash1*-deficient mice have no detectable pulmonary neuroendocrine cells and lack of expression of neuroendocrine markers including cGRP [198, 200]. Increased expression of *Mash-1* in non-endocrine airway epithelial cells, using the CC10 promoter, caused progressive airway hyperplasia and metaplasia and the double transgenic animals developed massive neuroendocrine lung tumors [203]. Conversely, Hes-deficient mice have increased numbers of PNEC and increased Mash1 and NeuroD expression while Notch-1 is suppressed. Thus, Mash1 and Hes1 regulate neuroendocrine versus non-neuroendocrine cell fate specification in the lung epithelium via Notch signaling [198, 201, 204, 205]. They play an essential role in the differentiation of PNECs, but not in lung branching morphogenesis, as the lung gross morphology in mice deficient for either factor appears normal [198, 200]. In addition, increased numbers of neuroendocrine cells were observed in the airways of mice overexpressing Sox2 (this thesis, [206]) and Sox17[207], indicating their potential roles in PNEC differentiation.

The Shh receptor Ptc was detected in the basal layer of the adult bronchial epithelium where it was co-localized with the PNEC marker cGRP. This suggest that PNECs might be responsive to Shh signaling from adjacent airway epithelial cells [86].

6 5. Sox family

A number of Sox genes are expressed in the developing lung, including Sox2, -4, -9, -11, 12, -17 and -21 [208, 209]. Sox9 is expressed in the respiratory epithelium and in the proximal mesenchyme throughout lung morphogenesis [209]. Sox9 expression changes dynamically during the course of lung development, suggesting a role for Sox9 in lung morphogenesis. Heterozygous mutations in human Sox9 gene cause campomelic dysplasia syndrome. Some patients die in the neonatal period due to respiratory distress or survive having recurrent lung infections [210]. Heterozygous Sox9 mutant mice present a similar phenotype and die shortly after birth from respiratory failure [211, 212]. Heterozygous Sox9 mice also have multiple anomalies of the skeleton and cartilage and have severe tracheal cartilage malformations [213]. These phenotypes suggest that Sox9 could be involved in lung development or lung function. Surprisingly, deletion of Sox9 in the respiratory epithelium did not alter lung morphogenesis or function [209]. One possible explanation of these data could be that the respiratory problems in the Sox9 heterozygous mice are caused by skeletal defects.

Sox11-deficient mice die at birth from congenital cyanosis caused by severe heart defects and present lung hypoplasia among other defects [214, 215]. Thus, Sox11 is required for proper lung formation but it is not known which genes and processes it regulates in lung formation regulate [2]. Another member of the SoxC group, Sox12, is widely expressed during embryonic development, including in the lung, but the Sox12-deficient mice

developed normally. There is probably functional redundancy with other SoxC members expressed in the lung (Sox4 and Sox11) that can compensate for its loss during mouse lung development [216].

Sox17 is required for the formation of the definitive endoderm that gives rise to various organs, including the lung [207]. Sox17 interacts with ß-catenin in the regulation of transcription of endodermal genes and is required for anterior-posterior axis formation in *Xenopus* [217]. Sox17 expression is first detected at E18 in the proximal epithelium of the conducting airways in the mouse. Thereafter, Sox17 expression is restricted primarily to ciliated cells [207]. Conditional ectopic overexpression of Sox17 in distal respiratory epithelium of embryonic lung disrupted peripheral lung formation. Consistent with its expression in proximal airway cells, Sox17 inhibited Sftpc and enhanced Foxj1 promoter activity *in vitro*[207]. When overexpressed in the adult lung, Sox17 induced hyperplastic clusters of cells with ectopic expression of proximal airway markers, supporting its potential role in repair and redifferentiation of the bronchiolar epithelium, following injury [218].

The dynamic expression pattern of Sox2 during lung development was studied in chick and mouse embryos [206, 208]. Initially, it is expressed in the rostral gut epithelium from the pharynx to the stomach [208], and during lung morphogenesis, Sox2 expression is switched off in the epithelia invaginating into the surrounding mesenchyme to form the primordium of the lung [206, 208]. This lead to the hypothesis that local downregulation of Sox2 may be necessary for initiation of lung bud morphogenesis. As soon as the lung bud is established, Sox2 expression is upregulated again in the epithelium of the primary bronchi. During lung branching morphogenesis, Sox2 is associated with the epithelium that is less morphogenetically active, and expression is lost at sites where the bronchial tree branches [208]. In the adult lung, Sox2 remains highly expressed in the proximal airway epithelium. Thus, loss of Sox2 expression in the epithelium seems to correlate with the initiation of morphogenesis making the epithelium receptive to branching inducing signals [208]. Since Sox2-deficient mice die early in development, before lung organogenesis [26], a target inactivation of Sox2 in the lung epithelium will help to understand its role in lung development. Interestingly, mouse embryos with overexpression of Sox2 in distal airway epithelium, under control of the SPC promoter, die at birth due to respiratory distress. The transgenic lungs presented inhibition of branching morphogenesis and defects in epithelium differentiation, with an increase number of PNECs and basal cells (**chapter 3** of this thesis, [206]).

References

1. Whitsett, J.A., Wert, S.E. & Trapnell, B.C. Genetic disorders influencing lung formation and function at birth. *Hum Mol Genet* 13 Spec No 2, R207-215 (2004).
2. Maeda, Y., Dave, V. & Whitsett, J.A. Transcriptional control of lung morphogenesis. *Physiol Rev* 87, 219-244 (2007).
3. Groenman, F., Unger, S. & Post, M. The molecular basis for abnormal human lung development. *Biol Neonate* 87, 164-177 (2005).
4. Galambos, C. & Demello, D.E. Regulation of alveologenesis: clinical implications of impaired growth. *Pathology* 40, 124-140 (2008).

5. Galambos, C. & Demello, D. Molecular mechanisms of pulmonary vascular development. *Pediatr Dev Pathol*, 1 (2007).
6. Hogan, B.L. Morphogenesis. *Cell* 96, 225-233 (1999).
7. Burri, P.H. Structural aspects of postnatal lung development - alveolar formation and growth. *Biol Neonate* 89, 313-322 (2006).
8. Ten Have-Opbroek, A.A. Lung development in the mouse embryo. *Exp Lung Res* 17, 111-130 (1991).
9. Warburton, D. *et al*. The molecular basis of lung morphogenesis. *Mech Dev* 92, 55-81 (2000).
10. Inselman, L.S. & Mellins, R.B. Growth and development of the lung. *J Pediatr* 98, 1-15 (1981).
11. Van Lommel, A., Bolle, T., Fannes, W. & Lauweryns, J.M. The pulmonary neuroendocrine system: the past decade. *Arch Histol Cytol* 62, 1-16 (1999).
12. Sunday, M.E. Pulmonary Neuroendocrine Cells and Lung Development. *Endocr Pathol* 7, 173-201 (1996).
13. Cutz, E. Neuroendocrine cells of the lung. An overview of morphologic characteristics and development. *Exp Lung Res* 3, 185-208 (1982).
14. Warburton, D., Wuenschell, C., Flores-Delgado, G. & Anderson, K. Commitment and differentiation of lung cell lineages. *Biochem Cell Biol* 76, 971-995 (1998).
15. Warburton, D., Zhao, J., Berberich, M.A. & Bernfield, M. Molecular embryology of the lung: then, now, and in the future. *Am J Physiol* 276, L697-704 (1999).
16. Williams, M.C. & Mason, R.J. Development of the type II cell in the fetal rat lung. *Am Rev Respir Dis* 115, 37-47 (1977).
17. Rosenbaum, R.M., Picciano, P., Kress, Y. & Wittner, M. Ultrastructure of in vitro type 2 epithelial cell cysts derived from adult rabbit lung cells. *Anat Rec* 188, 241-261 (1977).
18. Wells, J.M. & Melton, D.A. Vertebrate endoderm development. *Annu Rev Cell Dev Biol* 15, 393-410 (1999).
19. Cardoso, W.V. & Lu, J. Regulation of early lung morphogenesis: questions, facts and controversies. *Development* 133, 1611-1624 (2006).
20. Mizuno, T. & Yasugi, S. Susceptibility of epithelia to directive influences of mesenchymes during organogenesis: uncoupling of morphogenesis and cytodifferentiation. *Cell Differ Dev* 31, 151-159 (1990).
21. Kumar, M. & Melton, D. Pancreas specification: a budding question. *Curr Opin Genet Dev* 13, 401-407 (2003).
22. Bort, R., Martinez-Barbera, J.P., Beddington, R.S. & Zaret, K.S. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* 131, 797-806 (2004).
23. Serls, A.E., Doherty, S., Parvatiyar, P., Wells, J.M. & Deutsch, G.H. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development* 132, 35-47 (2005).
24. Que, J., Choi, M., Ziel, J.W., Klingensmith, J. & Hogan, B.L. Morphogenesis of the trachea and esophagus: current players and new roles for noggin and Bmps. *Differentiation* 74, 422-437 (2006).
25. Li, Y., Litingtung, Y., Ten Dijke, P. & Chiang, C. Aberrant Bmp signaling and notochord delamination in the pathogenesis of esophageal atresia. *Dev Dyn* 236, 746-754 (2007).
26. Avilion, A.A. *et al*. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 17, 126-140 (2003).
27. Minoo, P., Su, G., Drum, H., Bringas, P. & Kimura, S. Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. *Dev Biol* 209, 60-71 (1999).
28. Lazzaro, D., Price, M., de Felice, M. & Di Lauro, R. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development* 113, 1093-1104 (1991).
29. Kimura, S. *et al*. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev* 10, 60-69 (1996).
30. Wert, S.E., Glasser, S.W., Korfhagen, T.R. & Whitsett, J.A. Transcriptional elements from the human SP-C gene direct expression in the primordial respiratory epithelium of transgenic mice. *Dev Biol* 156, 426-443 (1993).
31. Bohinski, R.J., Di Lauro, R. & Whitsett, J.A. The lung-specific surfactant protein B gene promoter is a target for thyroid transcription factor 1 and hepatocyte nuclear factor 3, indicating common factors for organ-specific gene expression along the foregut axis. *Mol Cell Biol* 14, 5671-5681 (1994).
32. Di Palma, T. *et al*. The paired domain-containing factor Pax8 and the homeodomain-containing factor TTF-1 directly interact and synergistically activate transcription. *J Biol Chem* 278, 3395-3402 (2003).
33. Genevieve, D., de Pontual, L., Amiel, J., Sarnacki, S. & Lyonnet, S. An overview of isolated and syndromic oesophageal atresia. *Clin Genet* 71, 392-399 (2007).
34. Shannon, J.M. Induction of alveolar type II cell differentiation in fetal tracheal epithelium by grafted distal lung mesenchyme. *Dev Biol* 166, 600-614 (1994).

35. Lawson, K.A. Stage specificity in the mesenchyme requirement of rodent lung epithelium in vitro: a matter of growth control? *J Embryol Exp Morphol* 74, 183-206 (1983).
36. Wessells, N.K. Mammalian lung development: interactions in formation and morphogenesis of tracheal buds. *J Exp Zool* 175, 455-466 (1970).
37. Zhou, L., Lim, L., Costa, R.H. & Whitsett, J.A. Thyroid transcription factor-1, hepatocyte nuclear factor-3beta, surfactant protein B, C, and Clara cell secretory protein in developing mouse lung. *J Histochem Cytochem* 44, 1183-1193 (1996).
38. Zhou, H. *et al*. Expression of thyroid transcription factor-1, surfactant proteins, type I cell-associated antigen, and Clara cell secretory protein in pulmonary hypoplasia. *Pediatr Dev Pathol* 4, 364-371 (2001).
39. Stahlman, M.T., Gray, M.E. & Whitsett, J.A. Expression of thyroid transcription factor-1(TTF-1) in fetal and neonatal human lung. *J Histochem Cytochem* 44, 673-678 (1996).
40. Yuan, B. *et al*. Inhibition of distal lung morphogenesis in Nkx2.1(-/-) embryos. *Dev Dyn* 217, 180-190 (2000).
41. Kimura, S., Ward, J.M. & Minoo, P. Thyroid-specific enhancer-binding protein/thyroid transcription factor 1 is not required for the initial specification of the thyroid and lung primordia. *Biochimie* 81, 321-327 (1999).
42. Zhu, N.L., Li, C., Xiao, J. & Minoo, P. NKX2.1 regulates transcription of the gene for human bone morphogenetic protein-4 in lung epithelial cells. *Gene* 327, 25-36 (2004).
43. Bellusci, S., Grindley, J., Emoto, H., Itoh, N. & Hogan, B.L. Fibroblast growth factor 10 (FGF10) and branching morphogenesis in the embryonic mouse lung. *Development* 124, 4867-4878 (1997).
44. Colvin, J.S., White, A.C., Pratt, S.J. & Ornitz, D.M. Lung hypoplasia and neonatal death in Fgf9-null mice identify this gene as an essential regulator of lung mesenchyme. *Development* 128, 2095-2106 (2001).
45. White, A.C. *et al*. FGF9 and SHH signaling coordinate lung growth and development through regulation of distinct mesenchymal domains. *Development* 133, 1507-1517 (2006).
46. Peters, K.G., Werner, S., Chen, G. & Williams, L.T. Two FGF receptor genes are differentially expressed in epithelial and mesenchymal tissues during limb formation and organogenesis in the mouse. *Development* 114, 233-243 (1992).
47. Arman, E., Haffner-Krausz, R., Gorivodsky, M. & Lonai, P. Fgfr2 is required for limb outgrowth and lung-branching morphogenesis. *Proc Natl Acad Sci U S A* 96, 11895-11899 (1999).
48. Sekine, K. *et al*. Fgf10 is essential for limb and lung formation. *Nat Genet* 21, 138-141 (1999).
49. Min, H. *et al*. Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to Drosophila branchless. *Genes Dev* 12, 3156-3161 (1998).
50. De Moerlooze, L. *et al*. An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal-epithelial signalling during mouse organogenesis. *Development* 127, 483-492 (2000).
51. Peters, K. *et al*. Targeted expression of a dominant negative FGF receptor blocks branching morphogenesis and epithelial differentiation of the mouse lung. *Embo J* 13, 3296-3301 (1994).
52. Park, W.Y., Miranda, B., Lebeche, D., Hashimoto, G. & Cardoso, W.V. FGF-10 is a chemotactic factor for distal epithelial buds during lung development. *Dev Biol* 201, 125-134 (1998).
53. Weaver, M., Dunn, N.R. & Hogan, B.L. Bmp4 and Fgf10 play opposing roles during lung bud morphogenesis. *Development* 127, 2695-2704 (2000).
54. Chen, F. *et al*. Inhibition of Tgf beta signaling by endogenous retinoic acid is essential for primary lung bud induction. *Development* 134, 2969-2979 (2007).
55. Li, C., Xiao, J., Hormi, K., Borok, Z. & Minoo, P. Wnt5a participates in distal lung morphogenesis. *Dev Biol* 248, 68-81 (2002).
56. Sakiyama, J., Yamagishi, A. & Kuroiwa, A. Tbx4-Fgf10 system controls lung bud formation during chicken embryonic development. *Development* 130, 1225-1234 (2003).
57. Perl, A.K. & Whitsett, J.A. Molecular mechanisms controlling lung morphogenesis. *Clin Genet* 56, 14-27 (1999).
58. Ross, S.A., McCaffery, P.J., Drager, U.C. & De Luca, L.M. Retinoids in embryonal development. *Physiol Rev* 80, 1021-1054 (2000).
59. Duester, G. Genetic dissection of retinoid dehydrogenases. *Chem Biol Interact* 130-132, 469-480 (2001).
60. Zile, M.H. *et al*. Retinoid signaling is required to complete the vertebrate cardiac left/right asymmetry pathway. *Dev Biol* 223, 323-338 (2000).
61. Dickman, E.D., Thaller, C. & Smith, S.M. Temporally-regulated retinoic acid depletion produces specific neural crest, ocular and nervous system defects. *Development* 124, 3111-3121 (1997).
62. Perissi, V. & Rosenfeld, M.G. Controlling nuclear receptors: the circular logic of cofactor cycles. *Nat Rev Mol Cell Biol* 6, 542-554 (2005).

63. Malpel, S., Mendelsohn, C. & Cardoso, W.V. Regulation of retinoic acid signaling during lung morphogenesis. *Development* 127, 3057-3067 (2000).
64. Wang, Z., Dolle, P., Cardoso, W.V. & Niederreither, K. Retinoic acid regulates morphogenesis and patterning of posterior foregut derivatives. *Dev Biol* 297, 433-445 (2006).
65. Desai, T.J., Malpel, S., Flentke, G.R., Smith, S.M. & Cardoso, W.V. Retinoic acid selectively regulates Fgf10 expression and maintains cell identity in the prospective lung field of the developing foregut. *Dev Biol* 273, 402-415 (2004).
66. Desai, T.J. *et al*. Distinct roles for retinoic acid receptors alpha and beta in early lung morphogenesis. *Dev Biol* 291, 12-24 (2006).
67. Wilson, J.G., Roth, C.B. & Warkany, J. An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. Effects of restoration of vitamin A at various times during gestation. *Am J Anat* 92, 189-217 (1953).
68. Mollard, R., Ghyselinck, N.B., Wendling, O., Chambon, P. & Mark, M. Stage-dependent responses of the developing lung to retinoic acid signaling. *Int J Dev Biol* 44, 457-462 (2000).
69. Lum, L. & Beachy, P.A. The Hedgehog response network: sensors, switches, and routers. *Science* 304, 1755-1759 (2004).
70. Mendelsohn, C. *et al*. Function of the retinoic acid receptors (RARs) during development (II). Multiple abnormalities at various stages of organogenesis in RAR double mutants. *Development* 120, 2749-2771 (1994).
71. Chapman, D.L. *et al*. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev Dyn* 206, 379-390 (1996).
72. Lindsay, E.A. *et al*. Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* 410, 97-101 (2001).
73. Naiche, L.A. & Papaioannou, V.E. Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. *Development* 130, 2681-2693 (2003).
74. Bruneau, B.G. *et al*. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* 106, 709-721 (2001).
75. Cebra-Thomas, J.A. *et al*. T-box gene products are required for mesenchymal induction of epithelial branching in the embryonic mouse lung. *Dev Dyn* 226, 82-90 (2003).
76. Bogue, C.W. *et al*. Identification of Hox genes in newborn lung and effects of gestational age and retinoic acid on their expression. *Am J Physiol* 266, L448-454 (1994).
77. Bogue, C.W., Lou, L.J., Vasavada, H., Wilson, C.M. & Jacobs, H.C. Expression of Hoxb genes in the developing mouse foregut and lung. *Am J Respir Cell Mol Biol* 15, 163-171 (1996).
78. Volpe, M.V., Martin, A., Vosatka, R.J., Mazzoni, C.L. & Nielsen, H.C. Hoxb-5 expression in the developing mouse lung suggests a role in branching morphogenesis and epithelial cell fate. *Histochem Cell Biol* 108, 495-504 (1997).
79. Aubin, J., Lemieux, M., Tremblay, M., Berard, J. & Jeannotte, L. Early postnatal lethality in Hoxa-5 mutant mice is attributable to respiratory tract defects. *Dev Biol* 192, 432-445 (1997).
80. Joksimovic, M., Jeannotte, L. & Tuggle, C.K. Dynamic expression of murine HOXA5 protein in the central nervous system. *Gene Expr Patterns* 5, 792-800 (2005).
81. Warburton, D. *et al*. Molecular mechanisms of early lung specification and branching morphogenesis. *Pediatr Res* 57, 26R-37R (2005).
82. 82. Pepicelli, C.V., Lewis, P.M. & McMahon, A.P. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr Biol* 8, 1083-1086 (1998).
83. Weaver, M., Batts, L. & Hogan, B.L. Tissue interactions pattern the mesenchyme of the embryonic mouse lung. *Dev Biol* 258, 169-184 (2003).
84. Bellusci, S. *et al*. Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. *Development* 124, 53-63 (1997).
85. Hui, C.C., Slusarski, D., Platt, K.A., Holmgren, R. & Joyner, A.L. Expression of three mouse homologs of the Drosophila segment polarity gene cubitus interruptus, Gli, Gli-2, and Gli-3, in ectoderm- and mesoderm-derived tissues suggests multiple roles during postimplantation development. *Dev Biol* 162, 402-413 (1994).
86. Watkins, D.N. *et al*. Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* 422, 313-317 (2003).
87. Litingtung, Y., Lei, L., Westphal, H. & Chiang, C. Sonic hedgehog is essential to foregut development. *Nat Genet* 20, 58-61 (1998).

88. Motoyama, J. *et al*. Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat Genet* 20, 54-57 (1998).
89. Grindley, J.C., Bellusci, S., Perkins, D. & Hogan, B.L. Evidence for the involvement of the Gli gene family in embryonic mouse lung development. *Dev Biol* 188, 337-348 (1997).
90. Bai, C.B. & Joyner, A.L. Gli1 can rescue the in vivo function of Gli2. *Development* 128, 5161-5172 (2001).
91. Park, H.L. *et al*. Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. *Development* 127, 1593-1605 (2000).
92. Mahlapuu, M., Enerback, S. & Carlsson, P. Haploinsufficiency of the forkhead gene Foxf1, a target for sonic hedgehog signaling, causes lung and foregut malformations. *Development* 128, 2397-2406 (2001).
93. Kalinichenko, V.V. *et al*. Defects in pulmonary vasculature and perinatal lung hemorrhage in mice heterozygous null for the Forkhead Box f1 transcription factor. *Dev Biol* 235, 489-506 (2001).
94. Lim, L., Kalinichenko, V.V., Whitsett, J.A. & Costa, R.H. Fusion of lung lobes and vessels in mouse embryos heterozygous for the forkhead box f1 targeted allele. *Am J Physiol Lung Cell Mol Physiol* 282, L1012-1022 (2002).
95. Costa, R.H., Kalinichenko, V.V. & Lim, L. Transcription factors in mouse lung development and function. *Am J Physiol Lung Cell Mol Physiol* 280, L823-838 (2001).
96. Lebeche, D., Malpel, S. & Cardoso, W.V. Fibroblast growth factor interactions in the developing lung. *Mech Dev* 86, 125-136 (1999).
97. Chuang, P.T. & McMahon, A.P. Branching morphogenesis of the lung: new molecular insights into an old problem. *Trends Cell Biol* 13, 86-91 (2003).
98. Metzger, R.J., Klein, O.D., Martin, G.R. & Krasnow, M.A. The branching programme of mouse lung development. *Nature* 453, 745-750 (2008).
99. DeFelice, M. *et al*. TTF-1 phosphorylation is required for peripheral lung morphogenesis, perinatal survival, and tissue-specific gene expression. *J Biol Chem* 278, 35574-35583 (2003).
100. Dave, V., Childs, T. & Whitsett, J.A. Nuclear factor of activated T cells regulates transcription of the surfactant protein D gene (Sftpd) via direct interaction with thyroid transcription factor-1 in lung epithelial cells. *J Biol Chem* 279, 34578-34588 (2004).
101. Kumar, A.S., Venkatesh, V.C., Planer, B.C., Feinstein, S.I. & Ballard, P.L. Phorbol ester down-regulation of lung surfactant protein B gene expression by cytoplasmic trapping of thyroid transcription factor-1 and hepatocyte nuclear factor 3. *J Biol Chem* 272, 20764-20773 (1997).
102. Shaw-White, J.R., Bruno, M.D. & Whitsett, J.A. GATA-6 activates transcription of thyroid transcription factor-1. *J Biol Chem* 274, 2658-2664 (1999).
103. Ikeda, K., Shaw-White, J.R., Wert, S.E. & Whitsett, J.A. Hepatocyte nuclear factor 3 activates transcription of thyroid transcription factor 1 in respiratory epithelial cells. *Mol Cell Biol* 16, 3626-3636 (1996).
104. Ikeda, K. *et al*. Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells. *J Biol Chem* 270, 8108-8114 (1995).
105. Wert, S.E., Dey, C.R., Blair, P.A., Kimura, S. & Whitsett, J.A. Increased expression of thyroid transcription factor-1 (TTF-1) in respiratory epithelial cells inhibits alveolarization and causes pulmonary inflammation. *Dev Biol* 242, 75-87 (2002).
106. Stahlman, M.T., Gray, M.E. & Whitsett, J.A. Temporal-spatial distribution of hepatocyte nuclear factor-3beta in developing human lung and other foregut derivatives. *J Histochem Cytochem* 46, 955-962 (1998).
107. Monaghan, A.P., Kaestner, K.H., Grau, E. & Schutz, G. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* 119, 567-578 (1993).
108. Besnard, V., Wert, S.E., Hull, W.M. & Whitsett, J.A. Immunohistochemical localization of Foxa1 and Foxa2 in mouse embryos and adult tissues. *Gene Expr Patterns* 5, 193-208 (2004).
109. Kaestner, K.H., Katz, J., Liu, Y., Drucker, D.J. & Schutz, G. Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. *Genes Dev* 13, 495-504 (1999).
110. Shih, D.Q., Navas, M.A., Kuwajima, S., Duncan, S.A. & Stoffel, M. Impaired glucose homeostasis and neonatal mortality in hepatocyte nuclear factor 3alpha-deficient mice. *Proc Natl Acad Sci U S A* 96, 10152-10157 (1999).
111. Besnard, V., Wert, S.E., Kaestner, K.H. & Whitsett, J.A. Stage-specific regulation of respiratory epithelial cell differentiation by Foxa1. *Am J Physiol Lung Cell Mol Physiol* 289, L750-759 (2005).
112. Ang, S.L. & Rossant, J. HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* 78, 561-574 (1994).

113. Weinstein, D.C. *et al*. The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell* 78, 575-588 (1994).
114. Wan, H. *et al*. Foxa2 is required for transition to air breathing at birth. *Proc Natl Acad Sci U S A* 101, 14449-14454 (2004).
115. Wan, H. *et al*. Foxa2 regulates alveolarization and goblet cell hyperplasia. *Development* 131, 953-964 (2004).
116. Wan, H. *et al*. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* 280, 13809-13816 (2005).
117. Balemans, W. & Van Hul, W. Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev Biol* 250, 231-250 (2002).
118. ten Dijke, P. & Hill, C.S. New insights into TGF-beta-Smad signalling. *Trends Biochem Sci* 29, 265-273 (2004).
119. Wrana, J.L., Attisano, L., Wieser, R., Ventura, F. & Massague, J. Mechanism of activation of the TGF-beta receptor. *Nature* 370, 341-347 (1994).
120. Pelton, R.W., Johnson, M.D., Perkett, E.A., Gold, L.I. & Moses, H.L. Expression of transforming growth factor-beta 1, -beta 2, and -beta 3 mRNA and protein in the murine lung. *Am J Respir Cell Mol Biol* 5, 522-530 (1991).
121. Pelton, R.W., Saxena, B., Jones, M., Moses, H.L. & Gold, L.I. Immunohistochemical localization of TGF beta 1, TGF beta 2, and TGF beta 3 in the mouse embryo: expression patterns suggest multiple roles during embryonic development. *J Cell Biol* 115, 1091-1105 (1991).
122. Heine, U.I., Munoz, E.F., Flanders, K.C., Roberts, A.B. & Sporn, M.B. Colocalization of TGF-beta 1 and collagen I and III, fibronectin and glycosaminoglycans during lung branching morphogenesis. *Development* 109, 29-36 (1990).
123. Zhou, L., Dey, C.R., Wert, S.E. & Whitsett, J.A. Arrested lung morphogenesis in transgenic mice bearing an SP-C-TGF-beta 1 chimeric gene. *Dev Biol* 175, 227-238 (1996).
124. Bragg, A.D., Moses, H.L. & Serra, R. Signaling to the epithelium is not sufficient to mediate all of the effects of transforming growth factor beta and bone morphogenetic protein 4 on murine embryonic lung development. *Mech Dev* 109, 13-26 (2001).
125. Serra, R. & Moses, H.L. pRb is necessary for inhibition of N-myc expression by TGF-beta 1 in embryonic lung organ cultures. *Development* 121, 3057-3066 (1995).
126. Serra, R., Pelton, R.W. & Moses, H.L. TGF beta 1 inhibits branching morphogenesis and N-myc expression in lung bud organ cultures. *Development* 120, 2153-2161 (1994).
127. Zeng, X., Gray, M., Stahlman, M.T. & Whitsett, J.A. TGF-beta1 perturbs vascular development and inhibits epithelial differentiation in fetal lung in vivo. *Dev Dyn* 221, 289-301 (2001).
128. Zhao, J. *et al*. Abrogation of transforming growth factor-beta type II receptor stimulates embryonic mouse lung branching morphogenesis in culture. *Dev Biol* 180, 242-257 (1996).
129. Dickson, M.C. *et al*. Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice. *Development* 121, 1845-1854 (1995).
130. Shull, M.M. *et al*. Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease. *Nature* 359, 693-699 (1992).
131. Beer, H.D. *et al*. Mouse fibroblast growth factor 10: cDNA cloning, protein characterization, and regulation of mRNA expression. *Oncogene* 15, 2211-2218 (1997).
132. Tomlinson, D.C., Grindley, J.C. & Thomson, A.A. Regulation of Fgf10 gene expression in the prostate: identification of transforming growth factor-beta1 and promoter elements. *Endocrinology* 145, 1988-1995 (2004).
133. Kaartinen, V. *et al*. Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction. *Nat Genet* 11, 415-421 (1995).
134. Sanford, L.P. *et al*. TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. *Development* 124, 2659-2670 (1997).
135. Dick, A., Risau, W. & Drexler, H. Expression of Smad1 and Smad2 during embryogenesis suggests a role in organ development. *Dev Dyn* 211, 293-305 (1998).
136. Zhao, J., Lee, M., Smith, S. & Warburton, D. Abrogation of Smad3 and Smad2 or of Smad4 gene expression positively regulates murine embryonic lung branching morphogenesis in culture. *Dev Biol* 194, 182-195 (1998).
137. Sun, J. *et al*. Prenatal lung epithelial cell-specific abrogation of Alk3-bone morphogenetic protein signaling causes neonatal respiratory distress by disrupting distal airway formation. *Am J Pathol* 172, 571-582 (2008).

138. Takahashi, H. & Ikeda, T. Transcripts for two members of the transforming growth factor-beta superfamily BMP-3 and BMP-7 are expressed in developing rat embryos. *Dev Dyn* 207, 439-449 (1996).
139. King, J.A., Marker, P.C., Seung, K.J. & Kingsley, D.M. BMP5 and the molecular, skeletal, and soft-tissue alterations in short ear mice. *Dev Biol* 166, 112-122 (1994).
140. Warburton, D. *et al*. Growth factor signaling in lung morphogenetic centers: automaticity, stereotypy and symmetry. *Respir Res* 4, 5 (2003).
141. Weaver, M., Yingling, J.M., Dunn, N.R., Bellusci, S. & Hogan, B.L. Bmp signaling regulates proximal-distal differentiation of endoderm in mouse lung development. *Development* 126, 4005-4015 (1999).
142. Bellusci, S., Henderson, R., Winnier, G., Oikawa, T. & Hogan, B.L. Evidence from normal expression and targeted misexpression that bone morphogenetic protein (Bmp-4) plays a role in mouse embryonic lung morphogenesis. *Development* 122, 1693-1702 (1996).
143. Winnier, G., Blessing, M., Labosky, P.A. & Hogan, B.L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* 9, 2105-2116 (1995).
144. Shi, W., Zhao, J., Anderson, K.D. & Warburton, D. Gremlin negatively modulates BMP-4 induction of embryonic mouse lung branching morphogenesis. *Am J Physiol Lung Cell Mol Physiol* 280, L1030-1039 (2001).
145. Michos, O. *et al*. Gremlin-mediated BMP antagonism induces the epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. *Development* 131, 3401-3410 (2004).
146. Lu, J., Qian, J., Keppler, D. & Cardoso, W.V. Cathespin H is an Fgf10 target involved in Bmp4 degradation during lung branching morphogenesis. *J Biol Chem* 282, 22176-22184 (2007).
147. Quaggin, S.E. *et al*. The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development* 126, 5771-5783 (1999).
148. Kim, H.J. & Bar-Sagi, D. Modulation of signalling by Sprouty: a developing story. *Nat Rev Mol Cell Biol* 5, 441-450 (2004).
149. de Maximy, A.A. *et al*. Cloning and expression pattern of a mouse homologue of drosophila sprouty in the mouse embryo. *Mech Dev* 81, 213-216 (1999).
150. Minowada, G. *et al*. Vertebrate Sprouty genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. *Development* 126, 4465-4475 (1999).
151. Zhang, S., Lin, Y., Itaranta, P., Yagi, A. & Vainio, S. Expression of Sprouty genes 1, 2 and 4 during mouse organogenesis. *Mech Dev* 109, 367-370 (2001).
152. Mailleux, A.A. *et al*. Evidence that SPROUTY2 functions as an inhibitor of mouse embryonic lung growth and morphogenesis. *Mech Dev* 102, 81-94 (2001).
153. Perl, A.K., Hokuto, I., Impagnatiello, M.A., Christofori, G. & Whitsett, J.A. Temporal effects of Sprouty on lung morphogenesis. *Dev Biol* 258, 154-168 (2003).
154. Mucenski, M.L. *et al*. beta-Catenin is required for specification of proximal/distal cell fate during lung morphogenesis. *J Biol Chem* 278, 40231-40238 (2003).
155. Shu, W. *et al*. Wnt/beta-catenin signaling acts upstream of N-myc, BMP4, and FGF signaling to regulate proximal-distal patterning in the lung. *Dev Biol* 283, 226-239 (2005).
156. Zhang, Y. *et al*. A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. *Nat Genet* 40, 862-870 (2008).
157. Nusse, R. Wnt signaling in disease and in development. *Cell Res* 15, 28-32 (2005).
158. Nelson, W.J. & Nusse, R. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* 303, 1483-1487 (2004).
159. Tebar, M., Destree, O., de Vree, W.J. & Ten Have-Opbroek, A.A. Expression of Tcf/Lef and sFrp and localization of beta-catenin in the developing mouse lung. *Mech Dev* 109, 437-440 (2001).
160. Wang, Z., Shu, W., Lu, M.M. & Morrisey, E.E. Wnt7b activates canonical signaling in epithelial and vascular smooth muscle cells through interactions with Fzd1, Fzd10, and LRP5. *Mol Cell Biol* 25, 5022-5030 (2005).
161. Shu, W., Jiang, Y.Q., Lu, M.M. & Morrisey, E.E. Wnt7b regulates mesenchymal proliferation and vascular development in the lung. *Development* 129, 4831-4842 (2002).
162. Weidenfeld, J., Shu, W., Zhang, L., Millar, S.E. & Morrisey, E.E. The WNT7b promoter is regulated by TTF-1, GATA6, and Foxa2 in lung epithelium. *J Biol Chem* 277, 21061-21070 (2002).
163. Rajagopal, J. *et al*. Wnt7b stimulates embryonic lung growth by coordinately increasing the replication of epithelium and mesenchyme. *Development* 135, 1625-1634 (2008).
164. Okubo, T. & Hogan, B.L. Hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm. *J Biol* 3, 11 (2004).
165. De Langhe, S.P. *et al*. Dickkopf-1 (DKK1) reveals that fibronectin is a major target of Wnt signaling in branching morphogenesis of the mouse embryonic lung. *Dev Biol* 277, 316-331 (2005).

166. Li, C. *et al*. Wnt5a regulates Shh and Fgf10 signaling during lung development. *Dev Biol* 287, 86-97 (2005).
167. Yin, Y. *et al*. An FGF-WNT gene regulatory network controls lung mesenchyme development. *Dev Biol* 319, 426-436 (2008).
168. De Langhe, S.P. *et al*. Formation and differentiation of multiple mesenchymal lineages during lung development is regulated by beta-catenin signaling. *PLoS ONE* 3, e1516 (2008).
169. Morrisey, E.E., Ip, H.S., Lu, M.M. & Parmacek, M.S. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol* 177, 309-322 (1996).
170. Keijzer, R. *et al*. The transcription factor GATA6 is essential for branching morphogenesis and epithelial cell differentiation during fetal pulmonary development. *Development* 128, 503-511 (2001).
171. Morrisey, E.E. *et al*. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* 12, 3579-3590 (1998).
172. Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. & Grosveld, F. The transcription factor GATA6 is essential for early extraembryonic development. *Development* 126, 723-732 (1999).
173. Warburton, D. *et al*. Stem/progenitor cells in lung development, injury repair, and regeneration. *Proc Am Thorac Soc* 5, 703-706 (2008).
174. Kim, C.F. *et al*. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 121, 823-835 (2005).
175. Koutsourakis, M. *et al*. Branching and differentiation defects in pulmonary epithelium with elevated Gata6 expression. *Mech Dev* 105, 105-114 (2001).
176. Shannon, J.M., Nielsen, L.D., Gebb, S.A. & Randell, S.H. Mesenchyme specifies epithelial differentiation in reciprocal recombinants of embryonic lung and trachea. *Dev Dyn* 212, 482-494 (1998).
177. Cardoso, W.V. Molecular regulation of lung development. *Annu Rev Physiol* 63, 471-494 (2001).
178. Shannon, J.M. & Hyatt, B.A. Epithelial-mesenchymal interactions in the developing lung. *Annu Rev Physiol* 66, 625-645 (2004).
179. Tichelaar, J.W., Lim, L., Costa, R.H. & Whitsett, J.A. HNF-3/forkhead homologue-4 influences lung morphogenesis and respiratory epithelial cell differentiation in vivo. *Dev Biol* 213, 405-417 (1999).
180. Tichelaar, J.W., Wert, S.E., Costa, R.H., Kimura, S. & Whitsett, J.A. HNF-3/forkhead homologue-4 (HFH-4) is expressed in ciliated epithelial cells in the developing mouse lung. *J Histochem Cytochem* 47, 823-832 (1999).
181. You, Y. *et al*. Role of f-box factor foxj1 in differentiation of ciliated airway epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 286, L650-657 (2004).
182. Gomperts, B.N., Gong-Cooper, X. & Hackett, B.P. Foxj1 regulates basal body anchoring to the cytoskeleton of ciliated pulmonary epithelial cells. *J Cell Sci* 117, 1329-1337 (2004).
183. Huang, T. *et al*. Foxj1 is required for apical localization of ezrin in airway epithelial cells. *J Cell Sci* 116, 4935-4945 (2003).
184. Chen, J., Knowles, H.J., Hebert, J.L. & Hackett, B.P. Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. *J Clin Invest* 102, 1077-1082 (1998).
185. Zhang, M., Bolfing, M.F., Knowles, H.J., Karnes, H. & Hackett, B.P. Foxj1 regulates asymmetric gene expression during left-right axis patterning in mice. *Biochem Biophys Res Commun* 324, 1413-1420 (2004).
186. Yang, A. *et al*. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol Cell* 2, 305-316 (1998).
187. Levrero, M. *et al*. The p53/p63/p73 family of transcription factors: overlapping and distinct functions. *J Cell Sci* 113 (Pt 10), 1661-1670 (2000).
188. Ghioni, P. *et al*. Complex transcriptional effects of p63 isoforms: identification of novel activation and repression domains. *Mol Cell Biol* 22, 8659-8668 (2002).
189. Daniely, Y. *et al*. Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium. *Am J Physiol Cell Physiol* 287, C171-181 (2004).
190. Koster, M.I., Kim, S., Mills, A.A., DeMayo, F.J. & Roop, D.R. p63 is the molecular switch for initiation of an epithelial stratification program. *Genes Dev* 18, 126-131 (2004).
191. Mills, A.A. *et al*. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* 398, 708-713 (1999).
192. Okubo, T., Knoepfler, P.S., Eisenman, R.N. & Hogan, B.L. Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation. *Development* 132, 1363-1374 (2005).

193. Charron, J. *et al*. Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. *Genes Dev* 6, 2248-2257 (1992).
194. Sawai, S., Shimono, A., Hanaoka, K. & Kondoh, H. Embryonic lethality resulting from disruption of both N-myc alleles in mouse zygotes. *New Biol* 3, 861-869 (1991).
195. Sawai, S. *et al*. Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. *Development* 117, 1445-1455 (1993).
196. Stanton, B.R., Perkins, A.S., Tessarollo, L., Sassoon, D.A. & Parada, L.F. Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes Dev* 6, 2235-2247 (1992).
197. Artavanis-Tsakonas, S., Rand, M.D. & Lake, R.J. Notch signaling: cell fate control and signal integration in development. *Science* 284, 770-776 (1999).
198. Ito, T. *et al*. Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. *Development* 127, 3913-3921 (2000).
199. Post, L.C., Ternet, M. & Hogan, B.L. Notch/Delta expression in the developing mouse lung. *Mech Dev* 98, 95-98 (2000).
200. Borges, M. *et al*. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature* 386, 852-855 (1997).
201. Kageyama, R., Ishibashi, M., Takebayashi, K. & Tomita, K. bHLH transcription factors and mammalian neuronal differentiation. *Int J Biochem Cell Biol* 29, 1389-1399 (1997).
202. Chen, H. *et al*. Conservation of the Drosophila lateral inhibition pathway in human lung cancer: a hairy-related protein (HES-1) directly represses achaete-scute homolog-1 expression. *Proc Natl Acad Sci U S A* 94, 5355-5360 (1997).
203. Linnoila, R.I. *et al*. Constitutive achaete-scute homologue-1 promotes airway dysplasia and lung neuroendocrine tumors in transgenic mice. *Cancer Res* 60, 4005-4009 (2000).
204. Kageyama, R. & Ohtsuka, T. The Notch-Hes pathway in mammalian neural development. *Cell Res* 9, 179-188 (1999).
205. Ito, T. *et al*. Significance of proneural basic helix-loop-helix transcription factors in neuroendocrine differentiation of fetal lung epithelial cells and lung carcinoma cells. *Histol Histopathol* 16, 335-343 (2001).
206. Gontan, C. *et al*. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. *Dev Biol* 317, 296-309 (2008).
207. Park, K.S., Wells, J.M., Zorn, A.M., Wert, S.E. & Whitsett, J.A. Sox17 influences the differentiation of respiratory epithelial cells. *Dev Biol* 294, 192-202 (2006).
208. Ishii, Y., Rex, M., Scotting, P.J. & Yasugi, S. Region-specific expression of chicken Sox2 in the developing gut and lung epithelium: regulation by epithelial-mesenchymal interactions. *Dev Dyn* 213, 464-475 (1998).
209. Perl, A.K., Kist, R., Shan, Z., Scherer, G. & Whitsett, J.A. Normal lung development and function after Sox9 inactivation in the respiratory epithelium. *Genesis* 41, 23-32 (2005).
210. Mansour, S. *et al*. The phenotype of survivors of campomelic dysplasia. *J Med Genet* 39, 597-602 (2002).
211. Bi, W. *et al*. Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc Natl Acad Sci U S A* 98, 6698-6703 (2001).
212. Kist, R., Schrewe, H., Balling, R. & Scherer, G. Conditional inactivation of Sox9: a mouse model for campomelic dysplasia. *Genesis* 32, 121-123 (2002).
213. Mori-Akiyama, Y., Akiyama, H., Rowitch, D.H. & de Crombrugghe, B. Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. *Proc Natl Acad Sci U S A* 100, 9360-9365 (2003).
214. Sock, E. *et al*. Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. *Mol Cell Biol* 24, 6635-6644 (2004).
215. Wurm, A., Sock, E., Fuchshofer, R., Wegner, M. & Tamm, E.R. Anterior segment dysgenesis in the eyes of mice deficient for the high-mobility-group transcription factor Sox11. *Exp Eye Res* 86, 895-907 (2008).
216. Hoser, M. *et al*. Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. *Mol Cell Biol* 28, 4675-4687 (2008).
217. Sinner, D., Rankin, S., Lee, M. & Zorn, A.M. Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development* 131, 3069-3080 (2004).
218. Park, K.S. *et al*. Transdifferentiation of ciliated cells during repair of the respiratory epithelium. *Am J Respir Cell Mol Biol* 34, 151-157 (2006).

Chapter

2

Sox proteins and Sox2 functions in embryonic stem cells

Sox proteins and Sox2 functions in embryonic stem cells

I. Sox proteins

1. Transcription factors

Differences between cell types are established through differential regulation of gene expression, because almost all cells in an organism have identical genetic information. Approximately 5-10% of all genes in the mammalian genome code for transcription factors, which play an important role in this regulation. They act as molecular switches that determine fundamental decisions regarding cell fate, differentiation, or proliferation of distinct cell populations throughout development and adult life [1]. Transcription factors are grouped into transcription factor superfamilies by their type of DNA-binding domain, which is the domain that physically interacts with the DNA at specific sequences in promoters or enhancers, which are located within and around the gene [2]. There are large "families" of transcription factors such as HMG (High Mobility Group) box proteins, zinc finger proteins or homeodomain proteins. Apart from the DNA binding domain, transcription factors usually contain a separate domain that activates or represses gene transcription, depending on the context of the promoter and cellular condition. The transcriptional regulation of eukaryotic gene expression requires both the recruitment of tissue-specific transcription factors to a promoter and the proper establishment of local chromatin structure.

Embryonic development is regulated by several transcription factor families, which can affect the expression of down-stream target genes in developmental pathways [3]. One family of such development regulators is the Sox family. This thesis is focused on the transcription factor Sox2, which is expressed in a tissue specific manner during development.

2. The HMG box superfamily

HMG domain proteins are a large group of unrelated non-histone proteins that are associated with the chromatin and are thought to be derived from a common ancestral gene [4]. Due to their relatively small size, they migrate fast in polyacrylamide gels, and are therefore called "High Mobility Group" proteins. All factors of the HMG box superfamily bear a single or multiple HMG domains. The 79-amino-acid DNA-binding domain, known as the HMG box, forms an L-shape module composed of three helices that bind the minor grove of the DNA, which leads to sharp bend of 80° to 135° in the DNA helix [5].

The HMG domain superfamily can be subdivided into two diverse subgroups based on their sequence specificity of the DNA-binding and the number of HMG domains within a single protein [4] (see **Table 1**):

- **The "classical" HMG proteins group**, which include the non-histone chromatin-associated proteins HMG1 and HMG2 and ribosomal transcription factor UBF [6, 7]. Members of this group generally contain multiple HMG domains, which bind DNA with moderate affinity and little or no sequence specificity and are usually

expressed ubiquitously throughout embryogenesis [4, 6]. These factors play primarily an architectural role in transcriptional regulation i.e. to facilitate interactions between other factors bound at promoters or enhancers [6]. In addition to their ability to bind DNA, HMG domains have also been shown to mediate protein-protein interactions with other transcription factors. Classical HMG domain proteins are the oldest of the HMG domain superfamily and have been found within each of the major phyla [4].

- **The "Sox/TCF/MATA" group**, which includes members of the Sox family [8-10], the T-cell factor (TCF) and the family of lymphocyte enhancer factor 1 (LEF1) [11, 12]. In contrast to the "classical" HMG proteins, protein families in this group contain only a single HMG domain that bind to DNA in a highly sequence-specific manner [4], suggesting that their functions are directed to regulate specific target genes. Transcription factors in this category share about 25% sequence homology in the HMG box, with little or no similarity outside this box. Members of this group show highly restricted expression patterns. In contrast to classical HMG domain proteins, these protein families have only been found in animals [4].

HMG domain proteins can be further classified into families based on the degree of homology within the HMG box [13].

	Number of HMG domains	Specific DNA binding	DNA bending	Expression	Defined role in embryogenesis
"Classical" HMG proteins (e.g. HMG1 and 2, UBF)	Two or more	No	Yes	Abundant ubiquitous all eukaryotes	No
"Sox/TCF/MATA" proteins (e.g. Sox, TCF/LEF)	One	Yes	Yes	Non-abundant cell type-restricted meta-zoan specific	Yes

Tale 1. Comparison of the basic features of "classical" HMG proteins and the "Sox/TCF/MATA" proteins (adapted from [14]).

3. The Sox family

The *Sry* gene (Sex-determining region on the Y chromosome) [9, 10], was the first identified member of the Sox family of transcription factors in 1990. Sox family members are highly conserved across species and they were originally identified by homology, as they contain an HMG box closely related to that of the *Sry* gene [15]. Therefore, *Sry* gave the *Sox* gene family its name; Sry-related HMG box, hence 'Sox', followed by a number corresponding to the order of discovery [9].

Sox proteins have been identified in all animal species (birds, reptiles, amphibians, fish, insects, and nematodes) [16]. Mutations in several of the *SOX* genes have been implicated in the pathogenesis of human congenital syndromes. For example, certain

SRY mutations result in sex reversal and gonadal dysgenesis, while some *SOX9* mutations cause the sex reversal syndrome, campomelic dysplasia and skeletal dysmorphology [17-19]. Mutations in the *SOX2* gene are found in patients with microphthalmia and esophageal atresia [20].

Sox genes are expressed in diverse and dynamic patterns during embryogenesis. Members of the Sox family are expressed in almost every tissue of the embryo during development, and also in a number of adult tissues [13].The expression of a specific Sox transcription factor is not necessarily restricted to a particular cell type or lineage. Their expression pattern during development appears to correlate with early cell fate decisions. For example, *Sry* is expressed in the undifferentiated male gonad and is quickly down regulated once the decision is made to initiate male development [8, 21, 22]. Sox proteins have properties of both classical transcription factors and architectural proteins [23]. They function as classical transcription factors, either activating or repressing specific target genes through interaction with different partner proteins. All Sox factors bind DNA via their HMG domain and recognize the same consensus motif 5'-(A/T)(A/T)CAA(A/T)G-3' [24]. The transcriptional function of Sox proteins dependent on the cell type and the promoter context, and they often have functional redundancy among each other [23]. In contrast to other transcription factors which mainly target the major groove, Sox proteins interact with the minor groove of the DNA helix and, as a consequence, induce a sharp bend in the DNA [23]. The DNA bending capacity of Sox proteins can be functionally important for several reasons. It may bring different regulatory regions of the target gene into close proximity. Thereby, it facilitates the formation of enhanceosomes, i.e., functionally active complexes of transcription factors on different gene enhancer sequences [25]. It also allows the interaction of distant enhancer nucleoprotein complexes with the basal transcription machinery [26, 27]. The local changes in chromatin structure induced by Sox proteins may facilitate the recruitment of higher-order architectural factors (like polycomb or trithorax protein groups) [13]. Bending of DNA by Sox proteins could also act in a negative way by preventing the binding of other factors to adjacent sites in the major groove [13].

3 1. Classification of Sox proteins

Sox proteins were identified through their homology to the HMG box of *Sry* by PCR-based cloning and they were defined by the presence of the HMG domain signature sequence RPMNAFMVW [23, 28]. By consensus, HMG domains of Sox proteins have at least 50% amino acid sequence homology with the HMG domain of Sry. For example, the HMG domains of TCF/LEF1 family are only ≈30% identical to the HMG of Sry [13], and therefore not considered part of the Sox family.

Outside the HMG box, Sox protein sequences are variable, although common domains can be found among a number of Sox proteins, which suggests a recent shared ancestry [28]. The non-HMG regions have been shown to contain transcriptional activation or repression domains and interfaces for protein-protein interactions and thus may perform specialized, subgroup-specific functions [13, 23, 29]. At present, 20 Sox family members have

been identified in mouse and human that can be further divided into eight groups, designed A-H, on the basis of degree of homology, both within and outside the HMG box. Inside an individual subgroup, the amino acid sequence identity of the HMG domain is higher than 90% and outside of the HMG domain they share homologous regions [28]. In **Figure 1A,** all known mouse and human Sox proteins are listed. Sry is the sole member in its group (A) and is most similar to the members of group B. Group B is divided into two subgroups B1 (Sox1, 2 and 3) and B2 (Sox14 and 21) [30]. They all share the so-called "group B homology", which is an amino acid sequence situated in the C-terminal part of the HMG box, but the remaining parts are only conserved within their subgroup, B1 or B2 [29](**Figure 1B**).

In invertebrates, the number of *Sox* genes that has been identified is relatively small, five in *Drosophila* and eight in *Caenorhabditis elegans* and a single gene in these organisms typically corresponds to a group of *Sox* genes in vertebrates [28, 31]. Among vertebrates, individual Sox family members show a high degree of conservation across species, both within and outside their HMG domain. To identify orthologues of a specific *Sox* gene in invertebrates is generally not straightforward but in the case of *Sox2* the similarities are very high. Human SOX2 [32] shares 98% overall similarity with the mouse homologue Sox2 protein [33], 95% with the chicken homologue cSox2 [34] and 88% with the putative *Drosophila* homologue *Dicheate* [35, 36]. Sry is an exception because very little sequence homology can be detected outside the HMG domain, even between closely related species [37].

The 20 orthologous pairs of *Sox* genes in mice and humans show an almost identical genomic organization (with the exception of *Sox6* and *Sox13*) [38]. Further studies showed that the *Sox* genes are distributed evenly throughout the genome in both species with no gene clustering [13]. In addition, all *Sox* genes, with the exception of *Sox5, Sox9* and *Sox17*, are single exon genes, at least with respect to their coding region [16, 37]. These properties indicate a possible expansion of the Sox protein family via non-tandem duplication and retroposition [13].

3 2. Protein interactions

The HMG domain of Sox proteins functions as the major interface for protein-protein interaction and DNA-binding [16]. As previously mentioned, the HMG domain is highly conserved among all Sox proteins, and the variability between Sox proteins is only limited to other parts of the protein. It is even possible to swap the HMG domains between Sox proteins without disrupting the protein-partner interactions and target gene selection [40]. In addition, the binding affinity of Sox proteins to DNA is low. Sox proteins bind DNA with a Kd of ~10^{-7}–10^{-9} M *in vitro*, which is lower than most transcription factors (Kd ~10^{-9}–10^{-11} M *in vitro*). Due to this low affinity, Sox protein binding to its target site without the help of a partner protein is probably too unstable *in vivo* to regulate transcription [29].

The characteristics described above explain why a target site of a Sox-regulated gene can be bound in vitro by both the authentic Sox and by other Sox proteins [29]. However, individual Sox proteins appear to regulate a unique repertoire of target genes *in vivo*.

A

Group	Protein
A	Sry
B1	Sox1 **Sox2** Sox3
B2	Sox14 Sox21
C	Sox4 Sox11 Sox12 (Sox22)
D	Sox5 Sox6 Sox13
E	Sox8 Sox9 Sox10
F	Sox7 Sox17 Sox18
G	Sox15 (Sox20)
H	Sox30

B

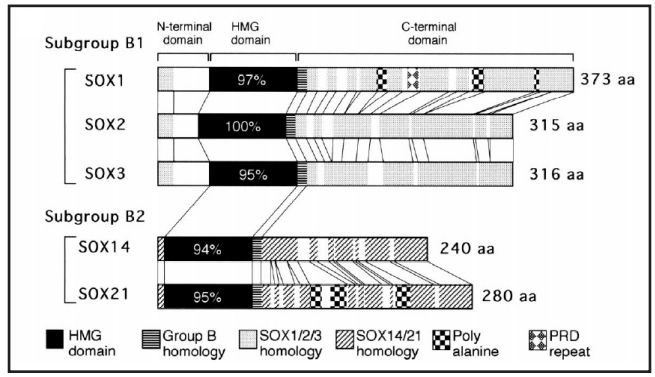


Figure 1. Classification of mouse and human Sox proteins [38]. The human orthologs of the mouse Sox12 and Sox15 proteins are called Sox22 and Sox20, respectively (**A**). Comparison of the domain organization between the Sox proteins of group B. The conserved subdomains are also indicated. Percent identity of the amino acid sequences of the HMG domain sequences relative to that of Sox2, and between the groups members, are indicated. Group B is further divided, where SoxB1 contains a transactivation domain, and SoxB2 contains a repressor domain. Between subgroups B1 and B2, the sequence similarity is found in the 'Group B homology' domain, in addition to the HMG domain. The regions and the domains conserved among Sox1, 2 and 3, and between Sox14 and 21 are indicated. The polyalanine stretch and PRD repeat are also indicated. Protein sizes are shown by the number of amino acids (**B**) (modified from [30, 39]).

Moreover, the same Sox protein may regulate a different set of target genes depending on the cell type and developmental status in which the protein is expressed [29]. These observations suggest that the HMG domain alone cannot be the only determinant for the *in vivo* target site selection. Thus, the question arises how specificity is generated among Sox factors. The answer appears to lie in the observation that Sox proteins form partnerships with specific transcription factors from a variety of protein families to regulate selective target genes [13, 23, 29].

In the current model for the specificity of Sox proteins, the interaction with appropriate partners is required to stabilize the DNA binding of a particular Sox protein to the promoter or enhancer region of the target gene and to exhibit its transcriptional activity *in vivo*. The specificity is generated partly by the combination of Sox proteins

and their potential partners, and partly by the juxtaposition of binding sites for the relevant factors in the target promoter/enhancer [16, 29]. It is also suggested that non-HMG Sox domains may be important to stabilize protein-protein interactions and/or to generate specificity by helping to discriminate between different partners when more than one is present [16].

The target gene specificity of a Sox protein in a particular tissue or cell type will therefore depend on the presence of specific partner proteins. Moreover, the same target gene might be activated or repressed by the same Sox protein depending on the interacting partners present in a particular cell type [16]. This model allows a tight and dynamic transcriptional control of the target genes by changes of the expression of the Sox protein and/or its interacting partner [29]. It also explains how a single Sox protein contributes to the differentiation of different cell types during embryonic development. Three examples of Sox2 partners are: the POU-domain transcription factor Oct4 in maintaining the pluripotent state of embryonic stem (ES) cells [41], Pax6 in the differentiation of lens cells [42] and the POU-domain protein Brn2 in the neural primordium [43].

Most of the cell types express more than one Sox protein at a certain time [33, 34, 44]. Although it is possible that co-expressed Sox proteins perform different functions, there may be functional redundancy, especially when they belong to the same subgroup and share homology outside the HMG domain. This functional redundancy might be a problem in gene inactivation experiments or dominant-negative interference strategies. For example, the lack of a more severe neural phenotype in human patients with a *SOX3* gene deletion may be explained by functional substitution of the other members of the same subgroup (*SOX1* and *2*) that are co-expressed in the central nervous system (CNS) [33, 45].

3 3. The *Sry* gene and sex determination

The first *Sox* gene to be identified was *Sry*, which is the prototype of the Sox protein family. In mammals, *Sry* is the master regulator of male sex determination and is located on the Y chromosome [9, 46]. Transgenic mice carrying the *Sry* gene were phenotypically male even with an XX karyotype [47] and when *Sry* was deleted from the Y chromosome, XY mice adopt a female phenotype [9, 48]. These experiments demonstrated that the *Sry* gene was necessary for male sex determination. Similar observations have been made in human sex-reversed patients with mutations within the *SRY* open reading frame [17, 18]. Most of the missense mutations identified in human patients occur within the HMG domain, predominantly affecting the ability of SRY to bind to and bend DNA [49]. Additionally, other functions of the HMG domain might also be affected, like its nuclear localization [50, 51]

The function of Sry in sex determination must be relatively recent in evolution, given that mammals are the only vertebrates expressing the Sry protein. The Sry HMG domain is the only functional domain that is conserved between different mammalian species, which is uncommon for Sox proteins [52]. In fact, Sry has such a high mutation rate that there is no homology outside the HMG box between mouse and human [37]. This suggests that the HMG domain mediates all the important functions of Sry. The function of the non-HMG part of the Sry protein is unclear. The gonad precursors in

mammalian embryos are initially bipotential, i.e. they are able to differentiate as testis or ovary [53]. Sry in male mice is only expressed for a short period of time (10.5-12 dpc) in the precursors of Sertoli cells [33], which triggers their differentiation and initiates the testis determination pathway of the gonads [21, 22, 54] . Sry may also act to induce male steroidogenesis, which in turn leads to the development of the male external genitalia. In the absence of Sry, the gonad develops, by default, as an ovary. An additional role for Sry might be in sex-specific differentiation of the brain, where it is also expressed in males [55, 56].

Another Sox gene involved in male sex determination in mammals is *Sox9*, a downstream target gene of *Sry*. *Sox9* is expressed in the pre-Sertoli cells shortly after initiation of *Sry* expression [57]. *Sox9* mutations result in XY sex reversal in Campomelic dysplasia patients [58] and its ectopic expression induces testis development [59].

3 4. Nuclear transport of Sox proteins

In order to regulate gene expression, transcription factors must be imported from the cytoplasm, where they are produced, into the nucleus. A family of transport receptors called importins mediates this active process and they belong to the superfamily of Ran-regulated nuclear transport receptors (NTRs) [60]. Importins bind cargoes in the cytoplasm, where RanGTP-levels are low, mediate cargo translocation through the nuclear pore complexes (NPCs) and release their load upon RanGTP-binding in the nucleus [60].

The HMG domain of Sox proteins, aside from its functions in DNA binding/bending and protein-protein interaction, also contains nuclear localization signals (NLS), which are responsible for the nuclear localization of the Sox proteins. Two NLS sequences have been identified at either end of the HMG domain, an N-terminal bipartite NLS and a C-terminal basic cluster NLS [50, 51]. Both NLS motifs are highly conserved, suggesting that all Sox proteins may have a common nuclear import mechanism [16, 51].

In addition, a small hydrophobic leucine-rich nuclear export signal (NES) was identified within the HMG box of the SoxE subgroup, situated between the two NLS sites [61, 62]. NES motifs are recognized by a family of transport receptors, called exportins, which also belong to the superfamily of NTRs but operate in an opposite manner to importins; they bind their cargo together with RanGTP in the nucleus and release it upon GTP-hydrolysis in the cytoplasm [60]. This NES is likely not limited to the SoxE subgroup, as the NES residues are conserved in the corresponding HMG domain regions of other Sox proteins from different subgroups, including Sox2 and Sry [61]. However, it has to be determined whether these sequences are functionally active.

Any protein with both an NLS and an NES can shuttle between the nucleus and the cytoplasm. In the case of transcription factors, controlling the balance of nuclear import and export is another level to regulate their transcriptional activity, as the protein can be sequestrated away from its target genes [63]. Recent studies suggest that active shuttling might be important for the function of some Sox proteins during development [63]. Whether nuclear import or export is indeed a general mechanism for regulation of Sox proteins activity *in vivo* is not clear at the moment.

Sox2 has also been suggested to shuttle between the cytoplasm and nucleus during early embryogenesis [64]. Maternal Sox2 actively enters the nucleus of the 2-cell stage embryo. In the later blastocyst stage Sox2 is nuclear only in the cells of the inner cell mass (ICM). This implies the existence of an active exporting mechanism in trophectoderm cells where Sox2 is exclusively cytoplasmic. The reason behind placing the protein in the cytoplasm, is presumably to prevent it from acting in the nucleus [64]. The nuclear import of Sox proteins will be further discussed in the **chapter 4** of this thesis.

3 5. Sox2 expression and function during embryonic development

The mouse *Sox2* gene has been mapped to chromosome 3 [65]. *Sox2* plays crucial roles during different stages of vertebrate embryonic development and its expression is temporally and spatially regulated.

Sox2 expression starts at the morula-stage of embryo development. In blastocysts it is specifically detected within the cells of the ICM. Expression continues in the epiblast, the tissue that will give rise to the embryo and germ cells [64]. Sox2 is also expressed in ES cells, which are derived from the ICM of blastocysts. During early gastrulation, Sox2 expression in the embryo is restricted to the anterior ectoderm, which gives rise to neuroectoderm and anterior surface ectoderm, while the extraembryonic expression becomes confined to the chorion [66]. At later stages of embryonic development, Sox2 is expressed in the brain, neural tube, sensory placodes, branchial arches, gut endoderm, and the germ cells [64, 67].

Sox2 is expressed at high levels in the murine embryonic nervous system and has been strongly implicated in neurogenesis [33]. Sox2 is the earliest definitive marker for the neural plate, which will form the entire CNS [68]. Sox2 is expressed in specific areas of the brain (telencephalon, mesencephalon and diencephalon). Outside the CNS it is also present in subsets of peripheral nervous system cells and in the primordia of sensory organs [30, 69, 70]. Sox2 is expressed in the nasal placode and epithelium [69]. In the developing eyes, Sox2 is expressed in the lens and the neural retina [71]. Sox2 is expressed in the primordium of the inner ear, the otic placode and in the sensory neuroepithelia [72]. Sox2 is homogeneously expressed in the inner ear sensory epithelia before hair cell differentiation, but it in the mature auditory and vestibular organs it has a unique pattern of expression [73]. Sox2 is also expressed in the branchial arches. When the arches develop, Sox2 continues to be expressed in the primitive foregut endoderm. Later, Sox2 is present in the epithelium of foregut-derived organs, including the tongue, esophagus, trachea, proximal lung and stomach [64, 67, 74, 75]. Sox2 is expressed in the urogenital ridge at the point of gonadal differentiation, similar to Sry. However, Sox2 is observed only in the germ cells, which excludes its participation in sex determination [33]. There are also high levels of Sox2 protein in adult ovary and oocytes, where is localized mainly in the cytoplasm, and in some adjacent stromal cells [64]. Several regulatory elements are responsible for directing Sox2 expression at specific sites in the embryo during development [68]. There are as many as 25 distinct conserved enhancers around the *Sox2* gene [76].

Several lines of evidence indicate that Sox2 has a function where it is expressed.

For example, *Sox2* is required for the maintenance of multipotency of the epiblast and ES cells [41, 64]. In addition, Sox2 is required for the development of the neuroepithelium of the retina and inner ear [71, 77]. Sox2 is also required for the differentiation of endoderm progenitor cells of the embryonic tongue into taste buds [74] and it is involved in the morphogenesis and differentiation of the embryonic lung [67, 78]. The role of Sox2 during lung development will be the subject of **chapter 3** of this thesis.

The lack of *Sox2* expression in mice results in early embryonic lethality [64]. *Sox2* null mutant mouse embryos implant but fail to develop an egg cylinder or epiblast, and they die before gastrulation because Sox2 is required in the ICM of the blastocyst [64]. Other mutations that only affect *Sox2* regulatory elements can cause deafness, defects in the inner ear, circling behavior, and a yellow coat color [65]. The use of two *Sox2* hypomorphic mutants showed a dose-dependent role of Sox2 in the development of the retina and the differentiation of the foregut endoderm [71, 79]. Heterozygous mutations in *SOX2* have been associated in human with severe structural malformations of the eye, bilateral anophthalmia (absent eye) and microphthalmia (small eye)[75], and anophthalmia-esophageal-genital (AEG) syndrome. In AEG infants the esophagus and trachea fail to separate normally and the trachea is connected to the stomach by an abnormal distal esophagus [75, 80, 81]. These symptoms underwrite the developmental functions for Sox2, as found in Sox2 hypomorphic mice described above.

II. Sox2 function in embryonic stem cells

1. Definition of an embryonic stem cell

Mouse embryonic stem cells (ES cells) are derived from the inner cell mass (ICM) of the blastocyst (**Figure 2**) and have two defining properties, self-renewal and pluripotency [82, 83]. Under the right culture conditions they can proliferate without differentiating for prolonged periods of time. In addition, ES cells have the ability to differentiate into any cell type of the mouse, and under certain *in vitro* conditions also the extraembryonic trophectoderm lineage, which will form the placenta [84]. However, ES cells lack the self-organizing capacity to generate a complete organism, as they need to be reintroduced into a blastocyst to generate the different progeny [84].

A number of functional assays are used to check whether ES cells are truly pluripotent. In order of increasing stringency, the first test is the ability of ES cells to differentiate *in vitro*. The next level is the formation of teratomas (where ES cells form germ cell tumors when injected subcutaneously in mice), followed by formation of chimeric mice (the ability to contribute to all tissue types in a chimeric mice, including germ line transmission). The highest level of pluripotency is the ability to generate an entire mouse from ES cells, when these ES cells are injected into a tetraploid blastocyst. Tetraploid cells will only contribute to the extraembryonic linages, so mice derived through this procedure are fully derived from injected ES cells [85].

Mouse and human ES cells share similar stem cell characteristics, but have different *in vitro* requirements in order to maintain an undifferentiated state. Mouse ES cells require

the presence of leukemia inhibitory factor (LIF) and bone morphogenetic protein 4 (BMP4) in the culture medium. In contrast, human ES cell proliferation is independent of LIF and BMP4, and instead requires FGF and activin/nodal signaling [86]. Mouse ES cells are essential tools for the generation of genetically engineered mouse models and the *in vitro* study of different biological process such as differentiation, pluripotency or self-renewal. Human ES cells are of great interest because of their potential in regenerative medicine [87].

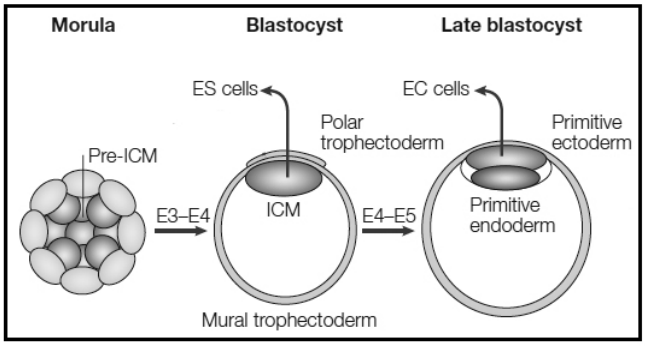


Figure 2. ES cell origin in the mouse embryo. At the blastocyst stage (E3-E4) pluripotent ES cells can be derived *in vitro* from the ICM. At the late blastocyst stage (E4-E5), the ICM develops into the primitive ectoderm (epiblast) and the primitive endoderm. Pluripotent embryonal carcinoma (EC) cells, which are similar to ES cells, can be isolated from the primitive ectoderm (adapted from [88]).

2. Core transcription factor network in ES cell

Over the past years numerous studies have focused on understanding the molecular network that regulates the establishment and maintenance of pluripotency and self-renewal in ES cells. The specific gene-expression program that regulates the ES cell properties is dominated by a few key transcription factors, chromatin regulators, specific signaling pathways and small RNA molecules [85]. It has been shown that ES cell identity is critically dependent on the combined action of three transcription factors Sox2, Oct4 and Nanog, which have a unique expression pattern. Oct4 and Nanog are exclusively expressed in ES cells, the ICM and in germ cells, while Sox2 is expressed in these cells, but also later in development. Sox2, Nanog and Oct4 play an essential role during early development, since deletion of any of these three genes causes early embryonic lethality, due to the loss of pluripotency of the ICM cells [41, 64, 89-91].

ES cells are very sensitive to variations in the expression levels of Sox2, Oct4 and Nanog. This suggests that ES cells require a specific balance of these factors to be able to maintain proliferation while blocking differentiation, by suppressing the expression of lineage-specific factors [87, 92]. Indeed, reduced expression of Oct4 leads to differentiation of ES cells into trophectoderm, whereas overexpression promotes differentiation into primitive endoderm and mesoderm [93]. ES cells depleted for Nanog are prone to

differentiation into extraembryonic endoderm, whereas Nanog overexpression bypasses the requirement for LIF to maintain pluripotency [89, 94]. A recent study has shown that the level of Nanog fluctuates in ES cell colonies and that the level of Nanog correlates with the self-renewal capacity, but that Nanog is not absolutely required for ES cell pluripotency [95].

The effect of Sox2 dosage on ES cell pluripotency has been less extensively studied than in the case of Oct4 and Nanog. Depletion of Sox2 in ES cells results in differentiation mainly into trophectoderm-like cells, similar to the results obtained with Oct4 depletion [41]. In line with this, it was suggested that the main function of Sox2 is to indirectly maintain the expression of Oct4. This was supported by the observation that ES cells can remain undifferentiated in the absence of Sox2 if Oct4 is ectopically expressed from a constitutive promoter [41]. Forced expression of Sox2 did not affect ES cell self-renewal, but upon *in vitro* differentiation, appears to drive ES cells specifically towards the neuroectodermal lineage [96] .

Beside these three master regulators, additional transcription factors that are important for maintaining ES cell properties have been recently identified, such as *Esrrb*, *Tbx3*, *Tcl1, Rif1, Tcf3, Sall4, Foxd3, Stat3*, *Klf4, Dax1, Nac1, Nr5a2* and *Myc* [87, 97, 98]. Depletion or downregulation of the expression of any of these factors of these factors impairs the ability of ES cells to proliferate or cause ES cells differentiation into specific lineages. The organization of the transcriptional network controlling ES cell status and the protein-protein interactions within this network are not fully understood. However, it seems clear that Sox2, Oct4 and Nanog are on top of the transcription factor hierarchy that controls the ES-cell-specific gene expression program [87] .

3. Promoter co-occupancy of pluripotency transcription factors

To understand how ES cells maintain their identity, we need to know more about the molecular mechanisms by which Sox2, Oct4 and Nanog regulate transcription in ES cells. These three master regulators form an auto-regulatory feedback loop by binding together to their own promoters [99-102]. This means that Sox2, Oct4 and Nanog tightly regulate their own expression and the expression of the other two, which maintains the appropriate levels of expression of all three factors (**Figure 3A**).

Recently a number of studies determined the genome-wide locations of several transcription factors by ChIP-chip [99] in human ES cells, and ChIP-PET [103], ChIP-chip [104] and ChIP-seq [105] in mouse ES cells. These different sets of data revealed that Sox2, Oct4 and Nanog co-occupy many of their target genes (**Figure 3B**). Two sets of target genes can be distinguished. The first set is formed by genes active in ES cells, such as Sall4, Dax1 and Rif1 transcription factors with a known function in maintaining ES cell status. The second set contains genes that are not expressed in ES cells, but will be expressed upon ES differentiation, like lineage-specific transcription factors [99, 103]. This suggests that Sox2, Oct4 and Nanog act together to maintain ES cell status, by activating genes associated with ES cell self-renewal and pluripotency and repressing genes associated with lineage specification differentiation. Interestingly, the majority of the silent target

genes co-occupied by Sox2, Oct4 and Nanog are also bound by the Polycomb repressive complexes (PRCs) [106-108]. Polycomb group (PcG) proteins facilitate gene silencing in metazoans by the epigenetic modification of the chromatin structure. PcG proteins play a crucial role during early embryonic development and are involved in maintaining the pluripotent state of ES cells [109]. PcG proteins function in at least two distinct multimeric complexes, PRC1 and PRC2. PRC2 creates the H3K27me3 (H3 lysine 27 tri-methylation) histone modification mark, which provide a recruitment site for PRC1 or maintenance complex. PRC1 in turn, creates the H2AK119ub1 (H2A lysine 119 monoubiquitination) histone modification mark, which contribute to the repressive chromatin state [110]. By contrast, the activating H3K4me3 (H3 lysine 4 tri-methylation) mark is created by the trithorax group (trxG) proteins [110]. Many promoters of key developmental genes in ES cells have 'bivalent chromatin domains', with overlapping activating (H3K4m3) and repressive (H3K27m3) marks [109]. It was suggested that the 'bivalent domains' silence genes in ES cells, while keeping them primed for activation upon differentiation [106].

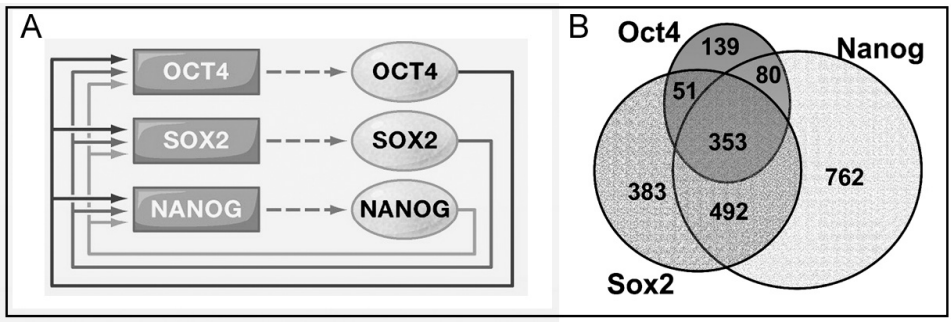


Figure 3. Oct4, Nanog, and Sox2 form an interconnected auto-regulatory feedback loop through binding together to their own promoters (**A**). Venn diagram representing the overlap of the targets of Sox2, Oct4 and Nanog in human ES cells (**B**) (adapted from [85, 99]).

Aside from the pluripotency factors Sox2, Oct4 and Nanog, a number of other transcription factors were analyzed in order to get a more complete picture of the transcription regulatory networks controlling ES cells status. The genomic binding sites for *Klf4, c-Myc, Nanog, Dax1, Rex1, Zpf281, Nac1, Stat3, Smad1, Zfx, n-Myc, Esrrb, Tcfcp2l1, E2f1* and *Ctcf* [104, 105] factors were determined. Interestingly, as in the previous studies, a high degree of co-localization of the different transcription factors on their putative target genes was observed. Two classes of target genes were distinguished, the inactive/repressed genes, which are bound by few transcription factors,and the active genes which are co-occupied by four or more transcription factors. Interestingly, the different transcription factors share almost the same binding site within the promoter, a motif similar to the consensus Sox2-Oct4 target sequences, which suggests that they form a protein complex [104]. This idea is supported by a study that used a proteomic approach to identified the interacting partner proteins of Nanog in ES cells, some of this partner proteins include the transcription factors that co-occupy target genes with Nanog, such as Oct4, Dax1, Zpf281 and Nac1 [87].

4. Induced pluripotent stem cells

Ectopic expression of Sox2, together with Oct4, c-Myc and Klf4 transcription factors is sufficient to reprogram different types of mouse and human somatic cells into induced pluripotent stem (iPS) cells, which resemble ES cells in both their behaviour and their expressed marker genes [111]. The first mouse iPS-like cells were generated using retroviral-mediated introduction of these four factors. These iPS cells formed teratomas but were unable to produce live chimeric mice [111]. Selection for better ES cell markers and a longer reprogramming period did yield real iPS cells that contributed to chimeric mice [112, 113]. Within two weeks after infection, the expression from the retrovirus becomes silenced and the iPS cells reactivated the endogenous expression of Oct4, Sox2 and Nanog to maintain their pluripotent state [85]. Human iPS cells have been generated using the same four transcription factors [114, 115] or a different combination of factors OCT4, SOX2, NANOG, and LIN28 [116].

The molecular mechanism behind the reprogramming process is poorly understood. The global epigenetic state of somatic cells is reset into a more pluripotent epigenetic state, which includes the reactivation of the silenced X chromosome in female cells [117]. Since iPS cells were first described, several groups have been working on improving the efficiency of creating iPS cells with the aim of future therapeutic applications in humans [113, 118]. For example, using a modified protocol it was possible to generate mouse and human iPs cells without the use of the c-Myc retrovirus [119]. If the silenced c-Myc virus is reactivated in the iPS cells or cells derived from them, it will cause tumors, as c-Myc is a potent oncogene, and c-Myc is therefore a safety concern. Although the efficiency of obtaining of iPS cells without c-Myc was lower, the cells were of high quality and were able to give rise to adult chimeras when transplanted into blastocysts [119]. Another way to overcome this problem was the use of adenovirus which does not integrate into the genome of their target host, decreasing the danger of creating tumors [120]. The use of only two factors, Oct4 and Sox2, combined with the addition of HDAC inhibitors and DNA methylase inhibitors increased the efficiency of iPS cell derivation and can compensate for the loss in efficiency due to the absence of c-Myc and Klf4 oncogenes [121]. Recently a protocol was published in which iPS cells were generated using transient transfection with plasmids without the use of any viruses, albeit with very low efficiency. Not using viruses for reprogramming will increase the safety of the reprogramming procedure for future clinical use[118]. In mouse, the first applications of iPS cells are being tested. iPS cells with a corrected globin gene defect could correct a human sickle cell anemia model in mouse [122]. Another way to use iPS cells for medical research is to generate iPS cells from patients with a genetic disease [123]. These iPS cells and their differentiated derivatives can now be used to study the disease progress and for drug development.

References

1. Latchman, D.S. Transcription factors: an overview. Int J Biochem Cell Biol 29, 1305-1312 (1997).
2. Mitchell, P.J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245, 371-378 (1989).
3. Prior, H.M. & Walter, M.A. SOX genes: architects of development. Mol Med 2, 405-412 (1996).
4. Laudet, V., Stehelin, D. & Clevers, H. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res 21, 2493-2501 (1993).
5. Ferrari, S. et al. SRY, like HMG1, recognizes sharp angles in DNA. Embo J 11, 4497-4506 (1992).
6. Grosschedl, R., Giese, K. & Pagel, J. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet 10, 94-100 (1994).
7. Jantzen, H.M., Admon, A., Bell, S.P. & Tjian, R. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature 344, 830-836 (1990).
8. Goodfellow, P.N. & Lovell-Badge, R. SRY and sex determination in mammals. Annu Rev Genet 27, 71-92 (1993).
9. Gubbay, J. et al. 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 (1990).
10. Sinclair, A.H. et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240-244 (1990).
11. Travis, A., Amsterdam, A., Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. Genes Dev 5, 880-894 (1991).
12. Waterman, M.L. & Jones, K.A. 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 (1990).
13. Pevny, L.H. & Lovell-Badge, R. Sox genes find their feet. Curr Opin Genet Dev 7, 338-344 (1997).
14. Dailey, L. & Basilico, C. Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J Cell Physiol 186, 315-328 (2001).
15. Pevny, L. & Placzek, M. SOX genes and neural progenitor identity. Curr Opin Neurobiol 15, 7-13 (2005).
16. Wilson, M. & Koopman, P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Genet Dev 12, 441-446 (2002).
17. Berta, P. et al. Genetic evidence equating SRY and the testis-determining factor. Nature 348, 448-450 (1990).
18. Foster, J.W. & Graves, J.A. An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. Proc Natl Acad Sci U S A 91, 1927-1931 (1994).
19. Jager, R.J., Anvret, M., Hall, K. & Scherer, G. A human XY female with a frame shift mutation in the candidate testis-determining gene SRY. Nature 348, 452-454 (1990).
20. Zenteno, J.C., Perez-Cano, H.J. & Aguinaga, M. Anophthalmia-esophageal atresia syndrome caused by an SOX2 gene deletion in monozygotic twin brothers with markedly discordant phenotypes. Am J Med Genet A 140, 1899-1903 (2006).
21. Hacker, A., Capel, B., Goodfellow, P. & Lovell-Badge, R. Expression of Sry, the mouse sex determining gene. Development 121, 1603-1614 (1995).
22. Koopman, P., Munsterberg, A., Capel, B., Vivian, N. & Lovell-Badge, R. Expression of a candidate sex-determining gene during mouse testis differentiation. Nature 348, 450-452 (1990).
23. Wegner, M. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. Mar 15;27(6), 1409-1420 (1999).
24. Harley, V.R., Lovell-Badge, R. & Goodfellow, P.N. Definition of a consensus DNA binding site for SRY. Nucleic Acids Res 22, 1500-1501 (1994).
25. Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y. & Pallavi, B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol 39, 2195-2214 (2007).
26. Werner, M.H. & Burley, S.K. Architectural transcription factors: proteins that remodel DNA. Cell 88, 733-736 (1997).
27. Wolffe, A.P. Architectural transcription factors. Science 264, 1100-1101 (1994).
28. Bowles, J., Schepers, G. & Koopman, P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol 227, 239-255 (2000).
29. Kamachi, Y., Uchikawa, M. & Kondoh, H. Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet 16, 182-187 (2000).

30. Uchikawa Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech. Dev. 84, 103-120. (1999).
31. Cremazy, F., Berta, P. & Girard, F. Genome-wide analysis of Sox genes in Drosophila melanogaster. Mech Dev 109, 371-375 (2001).
32. Stevanovic, M., Zuffardi, O., Collignon, J., Lovell-Badge, R. & Goodfellow, P. The cDNA sequence and chromosomal location of the human SOX2 gene. Mamm Genome 5, 640-642 (1994).
33. Collignon, J. et al. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. Development 122, 509-520 (1996).
34. Uwanogho, D. et al. Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. Mech Dev 49, 23-36 (1995).
35. Nambu, P.A. & Nambu, J.R. The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. Development 122, 3467-3475 (1996).
36. Russell, S.R., Sanchez-Soriano, N., Wright, C.R. & Ashburner, M. The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. Development 122, 3669-3676 (1996).
37. Whitfield, L.S., Lovell-Badge, R. & Goodfellow, P.N. Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature 364, 713-715 (1993).
38. Schepers, G.E., Teasdale, R.D. & Koopman, P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell 3, 167-170 (2002).
39. Kamachi, Y., Cheah, K.S. & Kondoh, H. Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. Mol Cell Biol 19, 107-120 (1999).
40. Bergstrom, D.E., Young, M., Albrecht, K.H. & Eicher, E.M. Related function of mouse SOX3, SOX9, and SRY HMG domains assayed by male sex determination. Genesis 28, 111-124 (2000).
41. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol 9, 625-635 (2007).
42. Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. Genes Dev 15, 1272-1286 (2001).
43. Tanaka, S. et al. Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. Mol Cell Biol 24, 8834-8846 (2004).
44. Stock, D.W., Buchanan, A.V., Zhao, Z. & Weiss, K.M. Numerous members of the Sox family of HMG box-containing genes are expressed in developing mouse teeth. Genomics 37, 234-237 (1996).
45. Stevanovic, M., Lovell-Badge, R., Collignon, J. & Goodfellow, P.N. SOX3 is an X-linked gene related to SRY. Hum Mol Genet 2, 2013-2018 (1993).
46. Polanco, J.C. & Koopman, P. Sry and the hesitant beginnings of male development. Dev Biol 302, 13-24 (2007).
47. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. Nature 351, 117-121 (1991).
48. Lovell-Badge, R. & Robertson, E. XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. Development 109, 635-646 (1990).
49. Koopman, P. Sry, Sox9 and mammalian sex determination. Exs, 25-56 (2001).
50. Poulat, F. et al. Nuclear localization of the testis determining gene product SRY. J Cell Biol 128, 737-748 (1995).
51. Sudbeck, P. & Scherer, G. Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. J Biol Chem 272, 27848-27852 (1997).
52. Pontiggia, A. et al. Sex-reversing mutations affect the architecture of SRY-DNA complexes. Embo J 13, 6115-6124 (1994).
53. Koopman, P. Sex determination: a tale of two Sox genes. Trends Genet 21, 367-370 (2005).
54. Jeske, Y.W., Bowles, J., Greenfield, A. & Koopman, P. Expression of a linear Sry transcript in the mouse genital ridge. Nat Genet 10, 480-482 (1995).
55. Lahr, G. et al. Transcription of the Y chromosomal gene, Sry, in adult mouse brain. Brain Res Mol Brain Res 33, 179-182 (1995).
56. Mayer, A., Lahr, G., Swaab, D.F., Pilgrim, C. & Reisert, I. The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. Neurogenetics 1, 281-288 (1998).
57. Sekido, R., Bar, I., Narvaez, V., Penny, G. & Lovell-Badge, R. SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. Dev Biol 274, 271-279 (2004).
58. Houston, C.S. et al. The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux et al in 1971. Am J Med Genet 15, 3-28 (1983).

59. Vidal, V.P., Chaboissier, M.C., de Rooij, D.G. & Schedl, A. Sox9 induces testis development in XX transgenic mice. Nat Genet 28, 216-217 (2001).
60. Gorlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol 15, 607-660 (1999).
61. Gasca, S. et al. Anuclear export signal within the high mobility group domain regulates the nucleocytoplasmic translocation of SOX9 during sexual determination. Proc Natl Acad Sci U S A 99, 11199-11204 (2002).
62. Rehberg, S. et al. Sox10 is an active nucleocytoplasmic shuttle protein, and shuttling is crucial for Sox10-mediated transactivation. Mol Cell Biol 22, 5826-5834 (2002).
63. Smith JM, K.P. The ins and outs of transcriptional control: nucleocytoplasmic shuttling in development and disease. Trends Genet. Jan;20(1), 4-8. (2004).
64. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17, 126-140 (2003).
65. Dong, S. et al. Circling, deafness, and yellow coat displayed by yellow submarine (ysb) and light coat and circling (lcc) mice with mutations on chromosome 3. Genomics 79, 777-784 (2002).
66. Wood, H.B. & Episkopou, V. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech Dev 86, 197-201 (1999).
67. Ishii, Y., Rex, M., Scotting, P.J. & Yasugi, S. Region-specific expression of chicken Sox2 in the developing gut and lung epithelium: regulation by epithelial-mesenchymal interactions. Dev Dyn 213, 464-475 (1998).
68. Papanayotou, C. et al. A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. PLoS Biol 6, e2 (2008).
69. Donner, A.L., Episkopou, V. & Maas, R.L. Sox2 and Pou2f1 interact to control lens and olfactory placode development. Dev Biol 303, 784-799 (2007).
70. Schlosser, G. & Ahrens, K. Molecular anatomy of placode development in Xenopus laevis. Dev Biol 271, 439-466 (2004).
71. Taranova, O.V. et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev 20, 1187-1202 (2006).
72. Knowlton, V.Y. Effects of extraembryonic membrane deficiency on differentiation of the embryonic avian brain and sense organs. Acta Anat (Basel) 66, 420-445 (1967).
73. Hume, C.R., Bratt, D.L. & Oesterle, E.C. Expression of LHX3 and SOX2 during mouse inner ear development. Gene Expr Patterns 7, 798-807 (2007).
74. Okubo, T., Pevny, L.H. & Hogan, B.L. Sox2 is required for development of taste bud sensory cells. Genes Dev 20, 2654-2659 (2006).
75. Williamson, K.A. et al. Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. Hum Mol Genet 15, 1413-1422 (2006).
76. Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. & Kondoh, H. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev Cell 4, 509-519 (2003).
77. Kiernan, A.E. et al. Sox2 is required for sensory organ development in the mammalian inner ear. Nature 434, 1031-1035 (2005).
78. Gontan, C. et al. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. Dev Biol 317, 296-309 (2008).
79. Que, J. et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 134, 2521-2531 (2007).
80. Fantes, J. et al. Mutations in SOX2 cause anophthalmia. Nat Genet 33, 461-463 (2003).
81. Hagstrom, S.A. et al. SOX2 mutation causes anophthalmia, hearing loss, and brain anomalies. Am J Med Genet A 138A, 95-98 (2005).
82. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156 (1981).
83. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78, 7634-7638 (1981).
84. Niwa, H. How is pluripotency determined and maintained? Development 134, 635-646 (2007).
85. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567-582 (2008).
86. Rossant, J. Stem cells and early lineage development. Cell 132, 527-531 (2008).
87. Wang, J. & Orkin, S.H. A protein roadmap to pluripotency and faithful reprogramming. Cells Tissues Organs 188, 23-30 (2008).
88. Boiani, M. & Scholer, H.R. Regulatory networks in embryo-derived pluripotent stem cells. Nat Rev Mol Cell Biol 6, 872-884 (2005).

89. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643-655 (2003).
90. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631-642 (2003).
91. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379-391 (1998).
92. Silva, J. & Smith, A. Capturing pluripotency. Cell 132, 532-536 (2008).
93. Niwa, H., Miyazaki, J. & Smith, A.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372-376 (2000).
94. Hatano, S.Y. et al. Pluripotential competence of cells associated with Nanog activity. Mech Dev 122, 67-79 (2005).
95. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-1234 (2007).
96. Zhao, S., Nichols, J., Smith, A.G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol Cell Neurosci 27, 332-342 (2004).
97. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533-538 (2006).
98. Zhou, Q., Chipperfield, H., Melton, D.A. & Wong, W.H. A gene regulatory network in mouse embryonic stem cells. Proc Natl Acad Sci U S A 104, 16438-16443 (2007).
99. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-956 (2005).
100. Kuroda, T. et al. Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol Cell Biol 25, 2475-2485 (2005).
101. Okumura-Nakanishi, S., Saito, M., Niwa, H. & Ishikawa, F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem 280, 5307-5317 (2005).
102. Tomioka, M. et al. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res 30, 3202-3213 (2002).
103. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38, 431-440 (2006).
104. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132, 1049-1061 (2008).
105. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106-1117 (2008).
106. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315-326 (2006).
107. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349-353 (2006).
108. Lee, T.I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125, 301-313 (2006).
109. Ku, M. et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 4, e1000242 (2008).
110. Pietersen, A.M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. Curr Opin Cell Biol 20, 201-207 (2008).
111. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-676 (2006).
112. Lewitzky, M. & Yamanaka, S. Reprogramming somatic cells towards pluripotency by defined factors. Curr Opin Biotechnol 18, 467-473 (2007).
113. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448, 318-324 (2007).
114. Lowry, W.E. et al. Generation of human induced pluripotent stem cells from dermal fibroblasts. Proc Natl Acad Sci U S A 105, 2883-2888 (2008).
115. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861-872 (2007).
116. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917-1920 (2007).
117. Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1, 55-70 (2007).

118. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. Science 322, 949-953 (2008).
119. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 26, 101-106 (2008).
120. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. Science 322, 945-949 (2008).
121. Huangfu, D. et al. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol 26, 1269-1275 (2008).
122. Hanna, J. et al. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science 318, 1920-1923 (2007).
123. Park, I.H. et al. Reprogramming of human somatic cells to pluripotency with defined factors. Nature 451, 141-146 (2008).

Chapter

3

Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation

Developmental Biology, 1; 317(1): 296-309, 2008

Developmental Biology 317 (2008) 296–309

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Sox2 is important for two crucial processes in lung development: Branching morphogenesis and epithelial cell differentiation

Cristina Gontan [a], Anne de Munck [a], Marcel Vermeij [b], Frank Grosveld [c], Dick Tibboel [a], Robbert Rottier [a,*]

[a] Department of Pediatric Surgery, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands
[b] Department of Pathology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands
[c] Department of Cell Biology and Genetics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received for publication 26 September 2007
Revised 11 February 2008
Accepted 14 February 2008
Available online 29 February 2008

Keywords:
Lung
Development
Sox2
Neuroendocrine cells
p63
Branching morphogenesis

ABSTRACT

The primary lung bud originates from the foregut and develops into the bronchial tree by repetitive branching and outgrowing of the airway. The Sry related HMG box protein Sox2 is expressed in a cyclic manner during initiation and branching morphogenesis of the lung. It is highly expressed in non-branching regions and absent from branching regions, suggesting that downregulation of Sox2 is mandatory for airway epithelium to respond to branch inducing signals. Therefore, we developed transgenic mice that express a doxycycline inducible Sox2 in the airway epithelium. Continuous expression of Sox2 hampers the branching process resulting in a severe reduction of the number of airways. In addition, the bronchioli transiently go over into enlarged, alveolar-like airspaces, a pathology described as bronchiolization of alveoli. Furthermore, a substantial increase was observed of cGRP positive neuroendocrine cells and ΔNp63 isoform expressing (pre-) basal cells, which are both committed precursor-like cells. Thus, Sox2 prevents airways from branching and prematurely drives cells into committed progenitors, apparently rendering these committed progenitors unresponsive to branch inducing signals. However, Sox2 overexpression does not lead to a complete abrogation of the epithelial differentiation program.

© 2008 Elsevier Inc. All rights reserved.

Introduction

The mechanisms involved in patterning, development and differentiation of the foregut and lung are only recently beginning to be understood at the molecular level. It has become clear that the intimate interactions between the mesenchyme and endodermal epithelium are mandatory for formation and completion of lung development. Gene inactivation and transgenic expression studies in mice have shown the importance of individual factors at different phases of lung development (Cardoso and Lu, 2006; Maeda et al., 2007). The primary bronchial tree develops by a process of repetitive branching of the airways, called branching morphogenesis. The secreted growth factor Fgf10 induces formation of the first morphological sign of the lung, the lung bud, by activating the Fgfr2b receptor (Bellusci et al., 1997b; Park et al., 1998; Sutherland et al., 1996; Weaver et al., 2000). Mice lacking either Fgf10 or the Fgfr2b do not form lungs, indicating the requirement of this growth factor signaling pathway for the lung initiation (De Moerlooze et al., 2000; Min et al., 1998; Sekine et al., 1999). Moreover, local Fgf10 expression in the mesenchyme induces the activation of the Fgfr2b receptor on cells of the surrounding airway epithelium, triggering the activation of downstream genes like *Spry2*, *Shh* and *Bmp4* (Litingtung et al., 1998; Mailleux et al., 2001; Weaver et al., 1999). These signaling molecules subsequently invoke a feedback mechanism that controls correct budding, growth and differentiation of the epithelium and mesenchyme. Continuous growth, budding and elongation of the airway epithelium are subject to interactions with the surrounding mesenchyme and involves various signaling cascades (reviewed in Cardoso and Lu, 2006; Maeda et al., 2007).

A crucial event in lung development after formation of the lung bud is the repetitive branching of the airways to develop into the primary bronchial tree. Based on the expression pattern, a putative role in the formation of the lung, trachea and esophagus for Sox2 was suggested at this stage of lung development (Ishii et al., 1998). Sox2 is part of a highly conserved family of transcription factors with the sex-determining gene *Sry* as the founding member. Although they share similar DNA binding properties, individual Sox proteins appear to regulate specific sets of target genes in vivo by restricted patterns of expression and in combination with specific interacting partners. Sox proteins are recognized as key players in the regulation of embryonic development and determination of cell fate and maintenance (Kamachi et al., 2000; Lefebvre et al., 2007; Schepers et al., 2002; Wilson and Koopman, 2002). Sox2 is cell-autonomously required in both embryonic and extra-embryonic cell types, and the lack of Sox2 results in peri-implantation lethality (Avilion et al., 2003). Sox2, together with Oct3/4 and Nanog, plays a pivotal role in the maintenance of ES cell pluripotency by regulating lineage commitment factors (Boyer et al., 2005; Ivanova et al., 2006; Rao and Orkin, 2006; Wang et al., 2006). Later in development, Sox2 is involved in spe-

* Corresponding author. Erasmus MC, Dept Pediatric Surgery/Cell Biology, PO Box 2040, 3000 CA Rotterdam, The Netherlands. Fax: +31 10 4089468.
E-mail address: r.rottier@erasmusmc.nl (R. Rottier).

cification and maintenance of neural stem cells during neurogenesis (Bylund et al., 2003; Graham et al., 2003; Pevny and Placzek, 2005). Sox2 also regulates neural progenitor cell differentiation in the retina and the formation of sensory hair cells and supportive cells in the cochlea (Kiernan et al., 2005; Taranova et al., 2006).

Sox2 is expressed in the endoderm of the foregut from the pharyngeal arches down to the future stomach, and later in organs derived form the foregut. Using hypomorphic mouse mutants with reduced *Sox2* expression, Que et al. (2007) demonstrated that Sox2 is involved in the morphogenesis of the trachea and esophagus, as well as the differentiation of the esophageal epithelium. The timing and localization of Sox2 expression in the foregut appears tightly regulated, as it is switched off at the site where the lung buds. Subsequently, after the outgrowth of the bud, Sox2 expression reappears, only to be shut down again the moment that this bud starts to branch. Again, Sox2 expression is reinstalled after these branches are growing out and the leading tip of the growing branch is moving away from the branching site (Ishii et al., 1998). Therefore, we hypothesized that downregulation of Sox2 at these branching sites is crucial for correct development of the lung.

We investigated the role of Sox2 during early lung development and differentiation of the airway epithelium using conditional overexpression of Sox2 in the epithelial cells of the lung. We show that Sox2 activity affects two processes during lung development. First, it is implicated in branching morphogenesis of the bronchial tree, since overexpression leads to severe disruption of the formation of the airway branches. Sox2 downregulation at branching sites is required for cells to respond to branch-inducing signals and to maintain these cells undifferentiated. Second, we show that Sox2 is required for differentiation of the epithelium of the airways, since overexpression leads to a premature commitment of cells towards differentiation. We show that aberrant Sox2 expression leads to an increase in the number of neuroendocrine cells, (pre-) basal cells and cells with a gastric/intestinal expression pattern.

Materials and methods

Generation of transgenic animals

A 3.6 kb *Hin*dIII–*Sal*I murine genomic fragment containing the Sox2 coding region was isolated and a myc epitope encoding sequence was cloned right after the endogenous ATG. This fragment was subsequently cloned into the pTRE-Tight vector (pTT::myc-sox2; Clontech). An *Aat*II–*Dra*I fragment was used for pronuclear injection to produce transgenic lines. Tail tip DNA of transgenic lines was initially genotyped by Southern blot analysis, after which positive lines were routinely checked by PCR, using a transgene-specific primer combination (sense 5′-GCAGAAGCTGATCTCGGAGG-3′; antisense 5′-CTGGTCATGGAGTTGTACTGCAGG-3′). Mice were kept under standard conditions and all experiments were performed according to the guidelines of the local ethics committee. Lung-specific overexpression of the myc-Sox2 transgene was obtained by crossing the myc-Sox2 lines with SPC-rtTA transgenic mice (gift of Jeffrey Whitsett, Cincinnati). Administration of doxycycline to dams from gestational age 6.5 onward in the drinking water (2 mg/ml doxycycline, 5% sucrose) resulted in lung epithelium-specific expression. Each experiment was executed on at least three litters containing double transgenic, single transgenic and wild type pups. Three independent transgenic lines have been tested with identical results.

Histology

Lungs were dissected and fixed in 4% PFA overnight at 4 °C before processing for paraffin embedding according to routine protocols (Canis Parera et al., 2005; van Tuyl et al., 2005). Antigen retrieval was performed with microwave treatment in 10 mM citric acid buffer pH 6.0. Sections were blocked with 5% BSA in PBS for 10 min and incubated with primary antibody diluted in 5% BSA in PBS overnight at 4 °C. The following antibodies were used: Sox2 (Chemicon), β-tubulin IV (Biogenex), proSP-C (Chemicon), cGRP (Sigma), CC10 (rabbit polyclonal, kindly provided by B. Stripp), T1α (University of Iowa Hybridoma bank, mAb 8.1.1; 1:100), myc (Roche), α-SM actin (NeoMarkers) Ttf1 (Neomarkers), Foxa2 (Seven Hills Bioreagents), p63 (Santa Cruz), phospho-SMAD1/5/8 (Cell Signaling Technology), phospho-ERK1/2 (Cell Signaling Technology), cleaved caspase-3 (Cell Signaling Technology), phospho-Histone H3 (Upstate Biotechnology), Krt14 (Monosan) and Fli-1 (Santa Cruz). The Sox2 immunohistochemistry was carried out using the Envision+ System (DAKO), StreptABC complex/HRP (DAKO) and HRP-DAB colorimetric detection. Secondary antibodies against the correct IgG species were conjugated with peroxidase (Dako), Alexa594 (molecular Probes) and FITC (Nordic Labs).

Whole-mount β-galactosidase staining to visualize the pulmonary vasculature was performed as previously described (Canis Parera et al., 2005). Whole-mount in situ hybridization was essentially performed as described (Ishii et al., 1998). The following mouse cDNAs were used as templates for the synthesis of digoxigenin-labeled riboprobes: a 1077 bp *Bmp4* cDNA and a 634 bp *Shh* cDNA, according to the protocol supplied by Roche Diagnostics.

Microarray analysis

Lungs of 6 littermates (pTT::myc-SOX2$^{+/-}$ * SPC-rtTA$^{+/-}$) were dissected at E18.5 and the middle and caudal lobes were used for total RNA isolation with Trizol reagent according to the manufacturer's instructions (Invitrogen life technologies, Carlsbad, CA, USA). RNA was purified using the RNeasy MinElute Cleanup kit (Qiagen, Valencia, CA, USA) and cDNA was synthesized from 3 μg RNA using the GeneChip Expression 3′-Amplification Reagents One-Cycle cDNA Synthesis kit (Affymetrix, Santa Clara, CA, USA). Biotin-labeled cRNA synthesis, purification and fragmentation were performed according to standard conditions. Fragmented biotinylated cRNA was subsequently hybridized onto Affymetrix Mouse Genome 430 2.0 microarray chips. After normalization, the data were analyzed with OmniViz software, version 3.6.0 (Omniviz, Inc., Maynard, MA, USA).

RT-PCR

Total RNA was isolated with Trizol reagent and 2 μg was used to prepare cDNA (Invitrogen). Primer sets for the PCR were myc-Sox2: 5′-GCAGAAGCTGATCTCGGAGG-3′ and 5′-CTGGTCATGGAGTTGTACTGCAGG-3′; Sox21: 5′-GCAGTTGGTGTTTGCTTTGC-3′ and 5′-ACTGACAACTTCCACCTCACA-3′; HPRT: 5′-GCTGGTGAAAAGGACCTCTC-3′ and 5′-CACAGGACTAGAACACCTGC-3′; all other primer sets were previously described (Li et al., 2005; Okubo and Hogan, 2004; Okubo et al., 2005; Perl et al., 2005; Que et al., 2007; Serls et al., 2005; Wan et al., 2005).

Results

Endogenous expression of sox2 in the lung

We first studied Sox2 expression during the different phases of murine embryonic lung development. Initially, it is expressed in the epithelial cells of the rostral part of the primitive gut up to the future stomach (Ishii et al., 1998). It is locally downregulated at the ventral side of the foregut at the side where the lung bud appears (E9.5; Fig. 1A). As soon as the lung bud is established, Sox2 protein is expressed again in the epithelial cells (Fig. 1B). The expression of Ttf1 is reciprocal from Sox2 expression, as has also been shown previously (Fig. 1A, insert; Que et al., 2007). When the right and left main bronchi are forming, Sox2 expression vanishes again, and reappears after these primary branches have developed (Figs. 1B, C). This specific expression pattern of Sox2, being upregulated and subsequently downregulated in the epithelial cells at sites where the bronchial tree branches, is repeated several times during the embryonal (Figs. 1A, B), pseudoglandular (Figs. 1C–E), canalicular (Fig. 1F), saccular (Fig. 1G) and alveolar phases of lung development (Fig. 1H). Sox2 remains highly expressed in the proximal airways and in the postnatal 3 weeks old lung it is primarily expressed in the epithelium of the conducting airways (Fig. 1H). Sox2 expression is excluded from the (primitive) alveoli, as can be seen from the strict border between the bronchioli and alveoli, the bronchioalveolar duct junction (inserts, Figs. 1F–H).

Conditional overexpression of Sox2 in embryonic lungs

Downregulation of *Sox2* expression appears a prerequisite for airways to branch and hence continued expression of *Sox2* should inhibit normal branching. We therefore generated an inducible myc-tagged *Sox2* transgene under the control of a tet-on promoter. Expression in embryonic lung epithelial cells was established by crossing the Sox2 line with the SPC-rtTA line (Perl et al., 2002). The SPC promoter drives expression of the rtTA gene in epithelial cells of the embryonal lung, and the rtTA protein can activate the tet-on promoter under the influence of doxycycline. Lungs isolated from single transgenic or non-induced double transgenic fetuses appeared healthy and showed no macroscopic or microscopic abnormalities. However, double transgenic pups that were given doxycycline succumbed within half

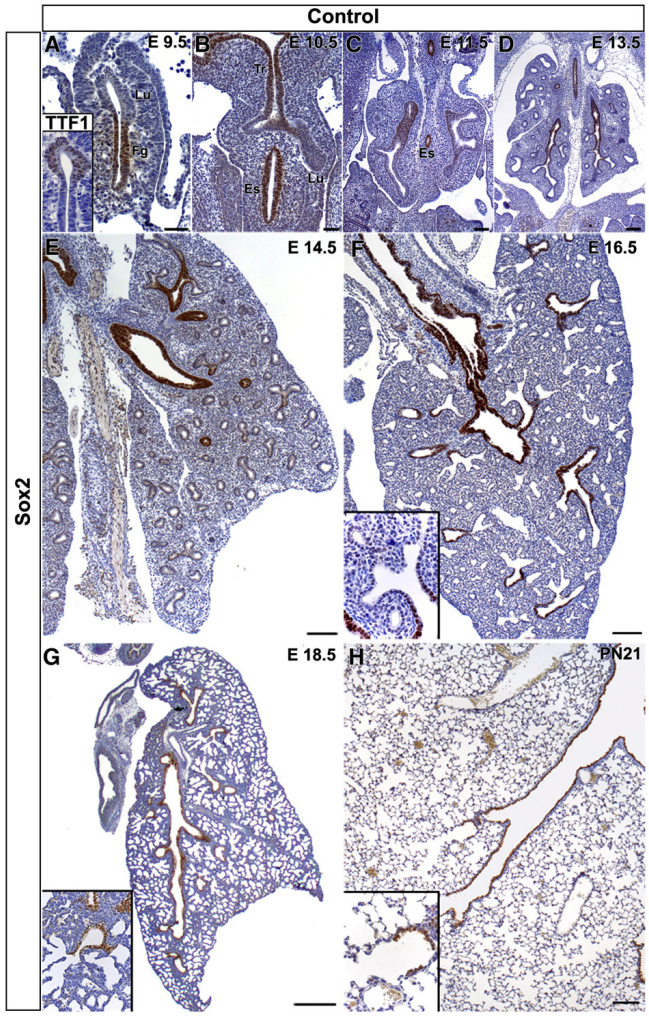


Fig. 1. Sox2 expression during wild-type lung development. (A) At E9.5, Sox2 was detected in epithelial cells of the foregut, reciprocal to Ttf1 expression in the lung bud (insert). Sox2 is exclusively expressed in the epithelial cells of the non-branching developing airways at E10.5 (B), E11.5 (C), E13.5 (D) and E14.5 (E). Sox2 remained expressed in epithelial cells of the conducting airways at E 16.5 (F), E18.5 (G) and PN21 (H) with a sharp boundary at the bronchoalveolar duct (see inserts). Es, esophagus; Lu, lung; Tr, trachea; Fg, foregut. Sections are frontal and scale bars are 50 μm (B), 100 μm (A, C–F) and 500 μm (E).

an hour after birth. When born by caesarian section at 18.5d of gestation, they were gasping for air and unable to oxygenize their blood.

Examination at E 15.5 and 18.5 showed that their lungs were smaller, contained enlarged airspaces and a marked decrease in the number of airways (Figs. 2A–G), especially in the periphery of the lobes. The abnormalities caused by Sox2 overexpression were even more pronounced in homozygous Sox2 transgenic lungs (compare Fig.1F with Fig.1G), showing a devastating, global lung defect. These enlarged airspaces were in contact with the trachea as pups born by caesarian section filled their lungs with air (not shown). However, the gas exchange surface is either too small, or oxygenation of blood is impossible due to structural changes. The presence of the four right lobes and one left lobe indicated that initial branching is unaffected, but the strong reduction of the number of primitive alveolar structures suggested that later branching is impaired.

The airspaces had very thin, primitive alveolar walls, composed of just a few cell layers, and a general airway epithelial dysplasia (Figs.

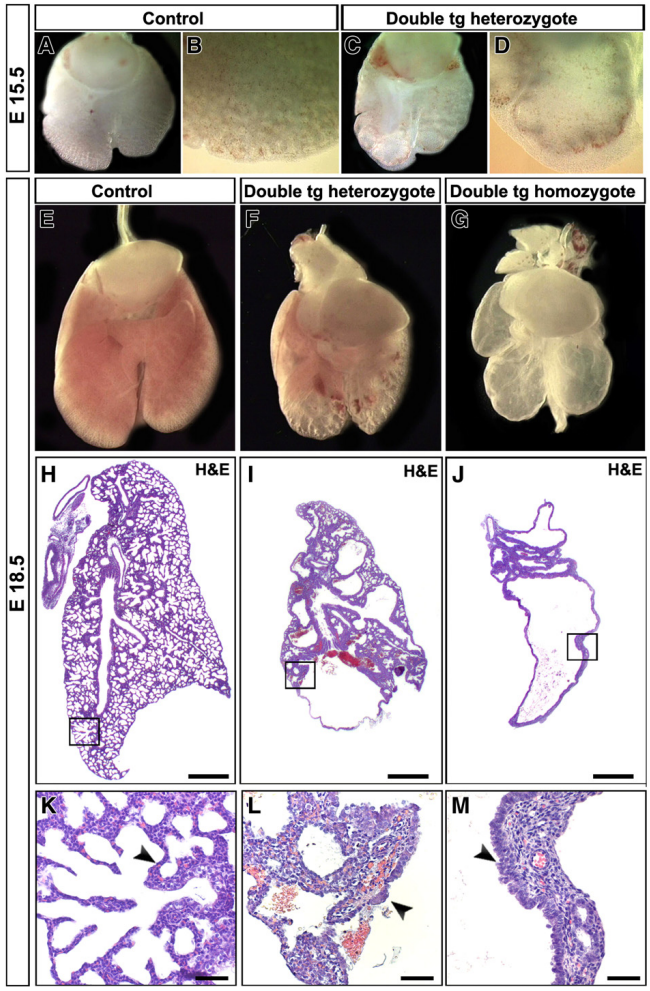


Fig. 2. Sox2 overexpression leads to abnormal primitive alveolar formation. External appearances of control (A, B, E), heterozygous (C, D, F) and homozygous Sox2 (G) overexpressing lungs displaying enlarged airspaces and a marked decrease in the number of airways. Low and high power images of sections stained with hematoxylin and eosin from normal (H, K), heterozygous (I, L) and homozygous transgenic lungs (J, M). Note the undifferentiated dysplastic epithelium in the double transgenic (K–M, arrowhead). The sharp boundary between bronchioles and primitive alveoli in wild-type (K, arrowhead) is absent in the transgenic lungs (L, M). Scale bars: 500 μm (H–J), 50 μm (K–M).

2H–M). The major bronchi were present, but subsequent branching in smaller bronchioli and primitive alveoli was severely hampered (Figs. 2H–J). The columnar epithelium of the bronchioli extends into the proximal part of the dilated air spaces and gradually changes into a dysplastic epithelium. The cells lining the enlarged airspaces are frequently cuboidal, but some regions contained multilayered atypical cells (arrows in Figs. 2L, M). Using the myc epitope present in the transgenic protein, we revealed that the transgene was already expressed at an early stage of lung development, E10.5 (Fig. 3D). Total Sox2 expression in the double transgenic lungs (Figs. 3A–C, G, H) consists of transgenic Sox2 in both proximal and distal airways (Figs. 3D–F, I, J), as well as the endogenous Sox2 in the proximal airways only (Fig. 1). However, the level varies between cells and the pattern is certainly not uniform in all cells of the airways (Figs. 3D, E). The lumen of the enlarged air spaces often contained cellular debris, that could originate either from shredded skin epithelial cells that are inhaled by fetal breathing movements, or by shredding of airway epithelial cells (inserts, Figs. 3H, J).

Next, we tested whether the initial effects on epithelial branching and differentiation in the Sox2 overexpressing lungs could be reversed when doxycycline was administered for different time periods. Indeed, when doxycycline treatment was started at the

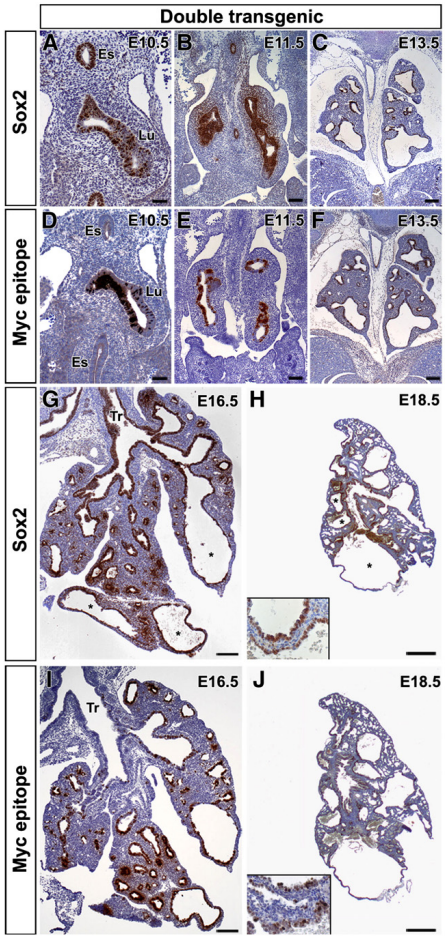


Fig. 3. Transgenic Sox2 extends expression pattern to the distal lung. Comparison of total (Sox2; A–C, G, H) and transgenic Sox2 (Myc epitope; D–F, I, J) expression from E10.5 to E18.5 in Sox2 overexpressing lungs. Sox2 staining was increased throughout much of the respiratory epithelium at E16.5 and E18.5 (G–J). Sox2 impaired branching morphogenesis and caused formation of dilated airspaces (asterisks in panels G, H). Note the difference in expression level and pattern of transgenic Sox2 protein in the airways. Es, esophagus; Lu, lung; Tr, trachea; Fg, foregut. Sections are frontal and scale bars are 50 μm (A, D), 100 μm (B, C, E, F, G, I) and 500 μm (H, J).

pseudoglandular phase, the lungs of double transgenic animals showed only small airsacs at the periphery of the lung (Figs. 4C, G). This indicates that the initial bronchial tree has developed normally, but is only disturbed by Sox2 overexpression later in development, resulting in the formation of airsacs at the periphery. Sections of these lungs revealed the absence of big airsacs, but the formation of primitive alveoli is hampered throughout the lung (Fig. 4G). When mice were fed doxycycline from the onset of lung development until the pseudoglandular phase, it appears that the lungs of the double transgenic mice recover partially from the Sox2 overexpression, as shown by the formation of small primitive, alveolar structures at the periphery (arrow in Figs. 4D, H).

Since the affected lungs clearly showed impaired branching at E12.5, we investigated genes that are expressed at branching sites. Whole-mount in situ hybridizations of E12.5 lungs showed no difference in expression between control and affected lungs of the branching inducing gene, *Fgf10* (Bellusci et al., 1997b; Weaver et al., 2000), which was expressed in the mesenchyme around the lung buds, and its receptor, *Fgfr2*, which was expressed in the airway epithelium (data not shown). The Fgf10 responsive genes, *Shh* and *Bmp4* (Bellusci et al., 1997a; Weaver et al., 1999), were expressed in the epithelium of the outgrowing buds (Figs. 5A–D and E–H, respectively). However, the expression of *Bmp4* seems reduced, although the area of expression is extended around the expanded airspaces (compare arrowheads in Figs. 5G, H). Quantitative PCR analysis of Bmp4 did not show differences in total level of expression between non-affected and affected lungs, which supports the finding of the in situ hybridization (Figs. 5E–H; data not shown). Since *Bmp4* is implicated in a negative feedback loop to counteract the branch outgrowth, this result could indicate a possible role for Bmp4 in the occurrence of the dilated airspaces. Therefore, we analyzed downstream mediators of Bmp and Fgf signaling with antibodies against the phosphorylated forms of SMAD1/5/8 and ERK1/2 (Phospho-p44/42 MAP Kinase), respectively (Figs. 5I–N). Control lungs showed clear, nuclear pERK staining in the tips of the growing airways, corresponding to active Fgf signaling (Fgf10), but no nuclear pSMAD staining was observed in the tips (Fig. 5K, arrowheads in M). However, in the Sox2 overexpressing lungs, no clear nuclear staining of pERK in the tips of the airways was observed, instead more nuclear pSMAD positive cells were present (Figs. 5L, N). Equal numbers of pSMAD positive cells were detected throughout the mesenchyme in both control and affected lungs (Figs. 5K, L).

Since the lungs appear to have a comparable phenotype to that seen when *Bmpr1a* or *Nmyc* is conditionally deleted, we investigated the proliferation and apoptosis in the Sox2 overexpressing lungs using antibodies directed against the phospho Histone H3, which detects cells in the mitotic phase of the cell cycle, and against the cleaved form of caspase3, which detects cells that are undergoing apoptosis (Eblaghie et al., 2006; Okubo et al., 2005). Very few apoptotic cells could be detected in lung tissue of controls and Sox overexpressors, even if some positive cells were present in the lumen of the affected lungs (arrowhead, Figs. 6C, D, H, I). Although not statistically significant, the Sox2 overexpressing lungs seem to have a slight reduction in proliferation, which could account for the reduced size of the affected lungs (Figs. 6E, F, J, K; quantification, Figs. 6G, H; lung size, see Figs. 2E–G).

Epithelial differentiation of Sox2 overexpressing lungs

Double transgenic lungs were subsequently analyzed with several lung-specific markers to identify whether the epithelium still contained differentiated cells (Fig. 7). The general lung epithelial marker Ttf1 appeared to be expressed in nearly all epithelial cells of the double transgenic lungs (Figs. 7A, B). Thus, Sox2 and Ttf1 are not mutually exclusive in airway epithelium, contrasting the reciprocal expression pattern in trachea and esophagus (Que et al., 2007). Respiratory epithelial cells (Foxa2), type II pneumocytes (Sftpc), ciliated cells (β-tubulin) and type I pneumocytes (T1α) are present in the double transgenic lungs, showing that differentiation of cells along the proximo-distal axis of the airway epithelium was still possible (Figs. 7C–J). The expression of the proximal Clara cell marker CC10 (Scgb1a1) extended into the enlarged airsacs, indicating that the border between the proximal airways and the distal airways is not as distinct as in control lungs, but seemed to be more gradual (Figs. 8A, B, arrowheads). Several other epithelial markers, *SftpA*, *SftpB*, *Aqp5* and *Scgb3a1*, were also analyzed by RT-PCR, confirming our finding that epithelial cells are still able to initiate differentiation into several

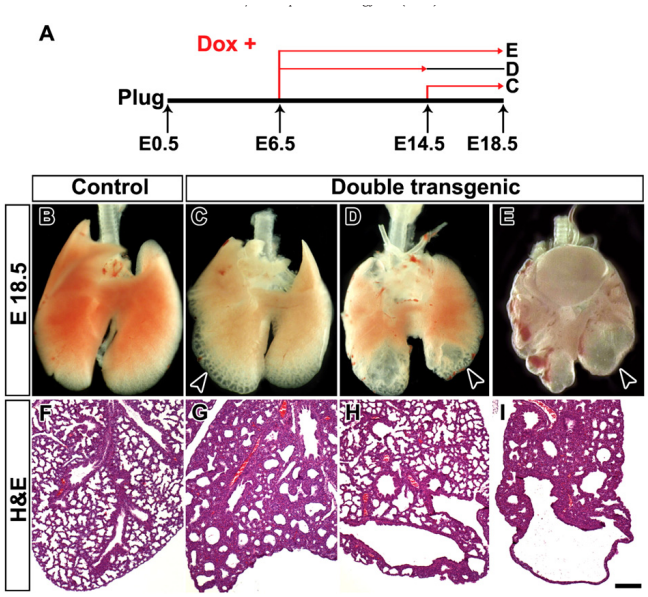


Fig. 4. Effect of different time frames of Sox2 induction. Sox2 was induced at different time points by manipulating the doxycycline administration, as indicated in panel A. External appearance and hematoxylin and eosin staining are shown for controls (B, F), Sox2 induced from the pseudoglandular phase onward (C, G), Sox2 induced until the pseudoglandular phase (D, H) or Sox2 induction throughout lung development (E, I). Note the clear differences in the degree of branching defects (arrowheads, C, D, E). All lungs were harvested by caesarian section at 18.5 days of gestation. Sections are frontal and scale bar is 200 μm.

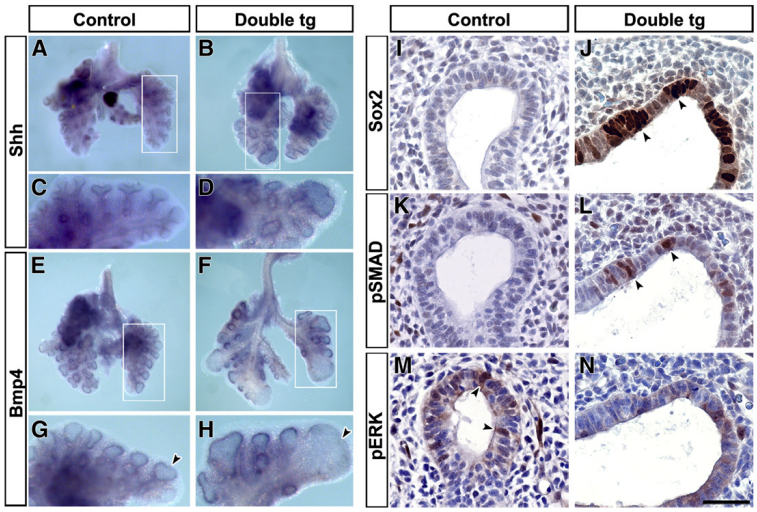


Fig. 5. Bmp4 expression is reduced around affected airways. Whole-mount *in situ* hybridization for *Shh* and *Bmp4* in embryonic E12.5 lungs of control and double transgenic lungs. Epithelial *Shh* expression in double transgenic (B, D) lungs is comparable to that of control lungs (A, C). Epithelial *Bmp4* expression is reduced around dilated airspaces in Sox2 overexpressing lungs (F, H, arrowhead) compared to control (E, G, arrowhead). Sequential sections of E12.5 control (I, K, M) and affected (J, L, N) lungs were stained with Sox2 (I, J), phospho-SMAD1/5/8 (K, L) and phospho-ERK1/2 (M, N). Note the clear difference in nuclear pERK1/2 and pSMAD1/5/8 staining between controls and affected lungs (K versus L, M versus N, arrowheads). Scale bar: 50 μm.

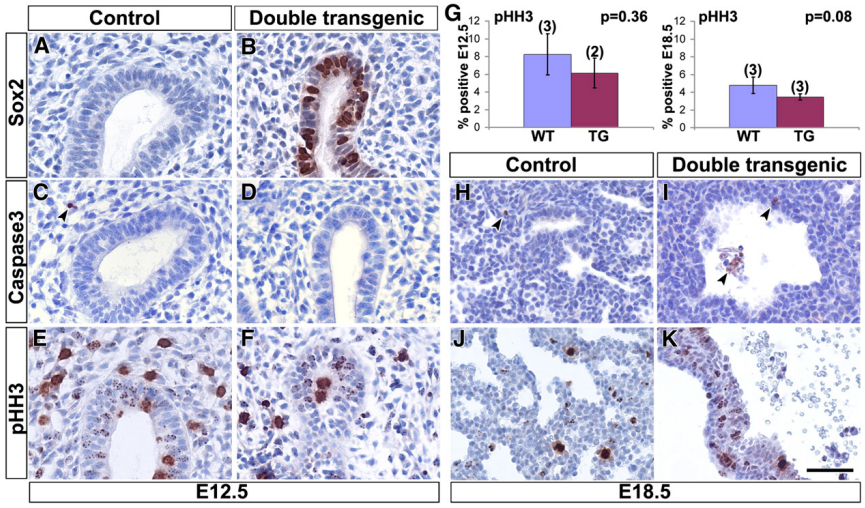


Fig. 6. Apoptosis and proliferation in Sox2 overexpressing lungs. E12.5 and E18 lungs were analyzed for apoptosis and proliferation using antibodies against the cleaved caspase3 (C, D, H, I) and phospho-Histone H3 (E, F, J, K), respectively. Sequential sections of E12.5 were stained with Sox2 (A, B). Proliferative cells in the epithelium of E12.5 and total E18.5 lungs were quantified by counting 3 microscopic fields. The number of lungs counted is indicated in parentheses and the results are means ± SEM.

lineages along the proximo-distal axis of the fetal lung (Figs. 10A). We also did not observe changes in expression between control and Sox2 overexpressing lungs of three genes involved in host defense, *Acid chitinase*, *LysozymeM* and *LysozymeP* (not shown).

The development and integrity of the pulmonary vasculature were analyzed by crossing the Tie2-LacZ transgenic mouse reporter line (Schlaeger et al., 1997) with the pTT-myc-sox2 and SPC-rtTA lines. Whole-mount lacZ staining showed that vessels still encompass the dilated airways and are connected to the major pulmonary arteries (Figs. 8C, D). However, these vessels seem to sprout less frequently compared to non-transgenic littermates, but capillaries containing erythrocytes could even be detected in the thinnest part of the airsac mesenchyme (Figs. 8E, F, arrows). The integrity of the vessels appeared to be intact as no edema or leakage of blood cells was observed. Muscularization of the airways and vasculature extended more distally than in control lungs (Figs. 8G, H). The enlarged airsacs have significantly more α-SMA positive staining than the primitive alveolar structures of control lungs, indicating that the proximal part of the airways has expanded more distally, supporting the finding of CC10 positive cells in the upper part of the enlarged airsacs. It could also be that these dilated airspaces are proximal airways that failed to form distal airways.

Three Sox2 overexpressing lungs and three control lungs were used for microarray hybridization (data not shown). One of the strongest upregulated genes in the double transgenic lungs is *cGRP*, a marker for neuroendocrine cells. Double transgenic lungs had a significant increased number of neuroepithelial cells and neuroepithelial bodies as shown by the presence of cGRP protein (Figs. 9A, B, compare to transgene co-expression in Fig. 9C). Other typical neuroendocrine cell markers, like chromogranin A and secretogranin V, were also significantly upregulated in the double transgenic samples (not shown). Interestingly, neuroendocrine cells are the first committed cells emerging during lung development (Warburton et al., 1998). The microarrays also showed that the proneural bHLH gene *Mash1* was significantly upregulated. Mash1 drives epithelial cells into neuroendocrine cell fate and Mash1 knockout mice lack neuroendocrine cells (Borges et al., 1997; Ito et al., 2000). Sox2 was shown to be involved in the upregulation of Mash1, which in turn induces *Sox21* (Bylund et al., 2003; Kiernan et al., 2005; Sandberg et al., 2005). Indeed, *Sox21* is specifically upregulated as shown by RT-PCR (Fig. 10A). Cells expressing the cGRP marker have lost the transgene expression, as shown by the absence of double positive cells. This indicates that the differentiation of epithelial cells into the neuroendocrine lineage looses expression of the SPC transgenic promoter. In addition, the Notch ligand Dll3, a target of Mash1, is also upregulated (Casarosa et al., 1999). Notch signaling is implicated in a process called lateral inhibition, in which cells committed to become neuroendocrine cells prevent neighboring cells from turning into the same cell fate. The Notch receptor is expressed in the non-neuroendocrine cells, and the neuroendocrine cells subsequently produce Notch ligands in a Mash1 dependent manner. In this respect, it is interesting that we also detected a strong increase in the expression of *Dner*, which is described as a neuron-specific Notch ligand (Eiraku et al., 2005). We did not find an altered expression of Sox4 and Sox9 (Fig. 10A) and based on the microarrays, further analysis of other Sox genes, like Sox17, was not required.

Based on the microarray analysis, we found in the Sox2 overexpressing lungs a significant increase of p63, a homologue of p53, which is involved in development and maintenance of stratified epithelia. p63 knockout mice die soon after birth without skin and epithelial appendages (Mills et al., 1999; Yang et al., 1999). In the development of the esophagus, trachea and bronchial epithelium p63 controls the switch from early stem cells to basal cells and their maintenance (Daniely et al., 2004). In control lung samples, p63 was only expressed in the basal cell layer of the tracheal epithelium (insert, Fig. 9D), whereas in double transgenic lungs, p63 expression was extended into the epithelia of the bronchioli and enlarged airspaces (Figs. 9D, E) supporting the findings that high Sox2 levels coincide with high p63 levels (Que et al., 2007). Staining sequential sections showed indeed that the transgenic Sox2 was co-expressed with p63 (Figs. 9E, F). The *p63* gene has two different promoters leading to two major isoforms, one with the N-terminal transactivation domain (TAp63) and one without (ΔNp63). It appeared that in the double transgenic lungs

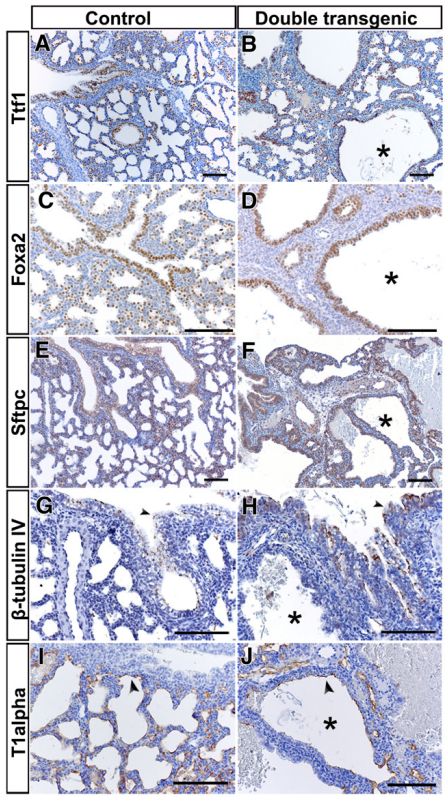


Fig. 7. Expression of epithelial differentiation markers in Sox2 overexpressing lungs. Expression of Ttf1 (A, B), Foxa2 (C, D), Sftpc (E, F), β-tubulin IV (G, H) and T1α (I, J) in normal and Sox2 overexpressing lungs at E18.5. Sites and extent of expression were comparable between controls (A, C, E, G, I) and Sox2 overexpressing lungs (B, D, F, H, J). Asterisks indicate the dilated airspaces, occasionally filled with debris. Scale bars: 100 μm.

ΔNp63 expression is markedly increased (Fig. 10A), whereas the TAp63 isoform is not expressed (not shown). The ΔNp63 isoform is believed to play a role in the maintenance of the proliferative potential of the basal layer in the epithelium, fitting the idea that Sox2 drives cells into a committed progenitor stage (Candi et al., 2007; Daniely et al., 2004).

$p63^+/Krt14^-$ cells are pre-basal cells, constituting an intermediate, precursor cell for the $p63^+/Krt14^+$ basal cells. In trachea, the normal side of p63 expression (see insert, Fig. 9D), Krt14 staining appeared around the nuclei of the p63 positive cells (Fig. 10F), indicative for basal cells, whereas in the distal cells the staining was absent (Fig. 10D). In the double transgenic lungs, Krt14 was not expressed in the p63 positive cells (Fig. 10E), suggesting that pre-basal cells ($p63^+$, $krt14^-$) are present in the dilated airspaces with exogenous p63 expression, which is in line with our observations that Sox2 induces the appearance of committed progenitor cells. We also found upregulation of genes normally expressed in secretory cells of the posterior glandular stomach, *Tff3*, and intestine, *Muc5b* (Fig. 10A). In addition, *Agr2* (posterior stomach and intestine) and *Reg3g* (intestine) showed a drastic increase by microarray analysis, but are only slightly elevated in the lungs of Sox2 overexpressing mice by RT-PCR. PAS staining revealed that more glycogen containing cells were present in the epithelium of the Sox2 overexpressing lungs, confirming the immature phenotype (Figs. 10B, C). These markers, in combination with high p63 and high Sox2, show that the affected lung epithelium has obtained epithelial cells with a gastric phenotype, even though there is no clear gastric epithelial morphology. We also found high levels of *Dcpp*, which recently was shown to be expressed in the submucosal glands originating from the tracheal epithelium (Okubo and Hogan, 2004).

Discussion

Sox2 is a transcription factor that plays a key role in cell fate decisions in a variety of cell types, like ES cell maintenance, neural progenitor specification, cochlear prosensory and neuronal specification, endodermal progenitor differentiation in the taste bud, and lung epithelium differentiation (Avilion et al., 2003; Bylund et al., 2003; Graham et al., 2003; Kiernan et al., 2005; Okubo et al., 2006; Pevny and Placzek, 2005; Que et al., 2007; Taranova et al., 2006). Recent studies with hypomorphic Sox2 mouse mutants with esophageal atresia and tracheoesophageal fistulas showed that the level of Sox2 expression is important in the development of the esophagus and trachea (Que et al., 2007). Loss and gain of function experiments additionally showed that Sox2 is crucial for the development of taste bud sensory cells (Okubo et al., 2006), while Sox2 levels are also important during the development of the sensory organ in the mammalian inner ear (Kiernan et al., 2005). Here we show that overexpression of *Sox2* in the airway epithelium leads to reduced branching of the airways. Although the initial branches are formed, subsequent branching is severely hampered, leading to dilated airspaces. We also show that the differentiation of the epithelial cells is affected, indicating that the expression level of Sox2 is critical in the differentiation of the pulmonary epithelium.

Sox2 and pulmonary branching morphogenesis

Sox2 expression is strictly regulated at sites where airways start to branch. Forced overexpression at these sites affects branching capacity of the airways, leading to a severely reduced number of airways (Fig. 2). However, the primary bronchial tree develops normally, so initial branching is not affected by the Sox2 transgene. At this early developmental stage, the level or distribution of transgenic Sox2 is not sufficient to prevent branching and epithelial cells are already entering the branching process at the time the transgene is switched on. Hence, cells may not have been triggered yet by exogenous Sox2 to differentiate into a branch-unresponsive cell, but instead remain committed to branching. At later stages, more cells express Sox2 and are influenced by Sox2 effects, leading to the observed branching defects. Mice homozygous for the Sox2 transgene (Fig. 2G), with higher Sox2 levels, have a more severe phenotype indicating that the level of expression is indeed important in blocking the branching process. The presence of morphologically normal airways in the heterozygote Sox2 overexpressors (Fig. 2I) is most likely due to the incomplete penetrance of the SPC-rtTA transgene.

How Sox2 is able to prevent cells from responding to branch-inducing signals is still elusive and under investigation. Expression analysis of the branch-inducing growth factor *Fgf10* and its receptor, *Fgfr2* showed no differences between control and Sox2 overexpressing lungs (data not shown), in accordance with Fgf signaling being upstream of Sox2 (Que et al., 2007). However, the expression level of the branch inhibitory gene *Bmp4* was reduced, although the region of expression expanded around the dilated airsacs (Fig. 5). Since Bmp4 inhibits branching in the lung tips, it could be that the reduced expression contributes to the occurrence of dilated airsacs through the lack of inhibition of branch outgrowth. Therefore, signaling by Fgf and Bmp was analyzed using antibodies against the phosphorylated forms

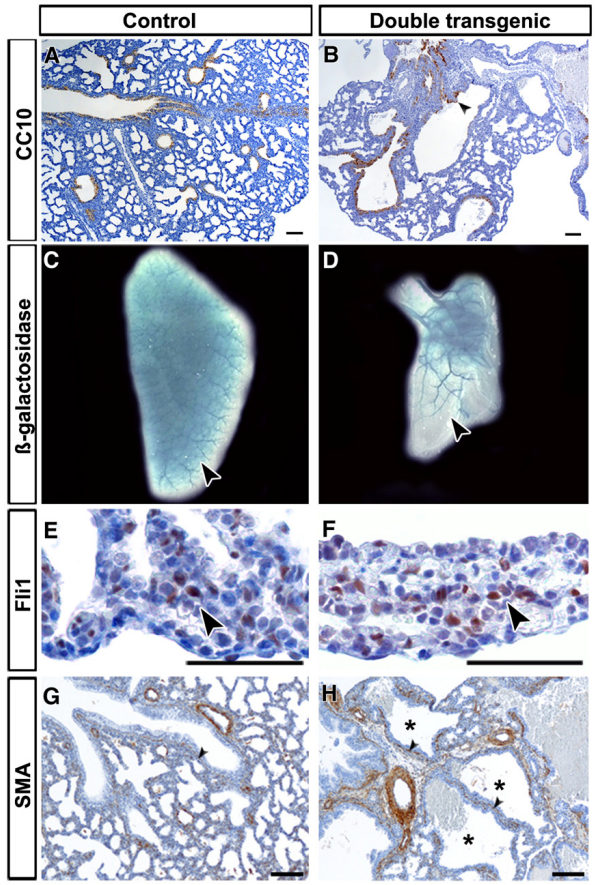


Fig. 8. Clara cells, vessels and muscularization in Sox2 overexpressing lungs: (A, B) Staining for CC10, a Clara cell marker, was expanded into the enlarged airsacs (arrowhead, B). LacZ staining shows *Tie2*-driven β-galactosidase expression in control (C) and Sox2 overexpressing (D) lungs. Note reduced sprouting of vessels in double transgenic lungs (arrows). Fli-1 staining shows the presence of small capillaries surrounding the distal dilated airways, indicating the integrity of the vasculature (E, F). Larger vessels and the main airways are muscularized in controls (G), Sox2 overexpressing lungs showed an increase in the muscularization of the main airways, the vasculature and the enlarged airsacs (H, arrowheads). Scale bars: 100 μm.

of the downstream mediators ERK1/2 and Smad1/5/8, respectively. In contrast to the controls, nuclear pERK1/2 was virtually absent from the very tips, indicating that Fgf signaling in the Sox2 overexpressing lungs was hampered (Figs. 5M, N). Although Bmp4 is expressed in the epithelium of control lungs, there is no detectable pSMAD1/5/8 in the very tip of the lung buds (Figs. 5E, G, K). This is in contrast with the Sox2 overexpressing lungs where the level of Bmp4 mRNA is reduced in the epithelium of the affected lung buds (Figs. 5F, H), but phosphorylated SMAD1/5/8 proteins are expressed in a subset of epithelial cells of the tips (Fig. 5L). However, paracrine Bmp4 signaling is not hampered, because we do detect equal numbers of pSMAD1/5/8/ positive cells in the mesenchyme of both control and Sox2 overexpressing lungs. This suggests that, in spite of the presence of Bmp4 mRNA, autocrine signaling in the epithelial cells of the branching tips is subject to other signaling cues or cofactors, present in the Sox2 overexpressing cells, but absent in the tip cells in control lungs. The exact nature of this signaling process and the role of Sox2 in this mechanism are under investigation. Although it is presently not clear how the expression of Bmp4 mRNA correlates with the presence of the phosphorylated forms of SMAD1/5/8, it is evident that in the branching tips of the Sox2 overexpressing lungs the expression of pSMAD1/5/8 is enhanced as compared to the controls. This fits the hypothesis that Sox2 overexpressing epithelial cells are unresponsive to branch-inducing signals and that Sox2 prematurely forces cells in the tips of the affected lungs into a more differentiated stage. Preliminary data from fetal lung explants showed that overexpression of Sox2 also inhibits branching in vitro. Culturing control and affected embryonal lungs in the presence of Fgf10 soaked beads revealed that the affected lungs do respond to the Fgf10, but the response is initiated later compared to control lungs. The affected epithelium does respond

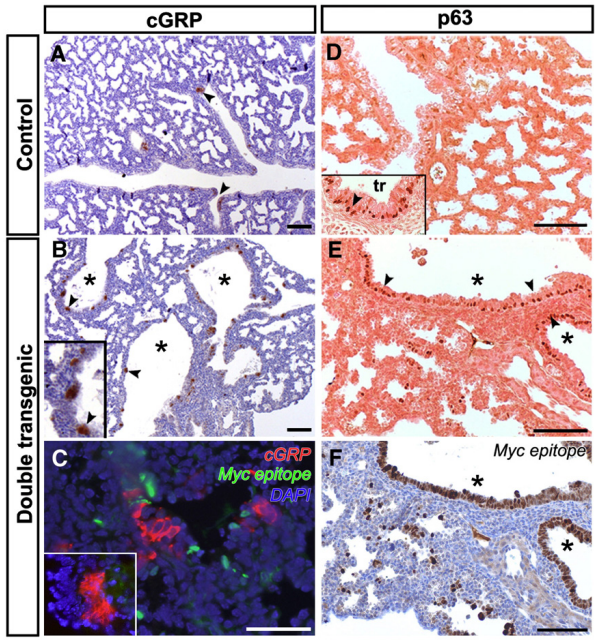


Fig. 9. Sox2 overexpression results in the increase in neuroepithelial cells and (pre-) basal cells. Staining of cGRP (A–C), p63 (D–F) and myc epitope (C, F) in lungs at E18.5 from control (A, D) and Sox2 overexpressing mice (B, C, E, F). cGRP positive staining (neuroepithelial cells) was only detected at the bronchoalveolar ducts in controls (A, arrowheads), and p63 (basal cells) was only expressed in the basal cell layer of the trachea (D, inset). Increased staining for cGRP (B) and p63 (E) was detected in the abnormal airway epithelium expanding into the enlarged airspaces. Note that cGRP positive cells are negative for the Sox2 transgene (C), while Sox2 transgenic expression colocalizes with p63 (E, F). Note that there is no myc-epitope expression in the control lungs (C, insert). Scale bars: 100 μm.

to the Fgf10, because the overexpression of Sox2 is not uniform and the epithelial cells not overexpressing Sox2 are able to respond to Fgf10.

The Sox2 overexpressing lungs appear to have a comparable phenotype to that seen when *Bmpr1a* or *Nmyc* are conditionally deleted (Eblaghie et al., 2006; Okubo et al., 2005). Recently, Shu et al. (2005) demonstrated by inhibiting Wnt/β-catenin signaling that Wnt signaling is upstream of Bmp, Nmyc and Fgf. In addition, Que et al. (2007) showed that Fgf10 signaling is upstream of Sox2 and negatively influences its transcription. Therefore, we hypothesize that Wnt and Fgf signaling has no (direct) influence on the Sox2 overexpression phenotype described. We did not observe differences in β-catenin staining pattern between lungs from controls and Sox2 overexpressors (data not shown). Nmyc is expressed in the proliferative cells of the distal epithelium, and overexpression leads to a loss of the distinct Sox2 boundary between the terminal bronchioli and the future alveoli, which is normally observed. This indicates that the expansion of proliferative cells diminishes the number of Sox2 positive cells, and thus the number of committed cells. When Nmyc is deleted from the epithelium, the gross appearance of the lungs resembles the overexpression of Sox2 phenotype. This could be due to the fact that cells are not longer maintained in the proliferative stage, but rather are initiating a differentiation program.

Bmp4 is expressed as the lung starts to branch, and may be required to keep cells proliferating, and prevent them from premature differentiation. Inhibiting Bmp4 signaling in the lung by conditional deletion of *Bmprla* results in a proximalization of the lung, resembling the Sox2 overexpressing lungs. This could indicate that Sox2 and Bmp4 have antagonistic effects, but this is merely speculative. Bmp4, like Nmyc, would maintain the cells in branching regions in a branch responsive stage, to allow them to branch again. Moreover, *Nmyc* expression is reduced in the conditional *Bmprla* mutant, suggesting that cell proliferation and survival in part are regulated through the synergistic action of Bmp and Nmyc. Sox2 on the other hand, will drive cells into the differentiation program and a loss to respond to branch inducing agents. However, the precise role of Bmp4 in early lung development is still elusive, because of contradictionary results obtained (Eblaghie et al., 2006; Cardoso and Lu, 2006). Overexpression of Bmp4 in lung epithelium results in a phenotype comparable to the Bmprla conditional mutant (Bellusci et al., 1996). Moreover, SPC promoter-driven overexpression of Noggin or Gremlin, two antagonists of Bmp, resulted in smaller lungs with a proximalized phenotype, slightly different from the Bmprla mutant (Lu et al., 2001; Weaver et al., 1999). These results indicate that the level of Bmp4 expression is critical for its activity and effect on cells.

So, both Nmyc and Bmp4 are necessary to maintain cells in a proliferative and branch responsive state early in lung epithelial branching, whereas Sox2 is expressed in cells that are committed for differentiation. Whether Nmyc, Bmp4 and Sox2 act in the same pathway to counteracts each other's function, or in parallel pathways that are influenced independently remains to be resolved.

The observation by Que et al. (2007) that Sox2 expression is upregulated in the Ttf1 (Nkx2.1) knockout mice, which has a failure to separate the foregut into trachea and esophagus, supports our find-

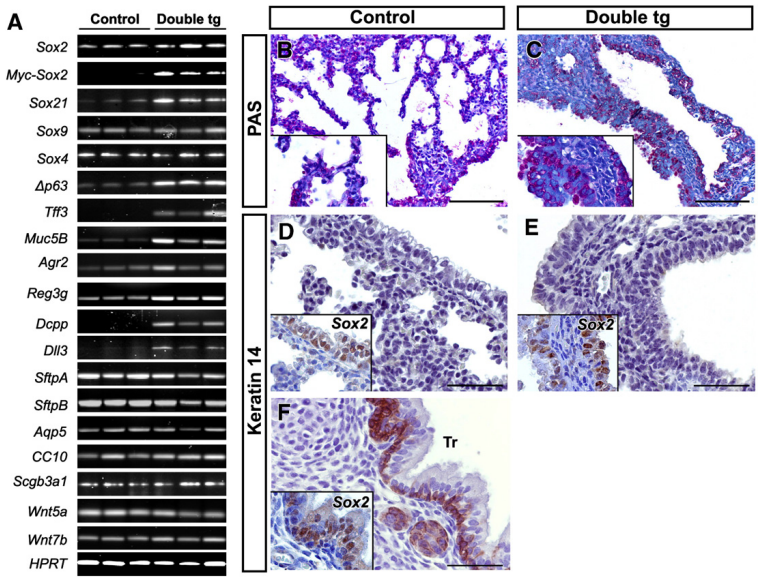


Fig. 10. Differential expression of specific markers. (A) Semi-quantitative RT-PCR for expression of *Sox2* (*endogenous and transgene*), *myc-Sox2* (*transgene-specific*), *Sox21*, *Sox9*, *Sox4* (Sox family), *ΔNp63* (p63 isoform), *Tff3*, *Muc5B*, *Agr2*, *Reg3g*, *Dcpp*, *Dll3* (gastric/intestinal markers), *SftpA*, *SftpB*, *Aqp5*, *CC10*, *Scgb3a1* (lung epithelial markers) and *Wnt5a*, *Wnt7b* and *HPRT* was performed on RNA isolated from three control and three double transgenic E18.5 lungs. Note the significant increase of *Sox2* in double transgenic lungs. Upregulation of *Sox21*, *ΔNp63*, *Tff3*, *Muc5B*, *Agr2*, *Reg3g*, *Dcpp* and *Dll3* confirmed our microarray data. PAS staining at E18.5 showed a marked increase in Sox2 overexpressing lungs (C) compared to controls (B). Keratin14 is absent in distal airways in control and double transgenic lungs (D, E), but present in tracheal epithelium (F). Sox2 is expressed in both epithelia (D–F, inserts). Tr, trachea. Scale bars: 100 μm (B, C), 50 μm (D–F).

ings. Normally, Ttf1 and Sox2 have a reciprocal expression pattern in the esophagus/trachea (Fig. 1A, insert) and the Ttf1 mouse mutant ectopically expresses Sox2 in the ventral part of the foregut. As a result, the primary lung buds do not branch into subsequent airways, but instead form air sacs, supporting our data that downregulation of Sox2 is necessary for the lung epithelium to branch (Minoo et al., 1999).

Overexpression of an inducible Sox17 transgene with the SPC-rtTA resulted in reduced branching morphogenesis, dilated tubules and decreased differentiation of distal epithelium (Park et al., 2006). However, the Sox17-induced defects are less severe than in the Sox2 overexpressing lungs. We also observed dilated tubules, but these tubules end in enlarged airspaces, whereas the sox17 lungs develop some alveolar structure. In contrast to their observation, we did not observe changes in the number of type I and type II pneumocytes. Differences in the effects of overexpressing Sox17 may result from the fact that endogenous Sox17 is only expressed in the mesenchyme and primary vessels. It is not until gestational age 18 that Sox17 is detected in the bronchioles. The observation that Sox17 induces an increase in the number of neuroepithelial cells could be caused by overlapping activities of Sox17 and Sox2. However, conditional deletion of Sox9 in the developing airways did not lead to overt morphological changes, so other Sox family members may have overlapping functions (Perl et al., 2005).

Sox2 and epithelial differentiation

The pulmonary epithelium of Sox2 overexpressing mice contains elevated levels of committed progenitors, like neuroendocrine cells and p63 positive cells. The neuroendocrine cells are the first type of cells that emerge during development and are partly a transient cell type, while the p63 positive cells are mainly expressed in the basal, progenitor-type of cells (Candi et al., 2007; Reynolds et al., 2000; Warburton et al., 1998). In other organ systems, like the developing neural tube, cochlea and tongue, overexpression of Sox2 leads to an increase of committed progenitor cells and a reduced number of fully differentiated cells (Bylund et al., 2003; Graham et al., 2003; Kiernan et al., 2005; Okubo et al., 2006; Pevny and Placzek, 2005; Que et al., 2007). This supports our findings that Sox2 drives epithelial cells prematurely into a committed progenitor cell type, resulting in cells with sustained Sox2 expression that are not responsive to branch-inducing agents. They are still able to differentiate into the various lung-specific cell lineages due to other signaling cues. While cells are differentiating, the transgene can be switched off in certain lineages, as we and others have observed (see double staining in Fig. 9C; Okubo and Hogan, 2004; Park et al., 2006).

Neuroendocrine cells

The increase in the number of neuroendocrine cells parallels the microarray results that the *cGRP* and proneural gene *Mash1* are considerably upregulated. Like in the inner ear (Kiernan et al., 2005), Sox2 upregulates *Mash1* expression in lung epithelium and subsequently drives cells into the neuroendocrine fate (Borges et al., 1997; Ito et al., 2000). Bylund et al. (2003) showed that B1 Sox genes (*Sox 1–3*) keep neural cells undifferentiated by blocking the downstream targets of pro-neural genes. Sandberg et al. (2005) showed that proneural genes induced *Sox21* expression and Sox21 has the opposite effect on neural development. It counteracts the activity of Sox2, thus the Sox2/Sox21 balance determines whether neural cells remain

progenitors or start to differentiate. Strikingly, we found a strong induction of Sox21, suggesting that a similar mechanism is active in the formation of neuroendocrine cells. The *Sox2* transgene is expressed from the SPC promoter, and hence Sox21 cannot switch off the *Sox2* transgene, but neuroendocrine cells loose SPC promoter activity (Fig. 9C).

Overexpression of *Mash1* in airway epithelium using the CC10 promoter leads to a generalized epithelial dysplasia and focal hyperplasia, which resembles some of our findings (Linnoila et al., 2000). This phenomenon is attributed to a disturbed balance between sustained proliferation in terminal bronchioles and resistance to apoptosis (Wang et al., 2007). Furthermore, the observed hyperplasia resembles a human condition known as bronchiolization of alveoli (BOA), a premalignant lesion of human lung carcinomas. Close examination of the dilated air sacs in our Sox2 mice showed that indeed the most proximal part of these alveolar structures resembled the phenotypic characteristics of the Mash1 overexpressing mice. Interestingly, the authors found that although overexpression of Mash1 leads to hyperplasia and metaplasia of the airway epithelium, no overt neuroendocrine differentiation was observed. This could be caused by the fact that these authors used the CC10 promoter, which is expressed at later stages and in different cells than the SPC promoter. Neuroendocrine hyperplasia was also observed in a variety of conditions, including hypoplasia of the lung in case of bronchopulmonary dysplasia, congenital diaphragmatic hernia and sudden infant death syndrome (Warburton et al., 1998).

p63 overexpressing cells

Overexpression of Sox2 in airway epithelium leads to a dramatic increase in the number of p63 expressing cells. Two groups showed that p63 plays a critical role in the development of stratified epithelium (Mills et al., 1999; Yang et al., 1999). Two different promoters lead to two major isoforms, one with an N-terminal Trans Activation domain (TAp63), and one without this domain (ΔNp63). Sox2 overexpression induces ΔNp63 expression, which is implicated in the maintenance of the proliferative potential of basal cells (Candi et al., 2006). However, another group indicated that ΔNp63 is necessary to counteract the TAp63 activity in order for cells to differentiate (Koster et al., 2004). In spite of the ongoing debate, partly caused by the usage of different approaches and mouse strains, most data seem to point at the fact that the ΔNp63 isoform is expressed and mandatory in a progenitor or committed precursor cell (Candi et al., 2007; Koster et al., 2007). This supports our findings that overexpression of Sox2 leads to a premature differentiation of cells into committed progenitors. In addition, Que et al. (2007) found a correlation between the level of Sox2 and p63, inversely correlating with Ttf1 expression in the esophagus. We clearly observe a high number of Sox2 and p63 positive cells, but Ttf1 is still expressed, indicating that ectopic overexpression of Sox2 may not necessarily result in a downregulation of Ttf1 in the lung.

Intestinal cells

Our microarray data unexpectedly showed genes that were characteristic for secretory cells of the posterior glandular stomach and intestine, like *Tff3*, *Muc5B*, *Agr2* and *Reg3g*. In the esophagus of mice that express reduced amounts of Sox2, the same markers are expressed, except for *Tff3* (Que et al., 2007). Although the esophagus epithelium differs from the airway epithelium, it is striking to observe the same genes to be upregulated. However, in mice that develop a tracheoesophageal fistula, which have the most severe reduction of Sox2 expression, it seems that these gastric markers are not prominently expressed (Que et al., 2007). In addition, overexpression of an upstream mediator of Sox2 expression, β-catenin, resulted in an increase of the same type of gastric and intestinal marker genes (Okubo and Hogan, 2004).

Concluding remarks

Our data clearly show that overexpression of Sox2 leads to a block in the ability of the airways to branch and form bronchial–alveolar structures. Furthermore, differentiation of the epithelium is abnormal and forced Sox2 expression results in formation of progenitor cells. We propose that Sox2 prematurely induces cells into a differentiation program, like neuroendocrine and (pre-) basal cells, thereby rendering these cells unresponsive to branch-inducing signals and causing a developmental arrest (Fig. 11). This phenotype agrees with previous Sox2 overexpressing studies in the neural tube, eye, cochlea and tongue, where elevated numbers of precursor or committed progenitors were detected (Bylund et al., 2003; Graham et al., 2003; Kiernan et al., 2005; Okubo et al., 2006; Pevny and Placzek, 2005; Que et al., 2007). The premature commitment of cells, however, does not lead to an absence of differentiation of neighboring cells like Clara cells,

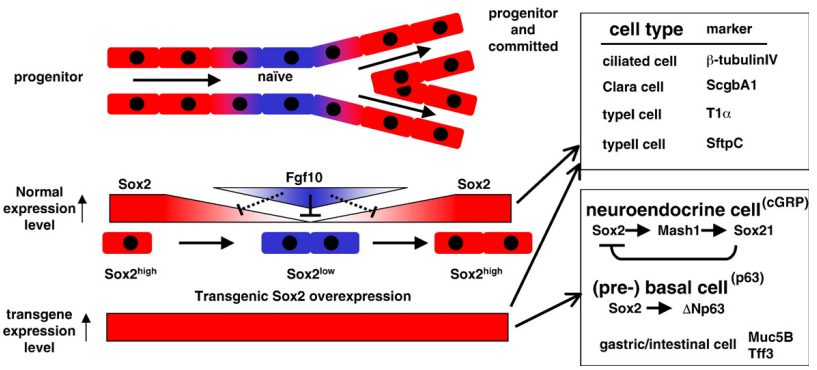


Fig. 11. Model of Sox2 action. Airway epithelial cells express Sox2 before they are induced to branch (progenitor). Fgf10 was shown to downregulate Sox2 (Que et al., 2007), thereby preparing cells to branch (naive). After branching, cells start to express Sox2 again and become either progenitor, which can branch further, or committed, which enters differentiation pathways (box, top right). Overexpression of Sox2 prematurely drives cells into committed progenitor cells, like neuroendocrine and (pre-) basal cells (box, bottom right), rendering these cells insensitive to branch inducing signals. See Discussion for a detailed description of the pathway.

ciliated cells, type I and type II cells, because the SPC-driven transgene is silenced in some committed progenitor cells as shown for neuroendocrine cells.

Acknowledgments

We thank Marjon Buscop-van Kempen and Frank van der Panne for technical assistance. We are also indebted to Dr. Paul Scotting and Martin Post for stimulating discussions, Jeffrey Whitsett for generously supplying the SPC-rtTA mice, Barry Stripp for providing the CC10 antibody, Ronald de Krijger for pathology and Sigrid Swagemaker for normalizing the microarray data.

This work was supported in part by the Sophia Foundation for Medical Research, SSWO nr. 413.

References

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

Bellusci, S., Henderson, R., Winnier, G., Oikawa, T., Hogan, B.L., 1996. Evidence from normal expression and targeted misexpression that bone morphogenetic protein (Bmp-4) plays a role in mouse embryonic lung morphogenesis. Development 122, 1693–1702.

Bellusci, S., Furuta, Y., Rush, M.G., Henderson, R., Winnier, G., Hogan, B.L., 1997a. Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. Development 124, 53–63.

Bellusci, S., Grindley, J., Emoto, H., Itoh, N., Hogan, B.L., 1997b. Fibroblast growth factor 10 (FGF10) and branching morphogenesis in the embryonic mouse lung. Development 124, 4867–4878.

Borges, M., Linnoila, R.I., van de Velde, H.J., Chen, H., Nelkin, B.D., Mabry, M., Baylin, S.B., Ball, D.W., 1997. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature 386, 852–855.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., Young, R.A., 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956.

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

Candi, E., Rufini, A., Terrinoni, A., Dinsdale, D., Ranalli, M., Paradisi, A., De Laurenzi, V., Spagnoli, L.G., Catani, M.V., Ramadan, S., Knight, R.A., Melino, G., 2006. Differential roles of p63 isoforms in epidermal development: selective genetic complementation in p63 null mice. Cell Death Differ. 13, 1037–1047.

Candi, E., Dinsdale, D., Rufini, A., Salomoni, P., Knight, R.A., Mueller, M., Krammer, P.H., Melino, G., 2007. TAp63 and DeltaNp63 in cancer and epidermal development. Cell Cycle 6, 274–285.

Canis Parera, M., van Dooren, M., van Kempen, M., de Krijger, R., Grosveld, F., Tibboel, D., Rottier, R., 2005. Distal angiogenesis: a new concept for lung vascular morphogenesis. Am. J. Physiol.: Lung Cell. Mol. Physiol. 288, L141–L149.

Cardoso, W.V., Lu, J., 2006. Regulation of early lung morphogenesis: questions, facts and controversies. Development 133, 1611–1624.

Casarosa, S., Fode, C., Guillemot, F., 1999. Mash1 regulates neurogenesis in the ventral telencephalon. Development 126, 525–534.

Daniely, Y., Liao, G., Dixon, D., Linnoila, R.I., Lori, A., Randell, S.H., Oren, M., Jetten, A.M., 2004. Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium. Am. J. Physiol.: Cell Physiol. 287, C171–C181.

De Moerlooze, L., Spencer-Dene, B., Revest, J., Hajihosseini, M., Rosewell, I., Dickson, C., 2000. An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal–epithelial signalling during mouse organogenesis. Development 127, 483–492.

Eblaghie, M.C., Reedy, M., Oliver, T., Mishina, Y., Hogan, B.L., 2006. Evidence that autocrine signaling through Bmpr1a regulates the proliferation, survival and morphogenetic behavior of distal lung epithelial cells. Dev. Biol. 291, 67–82.

Eiraku, M., Tohgo, A., Ono, K., Kaneko, M., Fujishima, K., Hirano, T., Kengaku, M., 2005. DNER acts as a neuron-specific Notch ligand during Bergmann glial development. Nat. Neurosci. 8, 873–880.

Graham, V., Khudyakov, J., Ellis, P., Pevny, L., 2003. SOX2 functions to maintain neural progenitor identity. Neuron 39, 749–765.

Ishii, Y., Rex, M., Scotting, P.J., Yasugi, S., 1998. Region-specific expression of chicken Sox2 in the developing gut and lung epithelium: regulation by epithelial–mesenchymal interactions. Dev. Dyn. 213, 464–475.

Ito, T., Udaka, N., Yazawa, T., Okudela, K., Hayashi, H., Sudo, T., Guillemot, F., Kageyama, R., Kitamura, H., 2000. Basic helix–loop–helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. Development 127, 3913–3921.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., Lemischka, I.R., 2006. Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533–538.

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

Kiernan, A.E., Pelling, A.L., Leung, K.K., Tang, A.S., Bell, D.M., Tease, C., Lovell-Badge, R., Steel, K.P., Cheah, K.S., 2005. Sox2 is required for sensory organ development in the mammalian inner ear. Nature 434, 1031–1035.

Koster, M.I., Kim, S., Mills, A.A., DeMayo, F.J., Roop, D.R., 2004. p63 is the molecular switch for initiation of an epithelial stratification program. Genes Dev. 18, 126–131.

Koster, M.I., Dai, D., Roop, D.R., 2007. Conflicting roles for p63 in skin development and carcinogenesis. Cell Cycle 6, 269–273.

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

Li, C., Hu, L., Xiao, J., Chen, H., Li, J.T., Bellusci, S., Delanghe, S., Minoo, P., 2005. Wnt5a regulates Shh and Fgf10 signaling during lung development. Dev. Biol. 287, 86–97.

Linnoila, R.I., Zhao, B., DeMayo, J.L., Nelkin, B.D., Baylin, S.B., DeMayo, F.J., Ball, D.W., 2000. Constitutive achaete-scute homologue-1 promotes airway dysplasia and lung neuroendocrine tumors in transgenic mice. Cancer Res. 60, 4005–4009.

Litingtung, Y., Lei, L., Westphal, H., Chiang, C., 1998. Sonic hedgehog is essential to foregut development. Nat. Genet. 20, 58–61.

Lu, M.M., Yang, H., Zhang, L., Shu, W., Blair, D.G., Morrisey, E.E., 2001. The bone morphogenic protein antagonist gremlin regulates proximal-distal patterning of the lung. Dev. Dyn. 222, 667–680.

Maeda, Y., Dave, V., Whitsett, J.A., 2007. Transcriptional control of lung morphogenesis. Physiol. Rev. 87, 219–244.

Mailleux, A.A., Tefft, D., Ndiaye, D., Itoh, N., Thiery, J.P., Warburton, D., Bellusci, S., 2001. Evidence that SPROUTY2 functions as an inhibitor of mouse embryonic lung growth and morphogenesis. Mech. Dev. 102, 81–94.

Mills, A.A., Zheng, B., Wang, X.J., Vogel, H., Roop, D.R., Bradley, A., 1999. p63 is a p53 homologue required for limb and epidermal morphogenesis. Nature 398, 708–713.

Min, H., Danilenko, D.M., Scully, S.A., Bolon, B., Ring, B.D., Tarpley, J.E., DeRose, M., Simonet, W.S., 1998. Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to *Drosophila* branchless. Genes Dev. 12, 3156–3161.

Minoo, P., Su, G., Drum, H., Bringas, P., Kimura, S., 1999. Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(−/−) mouse embryos. Dev. Biol. 209, 60–71.

Okubo, T., Hogan, B.L., 2004. Hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm. J. Biol. 3, 11.

Okubo, T., Knoepfler, P.S., Eisenman, R.N., Hogan, B.L., 2005. Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation. Development 132, 1363–1374.

Okubo, T., Pevny, L.H., Hogan, B.L., 2006. Sox2 is required for development of taste bud sensory cells. Genes Dev. 20, 2654–2659.

Park, W.Y., Miranda, B., Lebeche, D., Hashimoto, G., Cardoso, W.V., 1998. FGF-10 is a chemotactic factor for distal epithelial buds during lung development. Dev. Biol. 201, 125–134.

Park, K.S., Wells, J.M., Zorn, A.M., Wert, S.E., Whitsett, J.A., 2006. Sox17 influences the differentiation of respiratory epithelial cells. Dev. Biol. 294, 192–202.

Perl, A.K., Tichelaar, J.W., Whitsett, J.A., 2002. Conditional gene expression in the respiratory epithelium of the mouse. Transgenic Res. 11, 21–29.

Perl, A.K., Kist, R., Shan, Z., Scherer, G., Whitsett, J.A., 2005. Normal lung development and function after Sox9 inactivation in the respiratory epithelium. Genesis 41, 23–32.

Pevny, L., Placzek, M., 2005. SOX genes and neural progenitor identity. Curr. Opin. Neurobiol. 15, 7–13.

Que, J., Okubo, T., Goldenring, J.R., Nam, K.T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., Hogan, B.L., 2007. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 134, 2521–2531.

Rao, S., Orkin, S.H., 2006. Unraveling the transcriptional network controlling ES cell pluripotency. Genome Biol. 7, 230.

Reynolds, S.D., Giangreco, A., Power, J.H., Stripp, B.R., 2000. Neuroepithelial bodies of pulmonary airways serve as a reservoir of progenitor cells capable of epithelial regeneration. Am. J. Pathol. 156, 269–278.

Sandberg, M., Kallstrom, M., Muhr, J., 2005. Sox21 promotes the progression of vertebrate neurogenesis. Nat. Neurosci. 8, 995–1001.

Schepers, G., Teasdale, R., 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.

Schlaeger, T.M., Bartunkova, S., Lawitts, J.A., Teichmann, G., Risau, W., Deutsch, U., Sato, T.N., 1997. Uniform vascular-endothelial-cell-specific gene expression in both embryonic and adult transgenic mice. Proc. Natl. Acad. Sci. U. S. A. 94, 3058–3063.

Sekine, K., Ohuchi, H., Fujiwara, M., Yamasaki, M., Yoshizawa, T., Sato, T., Yagishita, N., Matsui, D., Koga, Y., Itoh, N., Kato, S., 1999. Fgf10 is essential for limb and lung formation. Nat. Genet. 21, 138–141.

Serls, A.E., Doherty, S., Parvatiyar, P., Wells, J.M., Deutsch, G.H., 2005. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. Development 132, 35–47.

Shu, W., Guttentag, S., Wang, Z., Andl, T., Ballard, P., Lu, M.M., Piccolo, S., Birchmeier, W., Whitsett, J.A., Millar, S.E., Morrisey, E.E., 2005. Wnt/beta-catenin signaling acts upstream of N-myc, BMP4, and FGF signaling to regulate proximal–distal patterning in the lung. Dev. Biol. 283, 226–239.

Sutherland, D., Samakovlis, C., Krasnow, M.A., 1996. branchless encodes a *Drosophila* FGF homolog that controls tracheal cell migration and the pattern of branching. Cell 87, 1091–1101.

Taranova, O.V., Magness, S.T., Fagan, B.M., Wu, Y., Surzenko, N., Hutton, S.R., Pevny, L.H., 2006. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev. 20, 1187–1202.

van Tuyl, M., Liu, J., Wang, J., Kuliszewski, M., Tibboel, D., Post, M., 2005. Role of oxygen and vascular development in epithelial branching morphogenesis of the developing mouse lung. Am. J. Physiol.: Lung Cell. Mol. Physiol. 288, L167–L178.
Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.L., Wert, S., Stahlman, M.T., Whitsett, J.A., 2005. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. J. Biol. Chem 280, 13809–13816.
Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., Orkin, S.H., 2006. A protein interaction network for pluripotency of embryonic stem cells. Nature 444, 364–368.
Wang, X.Y., Dakir el, H., Naizhen, X., Jensen-Taubman, S.M., DeMayo, F.J., Linnoila, R.I., 2007. Achaete-scute homolog-1 linked to remodeling and preneoplasia of pulmonary epithelium. Lab. Invest. 87, 527–539.
Warburton, D., Wuenschell, C., Flores-Delgado, G., Anderson, K., 1998. Commitment and differentiation of lung cell lineages. Biochem. Cell. Biol. 76, 971–995.
Weaver, M., Yingling, J.M., Dunn, N.R., Bellusci, S., Hogan, B.L., 1999. Bmp signaling regulates proximal-distal differentiation of endoderm in mouse lung development. Development 126, 4005–4015.
Weaver, M., Dunn, N.R., Hogan, B.L., 2000. Bmp4 and Fgf10 play opposing roles during lung bud morphogenesis. Development 127, 2695–2704.
Wilson, M., Koopman, P., 2002. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr. Opin. Genet. Dev. 12, 441–446.
Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R.T., Tabin, C., Sharpe, A., Caput, D., Crum, C., McKeon, F., 1999. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature 398, 714–718.

Chapter

4

Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors

Journal of Cell Biology, in press, 2009

Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors

Cristina Gontan[1,*], Thomas Güttler[2,*], Erik Engelen[3], Jeroen Demmers[4], Maarten Fornerod[5], Frank G. Grosveld[3], Dick Tibboel[1], Dirk Görlich[2], Raymond A. Poot[3,#] and Robbert J. Rottier[1,#]

[1]Department of Pediatric Surgery, Erasmus MC, Rotterdam, The Netherlands, [2]Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany, [3]Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands, [4]Proteomics Center, Erasmus MC, Rotterdam, The Netherlands, [5]Netherlands Cancer Institute, Amsterdam, The Netherlands

*: these authors contributed equally to this work

#: these authors contributed equally to this work as senior authors

Correspondence to Raymond A. Poot (r.poot@erasmusmc.nl) or Robbert Rottier (r.rottier@erasmusmc.nl)

ABSTRACT

SRY and other Sox-type transcription factors are important developmental regulators with various implications in human disease. Here, we identified Exportin 4 (Exp4) as an interaction partner of Sox2 in mouse embryonic stem cells and neural progenitors. We show that Exp4 acts, besides its established function in nuclear export, as a bona fide nuclear import receptor for Sox2 and SRY. Exp4 is thus an example of a nuclear transport receptor carrying distinct cargoes into different directions. Different from a published study, we observed that the import-activity of importin-αs towards Sox2 is negligible compared to Exp4. Instead, we found that Imp9 and the Impβ/7 heterodimer mediate nuclear import of Sox2 in parallel to Exp4. Import signals for the three pathways overlap and include conserved residues in the Sox2 HMG box that are also critical for DNA-binding. This suggests that nuclear import of Sox proteins is facilitated by several parallel import pathways.

INTRODUCTION

Sox proteins are important regulators of different stages in embryonic development such as sex determination, neural development, bone formation and neural crest development. They share the so-called non-canonical HMG box domain, an 80 amino acid domain that confers DNA-binding activity and sequence-specificity [1]. Sox2 is required in both embryonic and extra-embryonic tissues; the lack of Sox2 results in peri-implantation lethality [2]. Sox2, together with Oct4 and Nanog, is required for embryonic stem (ES) cell self-renewal and pluripotency [3], and is one of the factors needed for reprogramming somatic cells into induced pluripotent stem (iPS) cells [4]. Later in development, Sox2 is required for neural progenitor expansion and maintenance [5, 6], eye development [7], development of tastebuds [1], inner ear formation [8], and development of the trachea and esophagus [9]. Sox2 mutations in humans cause anophthalmia-esophageal-genital (AEG) syndrome [10]. It is believed that Sox proteins, including Sox2, exert their specificity by synergistic binding with other transcription factors to DNA. For Sox2, partners like Oct4, Pax6, and Brn2 have been found [1].

In this study, we identified Exportin 4 (Exp4) as an interaction partner of Sox2 in mouse ES cells and neural progenitors and show that it serves, besides its established function in nuclear export [11, 12], as a *bona fide* nuclear import receptor for Sox2 and SRY. In contrast to earlier findings [13], we found that the import activity of importin-α isoforms towards Sox2 is negligible compared to Exp4. Instead, we observed that Imp9 and the Impß/7 heterodimer operate as import mediators of Sox2 in parallel to Exp4, suggesting that regulation of Sox2 function is most likely not dependent on a single nuclear import mechanism. The import signals for the three pathways overlap and include conserved residues in the Sox HMG box that are critical for *in vivo* nuclear localisation and DNA binding.

RESULTS AND DISCUSSION

Exportin 4 is a novel interaction partner of Sox2

To identify Sox2-interactors by affinity-purification, we introduced a FLAG-Sox2 transgene into mouse ES cells and isolated individual clones stably expressing the fusion protein. A selected clone expressed FLAG-Sox2 at approximately 30% of the endogenous Sox2 level (**Fig. S1a**, lower left panel). These cells displayed normal ES cell growth behaviour and expressed the ES cell marker Oct4 (**Fig. S1a**). We then performed anti-FLAG immunoprecipitation experiments with extracts of both FLAG-Sox2 and control ES cells and analysed the bound fractions. One of the predominant Sox2 binding partners, migrating at ≈ 120 kD (**Fig. 1a**), was identified by mass spectrometry as Exportin 4 (Exp4, 39 unique peptides). We confirmed the identity of Exp4 by Western blotting (unpublished data) and showed that endogenous (untagged) Sox2 also interacts with Exp4 (**Fig. 1b**). We also analysed anti-FLAG-immunoprecipitates from extracts of FLAG-Sox2-ES cell-derived neural progenitors [14]) by mass spectrometry and again found the prominent presence of Exp4 (27 unique peptides), which was confirmed by co-immunoprecipitation of Exp4 with endogenous Sox2 (**Fig. 1c**). The addition of benzonase (benzo) or ethidium bromide (EtBr) did not abolish the interaction, which suggested that binding is independent of DNA (**Fig 1b,c**).

Exp4 is a nuclear export receptor for eIF5A and the transcription factor SMAD3 [11, 12]. It belongs to the superfamily of Ran-regulated nuclear transport receptors, which includes nuclear import-mediators (importins) and exportins [15-17]. Importins bind cargoes at low RanGTP-levels in the cytoplasm, facilitate translocation through nuclear pore complexes (NPCs), release their load upon RanGTP-binding in the nucleus, and finally return to the cytoplasm, where GTP-hydrolysis dissociates the Ran·importin-complex and allows the importin to bind and import another cargo-molecule. Exportins operate in the opposite manner; they bind cargo together with RanGTP in the nucleus and release their cargo upon GTP-hydrolysis in the cytoplasm. Surprisingly, addition of RanGTP almost completely abolished binding of Exp4 to Sox2 (**Fig. 1d**), suggesting that Exp4 acts as an importin for Sox2. The effect was specific for Ran, as the Rab6 GTPase did not dissociate the Sox2·Exp4 complex (**Fig. 1d**). In line with this observation, activation of the RanGTPase by exogenous RanGAP and RanBP1 [18] did not disrupt the Sox2·Exp4 interaction (**Fig. S1c**).

The observed Sox2·Exp4-binding could have been either direct or mediated by some additional component from the cell extract. We therefore tested whether recombinant GST-Sox2 and Exp4 interact within an *E. coli* extract lacking other components of the nuclear transport machinery. Indeed, GST-Sox2 bound Exp4 in the absence of recombinant RanGTP, but not in its presence (**Fig. 1e**). Interestingly, a Sox2 mutant (mt1, **Fig. 4a**) previously reported to be compromised in import [13] did not interact with Exp4 (**Fig. 1e**). Thus, Exp4 specifically and directly interacts with Sox2 in an importin-like manner.

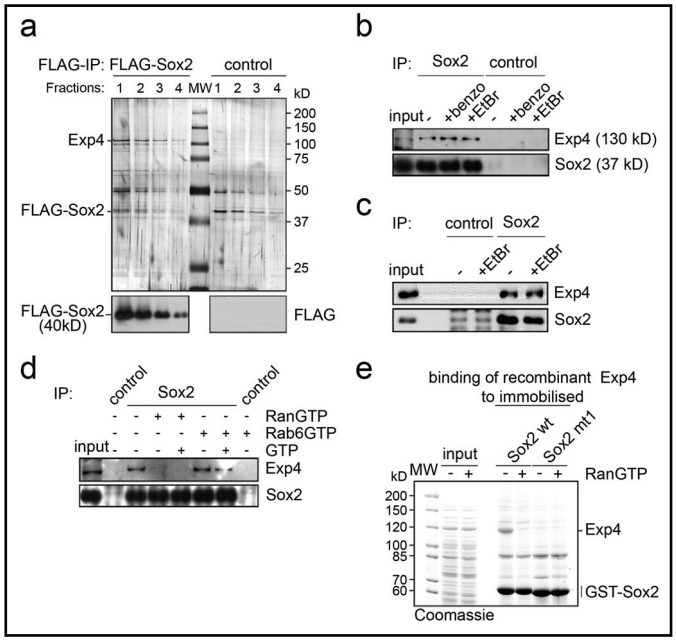


Fig. 1. Exp4 binds to Sox2. (a) Identification of Exp4 as a Sox2 interaction partner by FLAG-affinity purification. Upper panel: FLAG-affinity purifications from FLAG-Sox2 (left) or control-46C ES cell extract (right). Eluted fractions were analysed by SDS-PAGE and silver-staining. Bands representing FLAG-Sox2 and Exportin 4 (Exp4) are indicated. Lower panel: Anti-FLAG immunoblot showing FLAG-Sox2 in the eluted fractions. **(b)** Exp4 co-immunoprecipitated with endogenous Sox2 from ES cell extracts. Where indicated, benzonase ("benzo") or ethidium bromide ("EtBr") had been added to the extract to test for DNA-independence of the Sox2·Exp4 interaction. Input and bound fractions were analysed by SDS-PAGE and immunoblotting. **(c)** Exp4 co-immunoprecipitated with endogenous Sox2 as in (**b**), but from neural progenitor cell extracts. **(d)** Exp4 was co-immunoprecipitated with endogenous Sox2 as in (**b**), but with indicated additions. RanGTP disrupted the Sox2·Exp4 interaction, while the negative control, Rab6GTP, had no effect. **(e)** Recombinant Sox2 and Exp4 interact directly. Immobilised GST-Sox2, but not GST-Sox2 mt1 (see **Fig. 4**), binds recombinant Exp4 from an *E. coli* extract. RanGTP disrupted the GST-Sox2·Exp4 interaction. Input and bound fractions were analysed by SDS-PAGE and Coomassie-staining.

Exportin 4 facilitates nuclear import of Sox2

To test whether Exp4 imports Sox2, we prepared fluorophore-labelled GST-Sox2 for import assays with permeabilised HeLa cells [19, 20]. In the absence of transport receptors, GST-Sox2 failed to accumulate inside nuclei, but precipitated at the cytoplasmic remnants of the permeabilised cells (**Fig. 2a** and **Fig. 2b**) - similar to other nucleic acid-binding proteins such as histone H1 and ribosomal factors [21, 22]. Addition of Exp4, however, strongly stimulated nucleoplasmic accumulation of GST-Sox2 (**Fig. 2a** and **Fig. 2b**). Remarkably, Exp4 also suppressed cytoplasmic aggregation of Sox2 (**Fig. 2a** and **Fig. 2b**), an effect that appeared Exp4 dose-dependent (compare Exp4-panels of Fig. S2a and **Fig. S2b**) but did not require Ran (**Fig. S2b**; see discussion below). These effects were specific to Exp4, as transportin neither imported GST-Sox2 nor affected Sox2 aggregation (**Fig. 2a**). As for all Impβ-like importins, efficient multi-round nuclear import by Exp4 depended on both Ran and energy (**Fig. S2a**) reflecting the need for recycling Exp4·RanGTP complexes back to the cytoplasm. Thus, Exp4 is a *bona-fide* importin for Sox2 and mediates a novel nuclear import pathway. So far, only Imp13 from higher eukaryotes and Msn5 from *S. cerevisiae* have been recognised to transport distinct sets of cargoes into opposite directions through the NPC [23, 24]. It is now tempting to speculate that any exportin could operate as an importin, provided its affinity for RanGTP in the absence of export cargo is sufficiently high for displacing import-substrates into the nucleus. We are currently testing this hypothesis. In contrast to Impβ, but like transportin, Exp4 required RanGTP only to release its import cargo into the nucleus, but not to detach from the nuclear side of the NPC [25].

Importin ß/7 and Importin 9 also facilitate nuclear import of Sox2

Importin-cargo binding is highly specific and governed by nuclear import signals. Importin ß (Impβ) either binds its cargoes directly or pairs with Importin-αs that recognise so-called monopartite or bipartite classical nuclear localisation signals. Previous studies proposed that Sox proteins share a conserved bipartite NLS at the N-terminus of the HMG domain and a monopartite NLS at its C-terminal end (**Figs. 3a** and **4a**;[26, 27] and that Impβ/α3 and Impβ/α5 import Sox2 in neural progenitors [13]. It was also suggested that Impβ alone can import Sox2 in undifferentiated ES cells [13]. Therefore, it was fully unexpected to find Exp4 as a predominant interaction partner (**Fig. 1a**) and efficient nuclear import mediator of Sox2 (**Fig. 2a**). This situation prompted us to also test other members of the Impβ superfamily for import activity towards Sox2. While Exp4 strongly stimulated nuclear accumulation of GST-Sox2, we found no indication for Impβ, Imp4, Imp5, Imp7, Imp8 and Imp13 mediating import of the protein (**Fig. 2b** and **Fig. 2c**). However, the Impβ/7 heterodimer [21] and Imp9 imported GST-Sox2 into nuclei as efficiently as Exp4 (**Fig. 2c**). We obtained virtually identical results using permeabilised embryonic stem cells instead of HeLa cells (**Fig. S3**; see figure legend for the rationale). Thus, Imp9 and the Impβ/7 heterodimer can facilitate nuclear import of Sox2 in parallel to Exp4. Importantly, just as Exp4, Impβ/7 and Imp9 markedly reduced cytoplasmic precipitation

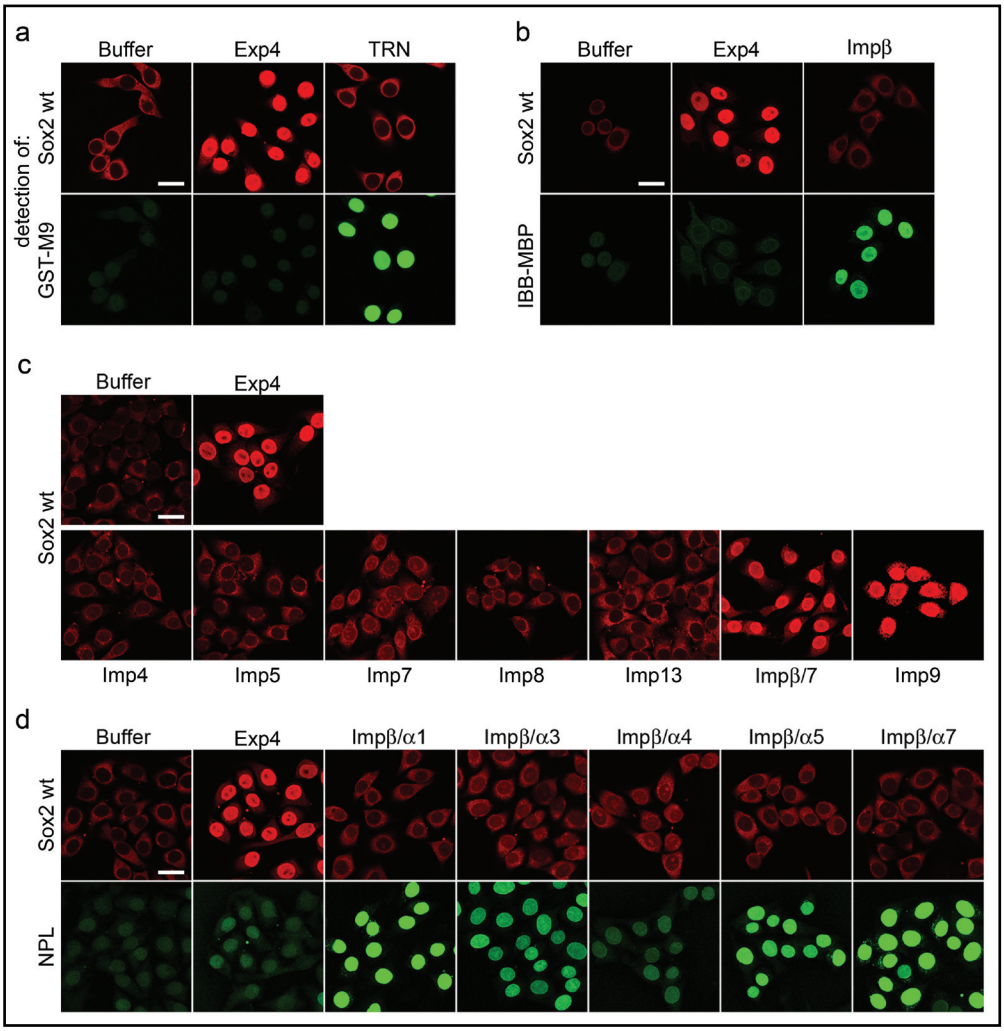


Fig. 2. Exp4, Impß/7 and Imp9, but not Impß alone or Impß/α, facilitate nuclear import of Sox2. Panels show nuclear import of fluorophore-labelled GST-Sox2 (red) and control cargo proteins (green) into nuclei of digitonin-permeabilised HeLa cells with either buffer or the indicated import receptors. All import reactions contained the components of the Ran-cycle and an ATP/GTP-replenishing system (see methods for details). **(a)** Exp4 imported GST-Sox2 into nuclei of permeabilised cells. The transportin (TRN)-substrate GST-M9 had been used as an internal specificity control. **(b)** Impß failed to import GST-Sox2, whereas it efficiently imported its substrate IBB-MBP. **(c)** Apart from Exp4, only Impß/7 and Imp9 efficiently imported GST-Sox2. Note that only the cognate import receptors suppressed cytoplasmic aggregation of GST-Sox2. (See also other panels of this figure, as well as **Fig. 4** and **Fig. S2b**). **(d)** The panel shows GST-Sox2 import by Exp4 compared to the import activity of the indicated Impß/α-combinations. The Impß/α-substrate nucleoplasmin (NPL) served as an internal positive control. Bars, 25 µm.

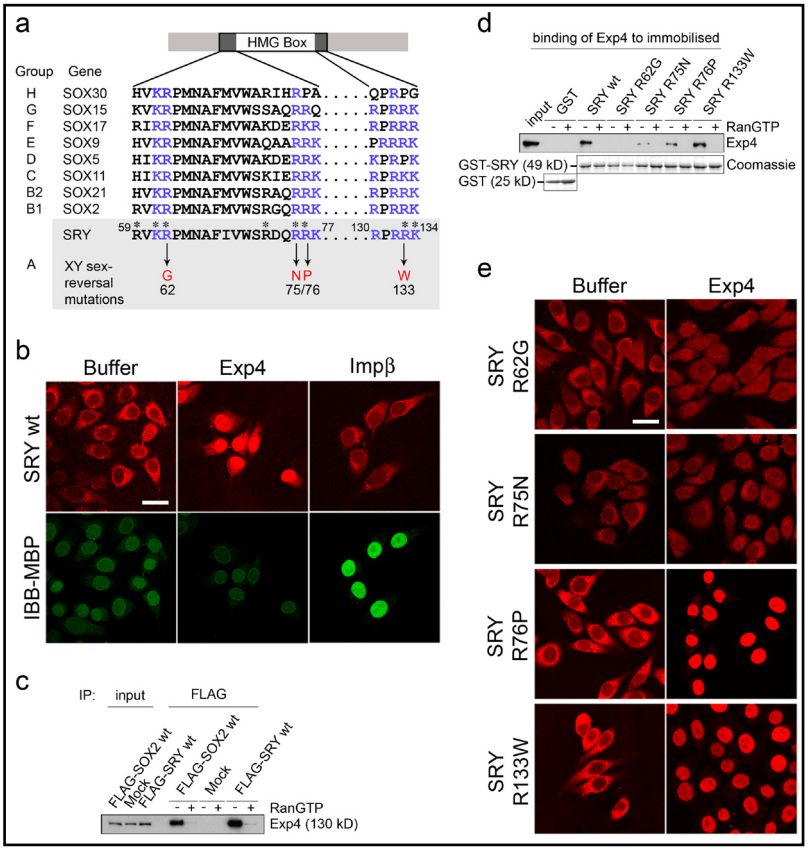


Fig. 3. Exp4 facilitates nuclear import of SRY. **(a)** Sequence alignment of N- and C-terminal conserved regions of the HMG box previously implicated in the nuclear localisation of Sox proteins (selected members of all nine subgroups are shown). Conserved basic residues are indicated in blue, basic residues binding DNA in the SRY·DNA complex [30] are marked (*). Amino acid substitutions of conserved arginines (R62G, R75N, R76P, and R133W) in the HMG box of SRY, originally identified in human patients with XY sex-reversal [43], are shown in red. **(b)** Exp4 mediated import of SRY but not IBB-MBP. The import experiment was as in **Fig. 2b**, but with fluorophore-labelled GST-SRY (red). Impβ imported IBB-MBP, but not SRY. Bar, 25 μm. **(c)** Exp4 was co-precipitated with FLAG-SRY in a RanGTP-sensitive manner. The panel shows FLAG-affinity purifications from extracts of ES cells, which had been transiently transfected with the indicated FLAG-plasmids. "Mock" refers to the "empty" control-FLAG-plasmid. Starting material and bound fractions were analysed by SDS-PAGE and Western blotting. Addition of RanGTP to the extract disrupted the SRY·Exp4 interaction. **(d)** SRY sex reversal mutations affected the SRY interaction with Exp4 to different degrees. The figure shows a GST-pulldown from ES cell extracts with GST-SRY or the indicated mutants. Upper panel: Eluted Exp4 was detected by SDS-PAGE and Western blotting. Lower panels: Eluted GST and GST-SRY fusions were detected by SDS-PAGE and Coomassie-staining. Addition of RanGTP to the extract disrupted the SRY·Exp4 interaction as in **(c)**.

of GST-Sox2 (**Fig. 2c**). We found no evidence for any Impα/ß combination stimulating Sox2-import, although they all imported the internal positive control, nucleoplasmin (**Fig. 2d**). Only with Impα3 and Impα4, GST-Sox2 showed slight nucleolar accumulation, which was not seen for endogenous or over-expressed Sox2 (unpublished data) and thus probably reflects mis-localised Sox2. In contrast to Exp4, none of the Impα/ß combinations suppressed cytoplasmic aggregation of GST-Sox2 (**Fig. 2d**). In summary, our comprehensive survey identified three import pathways for Sox2, all of which meet the criteria for cognate import receptors, namely stimulation of nuclear import and suppression of import substrate aggregation. In contrast, Impß alone and in conjunction with Impαs met none of these criteria.

SRY is imported by Exportin 4 through conserved residues of its HMG box

Of the three import pathways identified, Exp4 represents a novel mediator of nuclear import and it is therefore of interest to know whether Sox2 represents the sole import cargo of this pathway. Hence, we tested whether Exp4 also imports SRY - the prototypic Sox family member [28], which shares in its HMG box domain 85% identity with Sox2 (**Fig. 3a**) and whose loss of function leads to XY sex reversal in humans [1]. Indeed, fluorophore-labelled GST-SRY behaved identically to GST-Sox2, showing pronounced nuclear accumulation upon Exp4-addition and nuclear exclusion with cytoplasmic precipitation when only buffer or Impß had been added (**Fig. 3b**). Consistent with this result, FLAG-SRY co-immunoprecipitated Exp4 from ES cell extracts in a RanGTP-sensitive manner (**Fig. 3c**).

Interestingly, some SRY mutations in XY females have been reported to impair SRY nuclear localisation, with R62G and R75N causing the most drastic defect, while R76P and R133W only display a minor (if any) change (**Fig. 3a**; [29]). These effects were previously allocated to defects in Impß-dependent import (R62G, R133W) or as yet unidentified import pathways (R75N, R76P) [29]. However, as our data argue against a significant contribution of Impß to SRY nuclear localisation (**Fig. 3b**), we tested whether the SRY mutants are compromised in recruiting Exp4 from an ES cell extract. Strikingly, Exp4-binding was completely lost for the R62G mutant and strongly impaired for SRY R75N while SRY R76P retained a reduced Exp4-binding activity (**Fig. 3d**). R133W had no detectable effect on Exp4-binding (**Fig. 3d**). This correlation with the previously observed localisation defects [29] was perfectly re-capitulated by our nuclear import assays: while Exp4 efficiently facilitated nuclear import of GST-SRY wt (**Fig. 3b**), nuclear import was lost for GST-SRY R62G and GST-SRY R75N but not affected by the R76P or R133W mutation (**Fig. 3e**). In summary, our data suggest that Exp4 operates as a nuclear import receptor for Sox family transcription factors and that the Exp4·Sox protein recognition requires residues in the HMG box domain that are conserved in all Sox proteins (**Fig. 3a**).

(e) Sex reversal mutations affected SRY nuclear import by Exp4 to different degrees. The experiment was performed as in (**b**), but with the indicated GST-SRY sex-reversal mutants. Only a subset of the SRY mutations compromised Exp4-mediated nuclear import of SRY. Bar, 25 µm.

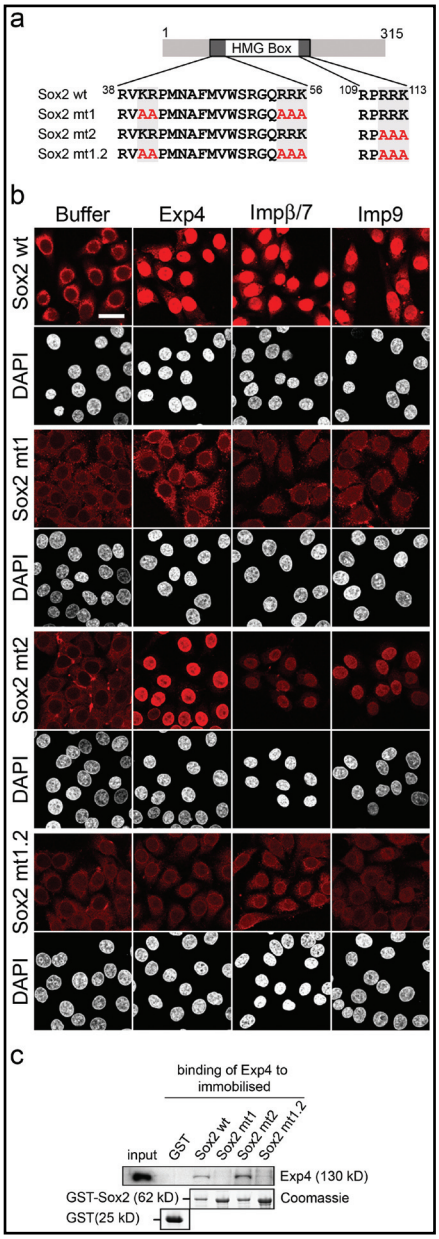


Fig. 4. The Exp4-, Impβ/7- and Imp9-dependent nuclear import signals overlap

(a) Schematic representation of N- and C-terminal regions of the HMG box that have been previously implicated in nuclear localisation of Sox2 (see **Fig. 3a**). The grey-shaded regions mark basic amino acid residues that were substituted for alanines (marked in red) in mutant 1 (mt1), mutant 2 (mt2) and mutant 1 + 2 (mt1.2) [13]. **(b)** Exp4-, Impβ/7- and Imp9-mediated Sox2 nuclear import is affected by the previously described Sox2 mutations [13]. The nuclear import assay was performed as in **Fig. 2c.** Note that the GST-Sox2 mutants showed reduced cytoplasmic aggregation in the absence of importins. DNA-staining with DAPI highlights nuclei. See text for further details. Bar, 25 μm **(c)** GST-Sox2 mutant 2, but not GST-Sox2 mutant 1, binds Exp4. The figure shows a GST-pulldown from ES cell extracts with GST-Sox2 or the indicated Sox2 mutants. Upper panel: Eluted Exp4 was detected by SDS-PAGE and Western blotting. Lower panels: Eluted GST and GST-Sox2 fusions were detected by SDS-PAGE and Coomassie-staining.

Nuclear import signals for Exp4, Impβ/7 and Imp9 overlap

Sox2 mutations within the HMG box domain have also been reported to cause localisation defects. GST-Sox2 mutants mt1, mt2, mt1.2 (**Fig. 4a**) were compromised in nuclear import when microinjected into the cytoplasm of HeLa cells [13]. We therefore tested whether these mutations affect Sox2-import by Exp4, Impβ/7 or Imp9. Strikingly, mutant 1 and mutant 1.2 completely failed to be imported by any of these pathways, while for mutant 2, only Impβ/7 and Imp9-mediated import was reduced (**Fig. 4b**). In line

with these observations, we detected binding of Exp4 only to wild-type Sox2 and mutant 2, but not to mutant 1 or mutant 1.2 (**Fig. 4c**). Therefore, the Exp4-dependent import signal involves parts previously shown to be important for Sox2 nuclear localisation and overlaps with the signal recognised by Impß/7 and Imp9. Given that some of the residues necessary for Sox protein import (for example, SRY R75) reside within one of the three alpha helices of the HMG-box domain [30], it appears that the import signature is not a linear sequence, but instead features a three-dimensional structure.

The above-described import determinants coincide conspicuously with DNA-binding residues (**Fig. 3a**; [30,43]). Thus, mutations of these residues, such as found in SRY-dependent sex-reversal conditions, have at least two consequences - they inhibit Sox DNA binding and thereby transcriptional activity directly and this effect is aggravated by the lower nuclear levels of the affected Sox-protein, due to inhibition of its import. Residues directly binding to DNA also appear to cause the observed cytoplasmic aggregation of Sox proteins, as the respective Sox2 mutations reduced precipitation in the absence of cognate importins (**Fig. 4b**). The mutations also compromised import (**Fig. 4b**) and we therefore propose that Exp4, Impß/7 and Imp9 act as chaperones for the exposed polycationic DNA-binding domain of Sox2 to prevent potentially deleterious aggregation with polyanions such as RNA [22]. Given this important function, it is not surprising that all cognate importins identified here recognise similar features of Sox2.

Sox2-dependent transcription is essential at different stages of embryonic development, including the maintenance of the inner cell mass, from which ES cells are derived [2] and neural progenitor expansion [5]. A recent study proposed that selective up-regulation of Impα3 and Impα5 would trigger the differentiation of ES cells into neural progenitors by mediating import of certain transcription factors, such as Sox2 [13]. Our finding that the import-activity of importin-αs towards Sox2 is negligible compared to Exp4, Imp9 or the Impß/7 dimer, argues against the scenario that up-regulated importin αs facilitate nuclear localisation of Sox2 during neural differentiation. Instead, our data are in agreement with the observation that Impα5-deficient mice undergo normal brain development [31]. It was suggested that up-regulated Impα4-expression may compensate for the loss of Impα5 [31], however, we found no supporting evidence for Sox2 import acitivity of any Impα isoform (**Fig. 2d**). Sox2 does not rely on a single nuclear import pathway, but can use three pathways all of whose import mediators are expressed in ES cells (**Fig. 1** and unpublished data). At least Impß/7 and Imp9 import essential mass proteins such as histones and ribosomal proteins [21, 22, 32]; they should thus represent constitutive import pathways that do not become limiting under physiological conditions. Whether other transcription factors required for neural differentiation can also access multiple, possibly constitutive import pathways remains to be tested. At least for Sox2, it appears very unlikely that import-regulation occurs by up- or down-regulating a specific nuclear import pathway. With a single known exception (Exp6, having only one substrate in *X. laevis* oocytes; [33]), regulation of transport is achieved by modifying the cargo, a strategy that does not affect myriads of other cargoes depending on the same pathway [34, 35]. If regulation of nucleocytoplasmic distribution plays a role in Sox-protein function, we would expect it to follow the just mentioned paradigm.

MATERIALS AND METHODS

Antibodies and cell culture

Anti-Exp4 antibodies (**Figs. 1c**, **3** and **4**) against residues 1015-1150 of the *Xenopus laevis* protein were raised in rabbits and affinity-purified via the antigen. For other figures, commercial anti-Exp4 was used (Santa Cruz, V-18). Other antibodies were against FLAG (M2; Sigma-Aldrich), Sox2 (Chemicon), Oct3/4 (Santa Cruz), Nestin and RC2 (DSHB).

Mouse 46C ES cells [36] and derivatives were grown under standard conditions [13]. FLAG-Sox2 46C ES cells were generated by electroporation of a 2xFLAG-tagged murine Sox2 under the control of the CAG promoter followed by neomycin selection. Differentiation into neural progenitors and immunostaining were as described [14]. HeLa cells were grown under standard conditions [11]. Lipofectamine 2000 (Invitrogen) was used for transient transfections .

Preparative FLAG-Sox2 purification and mass spectrometry

FLAG-Sox2 ES cell or neural progenitor nuclear extract (20 mM HEPES pH 7.6, 10% (w/v) glycerol, 100 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1x Complete protease inhibitor (Roche)) was incubated with anti-FLAG M2 agarose beads (Sigma-Aldrich). Bound material was eluted with a FLAG-tripeptide (Sigma-Aldrich) and analysed by SDS-PAGE and silver-staining. Mass spectrometrical analysis was by NanoflowLC-MS/MS on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap mass spectrometer (Thermo). Data analysis was as described [37].

Bacterial protein expression and purification

GST-fusions of human SRY (sub-cloned into a modified pGEX-6P-1 plasmid, GE Healthcare) and chicken Sox2 (both C-terminally His_6-tagged) were expressed in a protease-deficient *E. coli* strain and purified by nickel chelate affinity and gel filtration chromatography. The gel filtration step was essential to keep the background fluorescence in import assays sufficiently low to assess specific import effects. Alexa 488/568-maleimide-labelling was performed essentially as reported [38]. Other recombinant proteins were prepared as described [12, 22, 23, 38-42].

Immunoprecipitations and binding assays

ES cell or neural progenitor nuclear extract was incubated with Protein A Sepharose (Amersham) and anti-Sox2 antibody or control-rabbit-IgG (Santa Cruz), or with anti-FLAG M2-agarose (Sigma-Aldrich). Elution was with SDS sample buffer. Where indicated, 5 U/ml Benzonase (Novagen), 25 µg/ml ethidium bromide, 3-5 µM RanQ69L, 5 µM Rab6, 1 mM GTP, 12 µM RanGAP or 6 µM RanBP1 were added.

GST-pulldowns were performed by incubating GST-Sox2 bound to Glutathione Sepharose 4B (GE Healthcare) with an ES cell nuclear extract or MmExp4 (1 µM)-spiked *E. coli* lysate (50 mM Tris pH 7.5, 50 mM NaCl, 2 mM $Mg(OAc)_2$, 1 mM DTT, 0.005% digitonin) in the presence of an ATP/GTP-regenerating system and, if indicated, RanGTP (3 µM). Bound proteins were eluted with SDS sample buffer and analysed by immunoblotting and/or Coomassie-staining.

Import assays

Import assays were performed essentially as described [38]. If not noted differently, import mixes contained 1 µM of import cargoes and 1 µM of transport receptors (2 µM for Impαs). Where indicated, energy-dependence of nuclear import was tested by replacing the energy-regenerating system with 0.1 U/µl apyrase (Sigma-Aldrich). Fixed samples were mounted in Vectashield (Vector Laboratories); imaging was performed with an SP5 confocal laser scanning microscope system (Leica) using the 405-, 488- and 561-nm laser lines and a PlanApo oil objective (63x, NA 1.4, Leica). Confocal imaging was vital to minimize blurring of the strong cytoplasmic signal into nuclear regions. Figures were assembled in Photoshop or Illustrator (Adobe).

Further details are available on request.

Online supplemental material

Fig. S1 shows characteristics of the stable FLAG-Sox2 ES cells (**a**) and neural progenitor cells (**b**). Fig. S1c shows that the Sox2·Exp4 interaction is insensitive to RanGTPase activation. Fig. S2 shows that multiple rounds of nuclear import by Exp4 depend on Ran and energy (**a**) and that Exp4 requires RanGTP only to release its import cargo into the nucleus, but not for detachment from the nuclear side of the NPC (**b**). **Fig. S3** depicts that Exp4, Impβ/7 and Imp9 mediate nuclear import of Sox2 also in permeabilised ES cells.

ACKNOWLEDGEMENTS

We thank Dr. Vincent Harley (FLAG-SRY plasmid), Dr. Yoshihiro Yoneda (GST-Sox2 constructs), Dr. Matthias Köhler and Dr. Enno Hartmann (human Importin α plasmids), Dr. Amy Tang (pCBA vector), Dr. A. Akhmanova (Rab6GTP), Dr. Austin Smith (46C ES cells), Dr. Laura Riolobos for advice on import assays, Dr. Steffen Frey for helpful comments on the manuscript and Danilo Deichsel and Heinz-Jürgen Dehne for excellent technical assistance. The Rottier lab is supported by the Sophia Foundation for Medical Research (SSWO nr. 413), the Poot lab is supported by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO). Work in the Görlich lab is funded by the Max-Planck-Gesellschaft, T.G. was also supported by fellowships of the Boehringer Ingelheim Fonds and the Alfried Krupp Foundation.

REFERENCES

1. Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y. & Pallavi, B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol 39, 2195-2214 (2007).
2. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17, 126-140 (2003).
3. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol 9, 625-635 (2007).
4. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-676 (2006).

5. Ferri, A.L. et al. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development 131, 3805-3819 (2004).
6. Graham, V., Khudyakov, J., Ellis, P. & Pevny, L. SOX2 functions to maintain neural progenitor identity. Neuron 39, 749-765 (2003).
7. Taranova, O.V. et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev 20, 1187-1202 (2006).
8. Kiernan, A.E. et al. Sox2 is required for sensory organ development in the mammalian inner ear. Nature 434, 1031-1035 (2005).
9. Que, J. et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 134, 2521-2531 (2007).
10. Williamson, K.A. et al. Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. Hum Mol Genet 15, 1413-1422 (2006).
11. Kurisaki, A. et al. The mechanism of nuclear export of Smad3 involves exportin 4 and Ran. Mol Cell Biol 26, 1318-1332 (2006).
12. Lipowsky, G. et al. Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. EMBO J 19, 4362-4371 (2000).
13. Yasuhara, N. et al. Triggering neural differentiation of ES cells by subtype switching of importin-alpha. Nat Cell Biol 9, 72-79 (2007).
14. Conti, L. et al. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol 3, e283 (2005).
15. Mattaj, I.W. & Englmeier, L. Nucleocytoplasmic transport: the soluble phase. Annu Rev Biochem 67, 265-306 (1998).
16. Gorlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol 15, 607-660 (1999).
17. Pemberton, L.F. & Paschal, B.M. Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic (Copenhagen, Denmark) 6, 187-198 (2005).
18. Bischoff, F.R. & Gorlich, D. RanBP1 is crucial for the release of RanGTP from importin beta-related nuclear transport factors. FEBS Lett 419, 249-254 (1997).
19. Adam, S.A., Marr, R.S. & Gerace, L. Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors. J Cell Biol 111, 807-816 (1990).
20. Newmeyer, D.D., Finlay, D.R. & Forbes, D.J. In vitro transport of a fluorescent nuclear protein and exclusion of non-nuclear proteins. J Cell Biol 103, 2091-2102 (1986).
21. Jakel, S. et al. The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. EMBO J 18, 2411-2423 (1999).
22. Jakel, S., Mingot, J.M., Schwarzmaier, P., Hartmann, E. & Gorlich, D. Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. EMBO J 21, 377-386 (2002).
23. Mingot, J.M., Kostka, S., Kraft, R., Hartmann, E. & Gorlich, D. Importin 13: a novel mediator of nuclear import and export. EMBO J 20, 3685-3694 (2001).
24. Yoshida, K. & Blobel, G. The karyopherin Kap142p/Msn5p mediates nuclear import and nuclear export of different cargo proteins. J Cell Biol 152, 729-740 (2001).
25. Ribbeck, K., Kutay, U., Paraskeva, E. & Gorlich, D. The translocation of transportin-cargo complexes through nuclear pores is independent of both Ran and energy. Curr Biol 9, 47-50 (1999).
26. Poulat, F. et al. Nuclear localization of the testis determining gene product SRY. J Cell Biol 128, 737-748 (1995).
27. Sudbeck, P. & Scherer, G. Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. J Biol Chem 272, 27848-27852 (1997).
28. Sinclair, A.H. et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240-244 (1990).
29. Harley, V.R. et al. Defective importin beta recognition and nuclear import of the sex-determining factor SRY are associated with XY sex-reversing mutations. Proc Natl Acad Sci U S A 100, 7045-7050 (2003).
30. Werner, M.H., Huth, J.R., Gronenborn, A.M. & Clore, G.M. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell 81, 705-714 (1995).
31. Shmidt, T. et al. Normal brain development in importin-alpha5 deficient-mice. Nat Cell Biol 9, 1337-1338; author reply 1339 (2007).
32. Muhlhausser, P., Muller, E.C., Otto, A. & Kutay, U. Multiple pathways contribute to nuclear import of core histones. EMBO Rep 2, 690-696 (2001).

33. Bohnsack, M.T., Stuven, T., Kuhn, C., Cordes, V.C. & Gorlich, D. A selective block of nuclear actin export stabilizes the giant nuclei of Xenopus oocytes. Nat Cell Biol 8, 257-263 (2006).
34. Kaffman, A., Rank, N.M., O'Neill, E.M., Huang, L.S. & O'Shea, E.K. The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus. Nature 396, 482-486 (1998).
35. Kaffman, A., Rank, N.M. & O'Shea, E.K. Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. Genes Dev 12, 2673-2683 (1998).
36. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol 21, 183-186 (2003).
37. Sanchez, C. et al. Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics 6, 820-834 (2007).
38. Jakel, S. & Gorlich, D. Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. EMBO J 17, 4491-4502 (1998).
39. Dean, K.A., von Ahsen, O., Gorlich, D. & Fried, H.M. Signal recognition particle protein 19 is imported into the nucleus by importin 8 (RanBP8) and transportin. J Cell Sci 114, 3479-3485 (2001).
40. Kohler, M. et al. Evidence for distinct substrate specificities of importin alpha family members in nuclear protein import. Mol Cell Biol 19, 7782-7791 (1999).
41. Pollard, V.W. et al. A novel receptor-mediated nuclear protein import pathway. Cell 86, 985-994 (1996).
42. Ribbeck, K. & Gorlich, D. The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. EMBO J 21, 2664-2671 (2002).
43. Murphy, E.C., Zhurkin, V.B., Louis, J.M., Cornilescu, G. & Clore, G.M. Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. J Mol Biol 312, 481-499 (2001).
44. Lupu, F., Alves, A., Anderson, K., Doye, V. & Lacy, E. Nuclear pore composition regulates neural stem/progenitor cell differentiation in the mouse embryo. Dev Cell 14, 831-842 (2008).

SUPPLEMENTARY FIGURES

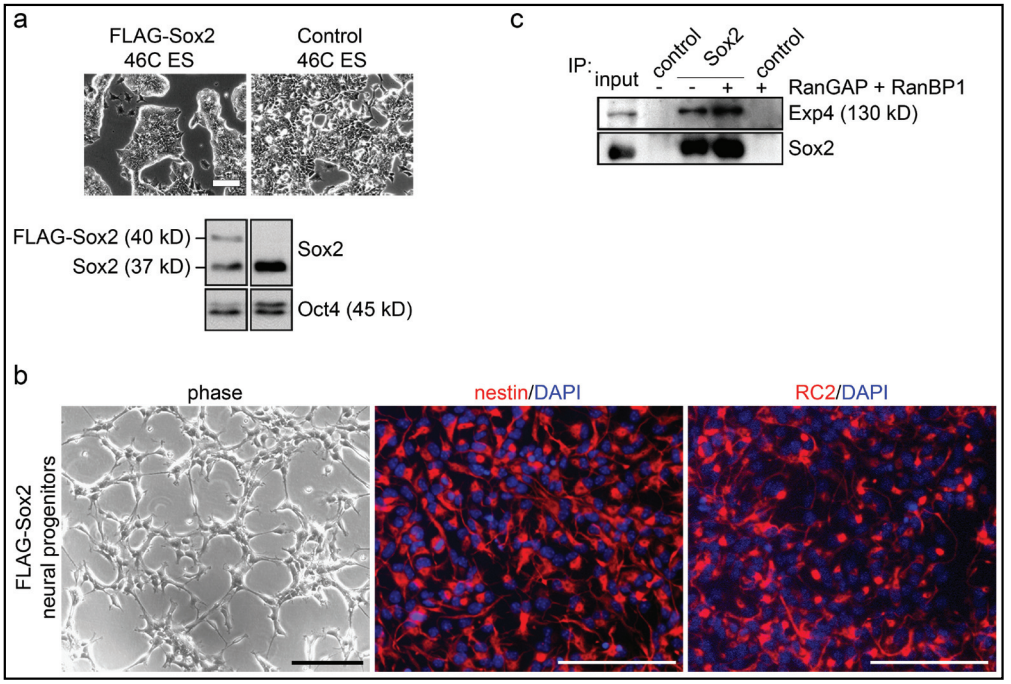
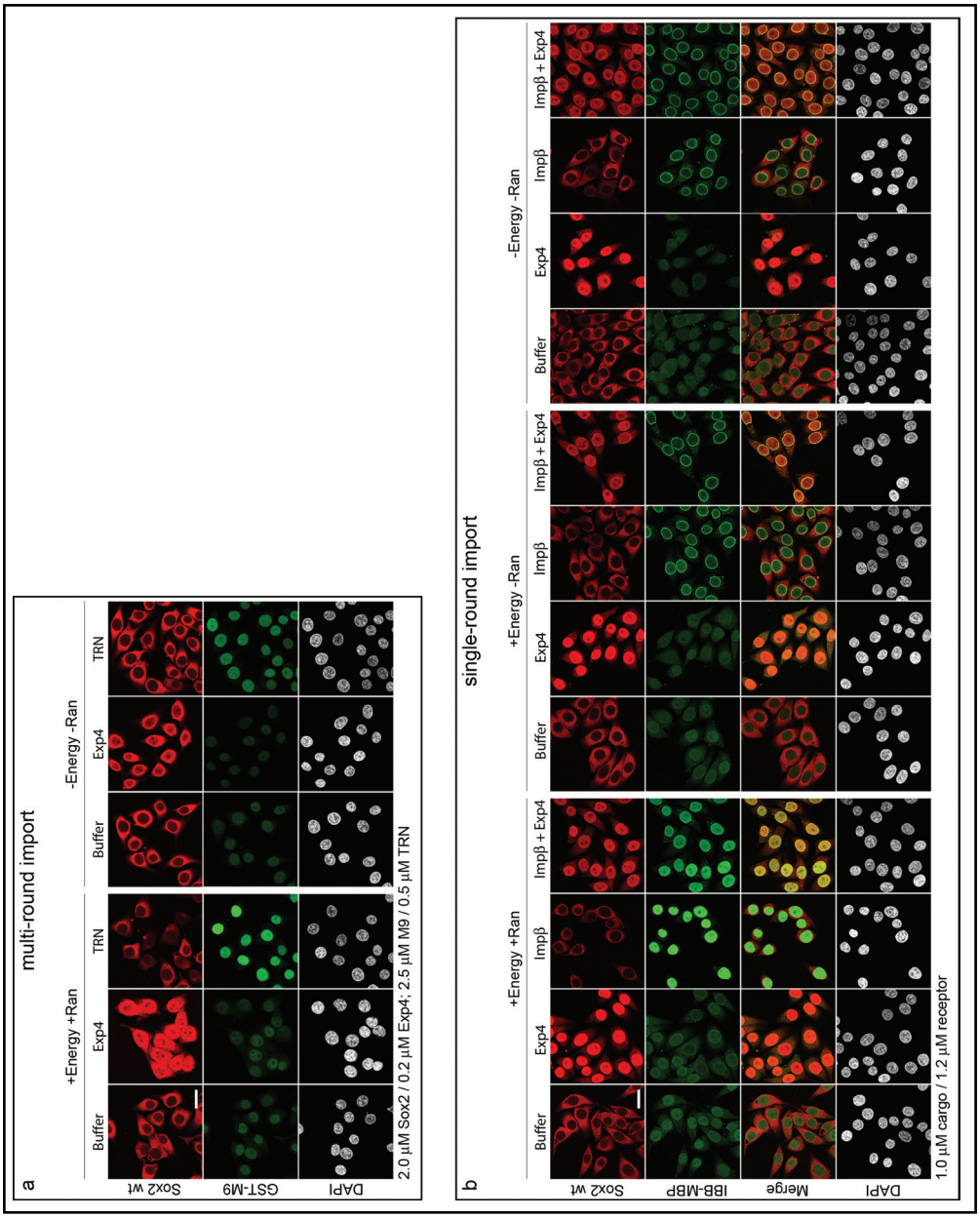


Fig. S1. (a) FLAG-Sox2 cells display normal ES cell growth behaviour and express the ES cell marker Oct4. Upper panels: Phase contrast images of control-46C ES cells and 46C ES cells stably transfected with FLAG-SOX2. Bar, 200 µm. Lower panels: Anti-Sox2 and Oct4 Western blots using extracts from FLAG-Sox2 ES cells (left) and control 46C ES cells (right). **(b)** Characterisation of ES-cell derived FLAG-Sox2 neural progenitors. The left panel shows a phase contrast image of FLAG-Sox2 neural progenitors. Immunostaining shows that the cells expressed the neural progenitor markers Nestin (middle panel) and RC2 (right panel). Bars, 100 µm. **(c)** The Sox2·Exp4 complex is insensitive to RanGTPase activation by RanGAP and RanBP1. Co-immunoprecipitation of Exp4 with endogenous Sox2 as in **Fig. 1b**, but with the indicated additions.

Fig. S2. (a) Efficient multi-round nuclear import by Exp4 depends on Ran and metabolic energy. The import experiment was performed as in Fig. 2a, but import substrates were in 10-fold (Sox2) or 5-fold (M9) excess over the receptors, making efficient nuclear accumulation of cargoes dependent on receptor recycling. In the absence of Ran and energy, nuclear accumulation of both GST-M9 and GST-Sox2 was impaired. **(b)** Single-round nuclear import of Exp4·Sox2 is independent of Ran and metabolic energy. The import experiment was performed as in **Fig. 2b**, but transport receptors were in slight excess over the substrates to render nuclear accumulation of the cargoes independent of receptor recycling. Omission of Ran or of Ran and energy did not affect GST-Sox2 import by Exp4, but did prevent the nuclear accumulation of Impß·IBB-MBP complexes (which instead arrested at nuclear pores). This was also evident when both Impß and Exp4 had been added simultaneously (right panels). Bars, 25 µm.


a
multi-round import
+Energy +Ran
-Energy -Ran
Buffer
Exp4
TRN
Buffer
Exp4
TRN
Sox2 wt
GST-M9
DAPI
2.0 µM Sox2 / 0.2 µM Exp4; 2.5 µM M9 / 0.5 µM TRN
b
single-round import
+Energy +Ran
+Energy -Ran
-Energy -Ran
Buffer
Exp4
Impβ
Impβ + Exp4
Buffer
Exp4
Impβ
Impβ + Exp4
Buffer
Exp4
Impβ
Impβ + Exp4
Sox2 wt
IBB-MBP
Merge
DAPI
1.0 µM cargo / 1.2 µM receptor

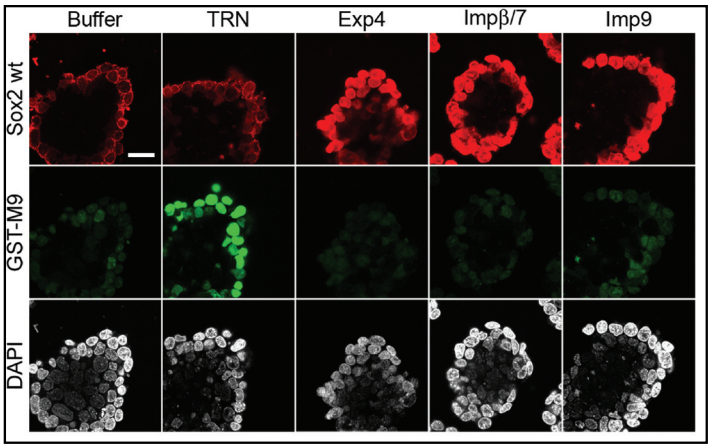


Fig. S3. Exp4, Impß/7 and Imp9 mediate nuclear import of Sox2 also in permeabilised embryonic stem cells. NPC composition may vary among different cell types and change during development [44]. As it cannot be excluded that such changes regulate nuclear import, we tested the import-activity of Exp4, Impß/7 and Imp9 with permeabilised 46C ES cells instead of HeLa cells. The import experiment was performed as in **Fig. 2a**. The transportin (TRN)-substrate GST-M9 had been used as an internal specificity control. Bar, 25 µm.

Chapter

5

The Sox2 protein interaction network in embryonic stem cells

Work in progress

The Sox2 protein interaction network in embryonic stem cells

ABSTRACT

Embryonic stem (ES) cells have unlimited proliferation potential while maintaining their pluripotent capacity. Sox2 is one of the key transcription factors in the molecular network that regulates the establishment and maintenance of pluripotency and self-renewal in ES cells. Sox2 also aids in the reprogramming of somatic cells into induced pluripotent-like stem (iPS) cells. Sox2 was shown to co-occupy many promoters with other factors associated with pluripotency, including Oct4 and Nanog. It is likely that Sox2 regulates its target genes with the help of interaction partners, but besides Oct4, no other Sox2-interacting proteins are known in ES cells. Here, we purified FLAG-Sox2 from ES cells stably expressing a FLAG-Sox2 transgene using an optimized FLAG-affinity purification protocol and identified many specific Sox2 interacting proteins by mass spectrometry. Putative Sox2 partners include ES cell self-renewal factors, chromatin remodeling complexes, several lineage-specific transcriptions factors and the Sox protein import receptor, Exportin 4. Our data suggest that protein-protein interactions are important for the maintenance of ES cell transcription networks.

INTRODUCTION

Embryonic stem cells (ES) cells are derived from the inner cell mass (ICM) of the blastocyst and have two defining properties, self-renewal and pluripotency [1, 2]. They can be maintained *in vitro* without differentiation for prolonged periods of time. In addition, ES cells have the ability to differentiate into different cells representing the three embryonic germ layers. Under specific *in vitro* conditions ES cells can also differentiate into trophectoderm cells, which in vivo form the placenta [3].

Over the past years numerous studies have focused on understanding the transcription factor network that maintains the pluripotency and self-renewal of ES cells. It was shown that this network is dominated by three transcription factors, Sox2 and the non-classical homeobox proteins Oct4 and Nanog [4, 5]. Sox2 belongs to the Sox protein family, which has over 20 members that all contain a typical HMG-box DNA binding domain [6]. Sox2 is expressed in ES cells, the ICM and in germ cells, but also later in development [7]. Sox2 plays an essential role during early development, since deletion of Sox2 causes early embryonic lethality, due to the failure of the ICM to develop [7]. Depletion of Sox2 in ES cells results in differentiation into mostly trophectoderm-like cells, similar to the effect of Oct4 depletion [4]. In line with this, it was suggested that the main function of Sox2 is to indirectly maintain the expression of Oct4 [4]. This was supported by the observation that ES cells can remain undifferentiated in the absence of Sox2 if Oct4 is ectopically expressed from a constitutive promoter [4]. Forced expression of Sox2 did not affect ES cell self-renewal, but upon loss of self-renewal conditions, drives ES cell differentiation towards the neuroectodermal lineage [8]. The stem cell inducing properties of Sox2 were further demonstrated by the reprogramming of both human and mouse somatic cells into induced pluripotent-like stem (iPS) cells by ectopic expression of Sox2 and Oct4 together with Klf4 and c-myc [9, 10].

Recently, a number of studies used different genome-wide chromatin immunoprecipitation [11] analyses to identify putative target genes of Sox2, as well as other transcription factors important for ES cell biology, in human [12] and mouse [13-15] ES cells. These different data sets suggest that Sox2 co-occupies many of its target promoters together with the other transcription factors, although it can not be excluded that different factors occupy the targets in different cells, as the ChIP analysis is performed on a population of cells that might not be homogeneous. It was suggested that Sox2 acts together with other transcription factors to maintain ES cell identity, by activating genes associated with ES cell self-renewal and pluripotency and repressing genes associated with lineage specification and differentiation [12-14].

The role of protein-protein interactions in maintaining the ES cell transcriptional network is poorly understood. Despite its critical role in ES cells, knowledge of Sox2 interacting proteins, besides Oct4 [4], is completely lacking. Here, we used two independent ES cell lines that express FLAG-tagged Sox2 from a transgene to affinity-purify FLAG-Sox2 and its interacting proteins from nuclear extract. Putative Sox2 binding proteins were subsequently identified by mass spectrometry. We found that Sox2 interacts with several ES cell self-renewal factors, including Oct4 and Nanog. Interestingly, many of the identified transcription factors were reported to co-occupy promoters with Sox2, which suggests that protein interactions are highly relevant for Sox2 transcriptional activity. In addition, our data indicate that Sox2 associates with several chromatin remodelling complexes and a number of lineage-specific transcriptions factors and efficiently binds to its nuclear import factor, Exportin 4 (described in **chapter 4** of this thesis).

RESULTS AND DISCUSSION

FLAG-Sox2 affinity-purification

To identify interacting proteins of Sox2, we introduced an amino-terminal tagged FLAG-Sox2 transgene into 46C ES cells and isolated individual neomycin-resistant clones that stably expressed FLAG-Sox2 (**Figure 1A**). Two clones, hereafter called F-Sox2.25 and F-Sox2.38 ES cells, were selected, which had a normal ES cell morphology, expressed the ES cell marker Oct4 and expressed FLAG-Sox2 at a level of approximately 30% of the endogenous Sox2 (**Figure 1B**). The moderate level of transgene expression was selected to avoid the disturbance of the stoichiometry of protein complexes and the creation of artefacts or non-physiological interactions [16]. FLAG-Sox2 ES cells were expanded and nuclear extracts prepared. Flag#3 purification was done from F-Sox2.38 cells using a normal FLAG-affinity protocol, whereas the purifications Flag#1 and Flag#2, from F-Sox2.25 ES cells and F-Sox2.38 respectively, were performed with an optimised protocol (**Figure 1C**, see material and methods for details). Silverstain analysis of a polyacrylamide gel containing the proteins purified from F-Sox2 cells identified a band running above the 37 kD marker, that was not present in purifications from control extracts and was recognised by a FLAG antibody, suggesting it represents the FLAG-Sox2 protein (data not shown). Mass spectrometry analysis of FLAG-Sox2 samples and

control samples confirmed the presence of the Sox2 protein in all three independent FLAG purifications. Putative Sox2-associated proteins identified by mass spectrometry are summarized in **Table 1A** and **1B**, following the criteria described in the material and methods. In this study, we focused on the role of Sox2 associated transcription factors, transcription related factors and chromatin remodelling complexes.

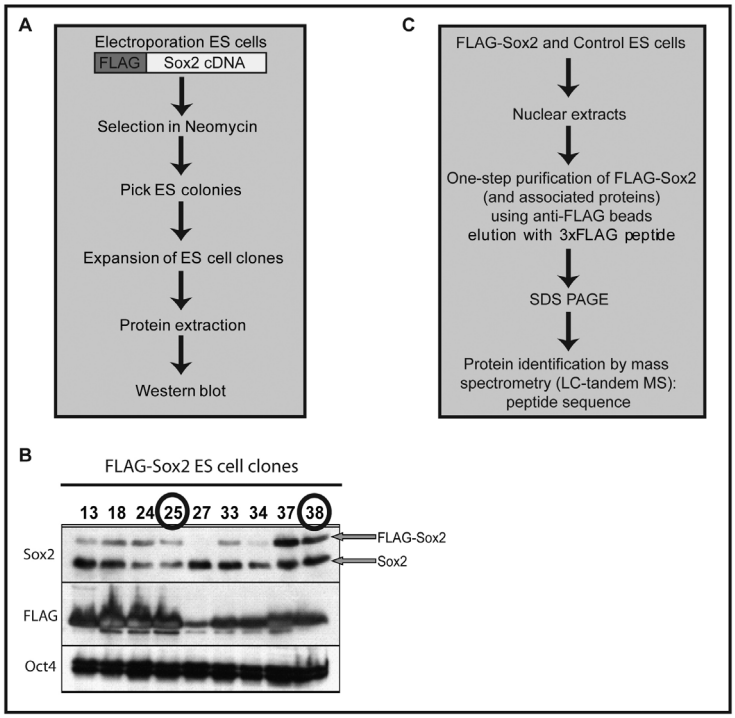


Figure 1. (A) Scheme for the generation of the stable FLAG-Sox2 ES cell clones. **(B)** Sox2, FLAG and Oct4 Western blots on extracts from FLAG-Sox2 ES cell clones. The selected ES cell clones, F-Sox2.25 and F-Sox2.38 ES cells, expressed FLAG-Sox2 at a level of approximately 30% of endogenous Sox2. FLAG-Sox2 ES cells express the ES cell marker Oct4. **(C)** Scheme for the FLAG-affinity purification from FLAG-Sox2 and control ES cells, the putative Sox2 interacting proteins were identified by mass spectrometry.

General remarks on identified Sox2-interacting proteins

We identified 78 putative Sox2 interacting proteins (**Table 1A**, **1B**), which suggests that Sox2 is a component of several protein complexes in ES cells. The list of putative Sox2 interacting proteins includes 27 proteins with a role in early embryonic development of ES cell self-renewal, as judged by their mouse knock-out phenotypes, or knock-down studies in ES cells (**Table 2**). There is no such information available for 27 proteins and 24 proteins have known mouse knock-out phenotypes, or knock-down phenotypes in ES cells that are not consistent with an early need in embryonic development. Thus, more

than 50% of the putative Sox2 interacting proteins with a known phenotype are essential for early development and/or ES cell properties. This enrichment suggests that Sox2 interacting proteins may be involved in the functioning of Sox2 as an essential ES cell factor.

We found that more than half of the genes that encode the proteins in the list of putative Sox2 interacting proteins have been reported to be bound by Sox2 in recent genome-wide chromatin immunoprecipitation studies in mouse ES cells [11,13,17] (**Table 2**). This may suggest a feedback mechanism whereby a Sox2-interacting protein may regulate its own gene as part of a complex with Sox2. From the list, there are 16 proteins of which the genes are bound by Sox2 and their expression is affected after Sox2 depletion (**Table 2**). Of these, 70% are down-regulated and around 30% are up-regulated.

ES cell self-renewal factors

We identified twenty transcription factors as putative Sox2 interacting partners (**Table 1**). Nine of these transcription factors have already been associated with a function in ES cell self renewal in previous studies (Oct4 [18], Nanog [19,20], Tbx3 [21], Esrrb [21], Sall4 [16], Nac1 [16], Dax1 [16,21], Zfp143 [22] and Klf5 [21]).

Sox2, together with Oct4 and Nanog form the central regulatory core of transcription factors maintaining pluripotency in ES cells [3]. The synergistic binding of Sox2 and Oct4 to promoters is well established and a model for their structure on DNA has been suggested [23]. A number of genes associated with pluripotency are regulated by Sox-Oct binding elements in their promoters [4]. It has been shown that Sox2 and Nanog co-occupy many promoters [13,14] but so far there are no data that show that they are part of the same protein complex. Interestingly, the predicted Nanog-bound sequences resembles the sox-oct enhancer [13], and the number of promoters bound by Sox2 and Nanog is even higher than that of Sox2 and Oct4 [14]. Our identified interaction between Nanog and Sox2 may play a role in their frequent co-localisation on promoters. Seven of the transcription factors identified as putative Sox2 interacting partners (Oct4, Esrrb, Sall1, Sall4, Nac1, Sp1 and Dax1) form part of the described Nanog interaction network in ES cells [16]. This suggests that Sox2 and Nanog have similar interaction partners, but it is unclear whether Sox2 and Nanog are together in (all) these complexes.

The genome-wide localisation of Sox2 and eight Sox2 interactors, Nac1, Oct4, Esrrb, Tcfcp2l1, Nanog, Dax1, Klf5 and Ring1B [13,14,24] (**Table 2**), was determined by chromatin immunoprecipitation and the results showed a high degree of overlapping genes bound by Sox2 and these factors. This suggests that protein-protein interactions between Sox2 and its interacting proteins may be important for linking the Sox2 transcriptional network to that of its putative interaction partners.

Lineage-specific transcriptions factors

Surprisingly, we found two transcription factors characteristic for lineage-specific differentiation, Gata4 and Runx1, as putative partners of Sox2. As Sox2 is preventing ES cell differentiation, these interactions were unexpected. The interaction between Sox2

and Gata4 may play a role in preventing ES cell differentiation towards the endoderm lineage. Gata4 is implicated in primitive endoderm specification and it has been shown that forced expression of Gata4 in ES cells is sufficient to induce their differentiation into extraembryonic endoderm [25]. In the ICM, Gata4 has a mosaic expression pattern, similar to Gata6, which is correlated with primitive endoderm fate determination [26]. Gata4 is not required for maintaining ES cell status, since Gata4-null ES cells proliferate normally, but they are defective in endoderm differentiation [27]. Interestingly, Sox2 binds the Gata4 promoter and upon Sox2 depletion, the *Gata4* gene is up-regulated (**Table 2**). This may suggest that Sox2, maybe in complex with Gata4, prevents endoderm differentiation by repressing *Gata4* expression to a level below the threshold necessary to initiate differentiation. Interestingly, the expression of *Fgf-3*, another gene involved in endoderm differentiation, is positively regulated by the binding of Sox7 and Gata4 to the same element in the promoter. When Sox2 was co-expressed with Gata4, it acted as a negative modulator of the strong activation by Gata4 [28]. A Sox2-Gata4 complex may therefore act as a repressor to prevent activation of endoderm specification genes.

We also identified the heterodimeric Runx1- CBFß transcription factor complex as a putative interactor of Sox2. Runx1- CBFß is required for the establishment of the definitive hematopoietic system during embryonic development [29] and it has been shown that this complex regulates a number of target genes essential for definitive hematopoiesis [30]. In line with this, homozygous disruptions of Runx1 or CBFß resulted in embryonic lethality due to the complete block in fetal liver hematopoiesis [31]. Runx1 is expressed at low levels in undifferentiated ES cells and Runx1$^{-/-}$ ES cells presented a deficient hematopoietic differentiation [31]. The *Runx1* gene is not a direct target of Sox2, but upon Sox2 depletion *Runx1* is upregulated (**Table 2**), which suggests that Sox2 might be indirectly repressing *Runx1*. The interaction between Sox2 and Runx1-CBFß may play a role in this regulation.

Transcription related factors

We found four components of the basal RNA polymerase II transcription machinery (Cdk9, Polr2e Gtf2f2 and Taf9) as putative interaction partners of Sox2. It is unclear how the transcription regulatory function of pluripotency factors such as Oct4, Nanog and Sox2 connects to the basal transcription machinery, i.e. the RNA polymerase II complex (PolII). The interaction of Sox2 with PolII components may suggest that Sox2, which has a known role as architectural and DNA bending factor, connects the Sox2 bound promoters or enhancers to the gene-bound PolII complex to regulate its activity. The putative interaction between Sox2 and Cdk9 is especially interesting. Cdk9 phosphorylates the C-terminal domain of PolII to allow for transcription-elongation [32]. In ES cells many active and inactive promoters are occupied by PolII and transcription appears to be regulated at the level of transcription-elongation [33]. An interaction between Sox2 and Cdk9 may suggest that Sox2 plays a regulatory role in this process.

The presence of Brca2 and Rad51 on the list of putative interacting proteins of Sox2 is very interesting. ES cells have a high proliferation rate and at the same time must preserve

their genomic integrity, since they are essential for embryo formation. Therefore, ES cells will undergo differentiation or apoptosis more rapidly than other cells to eliminate damage cells from the population [34]. It has been shown that ES cells are more sensitive to DNA damage than somatic cells, and they also seem to have a lower mutation frequency and a different DNA repair capacity than somatic cells [34]. Double strand breaks are the most dangerous type of DNA lesions, and ES cells repair those lesions faster than somatic cells using mainly the homologous recombination pathway. Rad51 is a key protein in this pathway and its DNA binding activity and cellular localization is mediated by Brca2 [35]. Considering the importance of homologous recombination in ES cells, it may not be surprising that null Rad51 or Brca2 ES cells are not viable (**Table 2**). The interaction with these proteins may suggest a role for Sox2 in maintaining the genomic stability in ES cells.

It has been reported that the ES cell transcription factor Klf5 interacts with the histone chaperone Anp32b, which leads to the transcriptional repression of a Klf5-downstream gene [36]. The interaction of Sox2 with both proteins may suggest that Sox2 plays a role in this process.

Chromatin remodelling complexes

A number of chromatin remodelling complexes were prominently present in our list of putative Sox2 interacting proteins such as the NuRD (nucleosome remodelling and histone deacetylation) complex, the SWI-SNF complex, the polycomb repressor complex 1 (PRC1) and the Lsd1 histone demethylase complex. Almost all subunits of the transcriptional repressor complex NuRD were present, most with high scores. Subunits of the NuRD complex were shown to bind and repress target genes of Oct4 [37]. As Oct4 is a partner of Sox2 in gene regulation [4], this may suggest that NuRD facilitates Sox2 to repress some of its targets. SWI-SNF is suggested to be both a co-activator and repressor in ES cells [38]. Recently, the DNA binding sites of Brg1, the catalytic subunit of SWI-SNF, were determined in ES cells [38]. Binding sites of Sox2 and Brg1 show a significant overlap, suggesting that the SWI-SNF complex may help Sox2 in regulating its targets. PRC1 was shown to be targeted by Oct4 to repress a subset of the Oct4 targets [39]. Analogous to this, PRC1 may also help Oct4 partner Sox2 in repressing some of its targets. Sox2 binding sites are enriched for the histone H3K27 tri-methylation mark, which is associated with polycomb silencing [14]. Together this suggests that the identified interaction between Sox2 and PRC1 may be functionally relevant for gene regulation by Sox2. The Lsd1 histone demethylase complex represses transcription by removing activating histone methylation marks [40]. The relevance of the interaction between Sox2 and Lsd1 complex is unclear as Lsd1 has not been studied yet in ES cells.

Concluding remarks

We have shown that Sox2 can interact with many putative partner proteins and partner complexes (**Figure 2**).

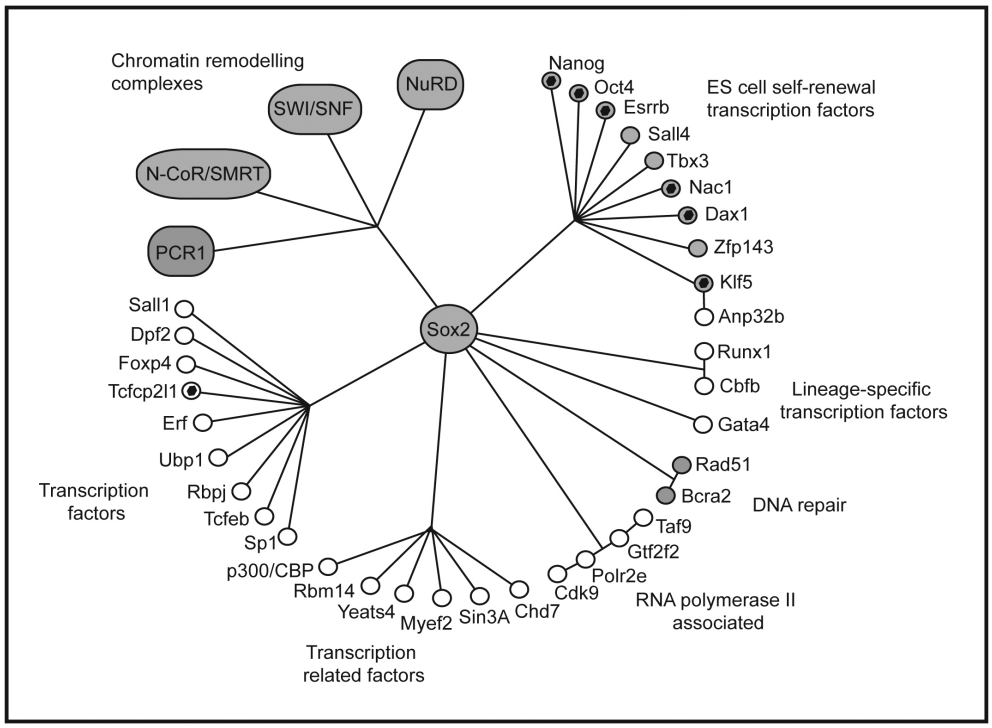


Figure 2. Sox2 protein interaction network in mouse ES cells, including Sox2 associated transcription factors, transcription related factors and chromatin remodelling complexes. Grey circles: proteins with a role in early embryonic development of ES cell self-renewal. White circles: proteins with unknown function in ES cells or proteins with known phenotypes that are not consistent with an early need in embryonic development. Black dots: proteins that co-localize with Sox2 on many promoters.

It is likely that these interactors help Sox2 in fulfilling its important role as a transcriptional regulator in maintaining ES cell identity and we have given a number of examples of reports that suggest that our interactions may be functional. However, a number of questions remain. It is unlikely that all Sox2 interacting proteins reside in one big complex. Future studies will have to dissect the composition of the different complexes that Sox2 forms by purifying the Sox2 interacting proteins. This leads to a second question. If Sox2 forms different complexes, how is the recruitment of the different Sox2 interacting proteins to different Sox2 target genes regulated? As Sox2 binds both complexes with an activating role and a repressing role in transcription, recruiting the correct complex is important. There are a number of possible ways to regulate interaction and promoter recruitment. Modifications of the Sox2 protein could allow it to selectively interact with some complexes and not with others and thereby could regulate a role for Sox2 as a repressor or an activator. Sox2 interacts with a number of transcription factors (**Tables 1** and **2**). For these proteins, the interaction with Sox2 on a promoter may also be stabilised by binding to its consensus site. An example of this is the Sox2-Oct4 partnership

where a weak interaction in solution is stabilised by binding to oct-sox elements in some promoters [41]. Another example is Esrrb, an interacting protein of both Oct4 [42] and Sox2 (**Table 1**), which is recruited to the Nanog promoter by both its interaction with Oct4-Sox2 and its interaction with a DNA binding site. As Sox2 is co-localising on promoters with many of its interacting transcription factors, most of which have consensus DNA binding sites[13], this mechanism of recruitment may be a more general principle. Our data also show that many ES cell transcription factors that are together on ES cell promoters are interacting with each other, via Sox2 or otherwise (**Table 2**). This correlation suggests that protein-protein interactions play an important role in maintaining transcription networks and their study is useful, both for understanding better the known network and discovering new links. In this respect it would be interesting to study the role in ES cell self-renewal of the many uncharacterised proteins that are in our list of putative Sox2 partners and expand the ES cell transcription network into new directions.

MATERIALS AND METHODS

Cell culture

Mouse 46C ES cells [43] and two derivative clones stably expressing FLAG-Sox2 (F-Sox2.25 and F-Sox2.38 ES cells, this thesis, **chapter 4**) were grown on gelatin-coated dishes without feeders, in BRL-conditioned medium supplemented with leukemia inhibitory factor, 20% FBS, non-essential amino acids, 50 µM beta-mercaptoethanol and penicillin/streptomycin.

Preparative FLAG-Sox2 purification and mass spectrometry

Nuclear extracts from FLAG-Sox2 ES cells and control 46C cells were prepared as described [44] and dialysed to buffer C-100 (20 mM Hepes pH 7.6, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol). Three independent purifications were performed. For purification Flag#1 F-Sox2.25 ES cells were used and Flag#2 and #3 used F-Sox2.38 ES cells. For Flag#3, FLAG-Sox2 was purified by rotating 10 ml of nuclear extracts from FLAG-Sox2 ES cells with 100 µl of anti-FLAG M2 agarose beads (Sigma-Aldrich), equilibrated in buffer C-100 (20 mM Hepes pH 7.6, 10% glycerol, 100 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1x *Complete EDTA-free* protease inhibitor, Roche) for 3 hrs at 4ºC in the presence of 1500 U of Benzonase (Novagen). Beads were than transferred into an Eppendorf tube and washed five times with 1 ml of C-100 buffer + 0.02% NP40 (C-100*). FLAG-Sox2 was eluted in four steps from the beads by addition of 100 µl elution buffer (C-100 + 0.2 mg/ml FLAG-tripeptide, Sigma-Aldrich) for 15 minutes at 4ºC. For Flag#1 and Flag#2 purifications, an optimised protocol was used. Here 60 µl of anti-FLAG M2 agarose beads was added to 1.5 ml of the nuclear extract in *no stick* microcentrifuge tubes (Alpha Laboratories), which give a higher protein recovery yield and a lower background. Fractions were analysed by SDS-PAGE (10% PAA) and silver-staining. Eluted fractions were concentrated by TCA precipitation and separated by SDS-PAGE (10% PAA) and

stained by colloidal Coomassie-staining. Gel lanes were cut, processed and analysed by NanoflowLC-MS/MS on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap mass spectrometer (Thermo) and data analysis was as described [45].

Sox2 protein interaction network and references

Criteria for exclusion/inclusion of proteins as FLAG-Sox2-interactors in **Table 1A** are: Mascot score of the identified protein needed to be higher than 45 in at least two of the three experiments, and three times higher than that in the corresponding control in Flag#1 and Flag#2 experiments, performed with the improved protocol, which gives a lower background. Proteins with reported membrane, cytoplasmic, or mitochondrial localization were removed. For the Gene Ontology analysis the PANTHER classification system was used. **Table 1B** contains proteins that follow the above criteria for only one purification, but are present in at least one other purification, albeit with a higher background.

The criteria used in **Table 2** are the following: correlation between transcription factor co-occupancy at promoters [13, 17] was scored as positive when > 0.2. Promoter occupancy by Klf5 was assigned as overlapping with Klf4 [17, 24]. Genes bound by Sox2 were assigned according to the detection of Sox2 at their promoter by bioChIP-Chip [17] or ChIP-seq data showing an association score > 0.1 [13]. Microarray data on genes regulated by Sox2 were derived from Sharov et al. [46]. Genes were scored as regulated by Sox2 if they showed at least 1.5-fold up or down regulation in mRNA levels within 96 hrs after shutdown of Sox2 transcription by addition of tetracycline to *Sox2*-inducible null cell line, 2TS22C ES cells [4].

References for subunit composition of Sox2-binding complex (**Tables 1** and **2**) are NuRD complex [47], SWI/SNF complex [48], N-CoR/SMRT complex [49], PRC1 complex [50] and LSD1 complex [40]. Interaction between Anp32b and Klf5 [36], between Runx1 and Cbfb [29] and between Rad51 and Brca2 [35].

Table 1A								
		Flag#1		Flag#2		Flag#3		
Protein	Accession	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Average mascot
Sox2	gi|127140986	625	6	650	7	143	2	
Transcription factors								
Sall4	gi|81913723	2602(641)	31(8)	2414(206)	30(3)	870(353)	12(5)	1962
Sall1	gi|14164331	1105	16	1195	16	158(147)	2(2)	819
Tbx3	gi|37622347	594	7	374	5	-	-	484
Ubp1	gi|7305605	367	4	465	6	-	-	416
Nac1	gi|81886163	276	4	378	5	45	1	350
Tcfeb	gi|4454797	220	4	78	2	-	-	259
Esrrb	gi|6166153	347	5	-	-	157(79)	2(1)	252
Runx1	gi|2498127	187	4	-	-	278	4	233
Erf	gi|3023704	406	4	-	-	47	1	227
Tcfcp2l1	gi|134053939	462	7	102	2	69	1	211
Rbpj	gi|94400775	285	6	-	-	85	2	185
Gata4	gi|3183530	69	1	213	3	255	3	179
Nanog	gi|31338864	263	3	169	3	73	1	168
Klf5	gi|31981873	269	3	-	-	65	1	167
Sp1	gi|2997601	210	2	97	1	-	-	154
Dpf2 (requiem)	gi|6755314	185	2	-	-	99(168)	1(2)	142
Zfp143	gi|22902397	184	2	68	1	-	-	126
Nr0b1 (Dax1)	gi|6671531	180	3	64	1	120	2	121
Foxp4	gi|81881776	110	3	136	4	-	-	123
Gtf2f2	gi|39930425	98	2	150	3	-	-	124
Transcription related factors								
Sin3A	gi|726286	232	5	512	10	-	-	488
Rbm14	gi|86262142	466	7	454	8	220	3	380
Myef2	gi|148696180	341	6	411	8	277	4	343
Crebbp (p300/CBP)	gi|495128	406	5	254	3	-	-	330
Chd7	gi|124487249	189	3	464	9	168(70)	4(2)	274
Yeats4	gi|13386064	182	3	87	1	-	-	135
Taf9	gi|18079237	-		186	3	96	2	141
Cdk9	gi|18699998	106	2	128	3	-	-	117
Polr2e	gi|20071300	99	2	-	-	105	1	102
Klf5 complex								
Anp32b	gi|18700032	265	3	159	4	217	3	214

Table 1A								
		Flag#1		Flag#2		Flag#3		
Protein	Accession	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Average mascot
Runx1 complex								
CbfB	gi|61557239	182	2	185	2	-	-	184
Rad51 Complex								
Rad51	gi|6755276	216	3	54	1	-	-	135
Brca2	gi|1743860	89	3	-	-	174	3	132
Transport								
Xpo4	gi|37360524	3369	39	2940	36	975	12	2428
Ranbp2	gi|153792534	3923	55	1315	17	225(53)	5(1)	1821
Rangap1	gi|15928791	728	10	133	2	294	5	385
Rcc1	gi|34784343	631	7	469	7	-	-	550
Kpnb1	gi|30931411	451	6	84	2	61	1	298
Other								
mCG14769	gi|148698953	749(143)	10(2)	633	9	-	-	691
Set	gi|13591862	455	5	469	6	-	-	462
Raly	gi|423560	492(47)	7(1)	637	9	106	1	412
Akap8	gi|31560394	651	9	326	5	152(104)	2(1)	376
Lmna	gi|162287370	767	10	57	1	102	2	309
Dnttip1	gi|19526976	462	7	102	2	-	-	282
Ppp2cb	gi|4758952	422	7	130(35)	2(1)	-	-	276
Ogt	gi|27499606	287(71)	4	222	5	249	3	253
Ube2i (UBC9)	gi|4507785	277	5	57	1	-	-	167
Tet2	gi|68342456	174	4	189	2	126	3	163
Csnk2b	gi|7106277	246	3	57	1	-	-	152
0610010K14Rik	gi|81917220	179	2	114	1	-	-	147
Wdr68	gi|28879003	99	1	104	1	-	-	102
Neb	gi|148222065	81	5	44	2	-	-	63
Cell-division cycle								
Rif1	gi|47078460	355	7	405	5	512(528)	9(8)	424
Cdc2a (CDK1)	gi|31542366	379	6	-	-	64	1	222
Pcnp	gi|71480098	80	1	161(42)	3(1)	-	-	161
NuRD complex								
Mi-2beta (Chd4)	gi|39204553	2414	39	3617(35)	54(1)	432(259)	9(6)	2154
Mta2	gi|51491880	1321	19	1023	15	623(107)	9(2)	989
Mta1	gi|15077051	1110	14	907	14	372(73)	6(1)	796
Gatad2b	gi|21314854	877(121)	11(2)	759	10	332(104)	5(2)	656
Mbd3	gi|7305261	871	10	647	9	358(116)	4(2)	625

Table 1A								
		Flag#1		Flag#2		Flag#3		
Protein	Accession	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Average mascot
Gatad2a	gi|30795222	896(97)	11(1)	678	11	161	4	578
Mta3	gi|18381007	681(104)	9(2)	419	6	251(113)	4(2)	450
SWI/SNF								
Smarca4 (Brg1)	gi|76253779	539	7	825(80)	15(1)	180(121)	2(1)	515
Smarcc1 (Baf155)	gi|30851572	458	8	1032(97)	17(1)	45(150)	1(3)	511
N-CoR/SMRT								
Tbl1xr1	gi|12006108	342	4	368	5	-	-	355
Tbl1x	gi|33468969	326	5	309	5	-	-	318
Ncor2	gi|4559296	88	2	145	3	-	-	117
Hdac3	gi|3676558	-	-	109	1	98	1	104
PRC1 complex								
Phc1	gi|110835733	457	6	187	2	-	-	322
Ring1B (Rnf2)	gi|2239142	468	6	301	5	-	-	385
Rybp	gi|5381327	190	2	-	-	115	1	153
LSD1 complex								
Aof2 (Lsd1)	gi|51315882	352	6	205	2	65	1	207
Rcor2	gi|17298682	103	1	86	1	78	1	89

Table 1A. Sox2-interacting proteins identified by mass spectrometry analysis of purified FLAG-Sox2 samples. Shaded cells in FLAG#3 purification indicates that the Mascot score of identified peptides were less than three times the Mascot score in the corresponding control.

Table 1B								
		Flag#1		Flag#2		Flag#3		
Protein	Accession	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Mascot [a]	Pept.[b]	Average mascot
Transcription factors								
Oct4 (Pou5f1)	gi|53501	357(109)	4(1)	339(226)	4(3)	202	2	299
NuRD complex								
Rbbp7	gi|157909799	824(330)	11(7)	966(227)	13(4)	264(151)	4(2)	685
Hdac2	gi|3023934	796(306)	11(6)	424(71)	6(1)	200	3	473
SWI/SNF								
Baf60a (Smarcd1)	gi|1549249	233(249)	4(5)	522	10	-	-	378
Ini1 (Smarcb1)	gi|55956801	196(282)	2(1)	227	3	-	-	212

Table 1B. Sox2-interacting proteins identified by mass spectrometry analysis of purified FLAG-Sox2 samples. Shaded cells indicates that the Mascot score of identified peptides were less than three times the Mascot score in the corresponding control.

[a] Mascot score of the specified protein in the FLAG-Sox2 sample, purified by FLAG- affinity, mascot score of the corresponding control purification is between brackets.

[b] Number of unique peptides for the specified protein in the FLAG-Sox2 sample, number of unique peptides in the control purification is between brackets.

Table 2

Protein	Description	Promoter Co-occup. with Sox2 [a]	Gene bound by Sox2 [a]	Expression changes upon Sox2 depletion [a]	ESC depletion phenotype	Developmental phenotype (E of lethality)
Sox2	SRY-box containing gene 2	Yes	Yes	-	Differentiation [4]	Before E7.5[7]
Transcription factors						
Sall4	Sal-like 4	-	Yes	-	Differentiation [51]	Before E5.5 [51]
Sall1	Sal-like 1	-	Yes	-	-	Perinatal [52]
Tbx3	T-box 3	-	Yes	Down	Differentiation [21]	~ E12.5 [53]
Ubp1	upstream binding protein 1	-	No	-	-	~ E11.5 [54]
Nac1	nucleus accumbens associated 1, BEN and BTB (POZ) domain containing	Yes	-	-	Differentiation [16]	Viable [55]
Oct4 (Pou5f1)	POU domain, class 5, transcription factor 1	Yes	Yes	Down	Differentiation [21]	~ E3.5 [56]
Tcfeb	transcription factor EB	-	Yes	-	-	~ E10.5 [57]
Esrrb	Estrogen related receptor, beta	-	Yes	Up	Differentiation [21]	~ E10.5 [58]
Runx1	runt related transcription factor 1		No	Up		E12.5 [59]
Erf	Ets2 repressor factor	-	Yes	-	-	~ E10.5 [60]
Tcfcp2l1	Transcription factor CP2-like 1	Yes	Yes	-	-	Postnatal lethality [61]
Rbpj	recombination signal binding protein for immunoglobulin kappa J region	-	Yes	-	Not detected [62]	Before E10.5 [62]
Gata4	GATA binding protein 4	-	Yes	Up	Not detected [27]	~ E9.5 [63]
Nanog	Nanog homeobox	Yes	Yes	Down	Differentiation [19, 20]	~ E4.5 [19, 20]
Klf5	Kruppel-like factor 5	Yes	Yes	-	Differentiation [64]	Before E8.5 [65]
Sp1	trans-acting transcription factor 1	-	No	-	-	~ E11 [66]
Dpf2 (requiem)	D4, zinc and double PHD fingers family 2	-	Yes	-	-	-
Zfp143	zinc finger protein 143	-	No	-	Differentiation [22]	-
Nr0b1 (Dax1)	nuclear receptor subfamily 0, group B, member 1	Yes	Yes	Down	Differentiation [67]	-
Foxp4	forkhead box P4	-	Yes	-	-	Before E12.5 [68]
Gtf2f2	general transcription factor IIF, polypeptide 2	-	No	-	-	-
Transcription related factors						
Sin3A	transcriptional regulator, repressor	-	No	Down	-	Before E6.5 [69]
Rbm14	RNA binding motif protein 14	-	Yes	-	-	-
Myef2	myelin basic protein expression factor 2, repressor	-	No	Down	-	-

Table 2						
Protein	**Description**	**Promoter Co-occup. with Sox2** [a]	**Gene bound by Sox2** [a]	**Expression changes upon Sox2 depletion** [a]	**ESC depletion phenotype**	**Developmental phenotype (E of lethality)**
Crebbp (p300/CBP)	CREB binding protein	-	Yes	-	-	~ E9.5 [70]
Chd7	chromodomain helicase DNA binding protein 7	-	Yes	-	-	E10.5 [71]
Yeats4	YEATS domain containing 4	-	No	-	-	-
Taf9	TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor	-	No	-	-	-
Cdk9	cyclin-dependent kinase 9 (CDC2-related kinase)	-	No	-	-	-
Polr2e	polymerase (RNA) II (DNA directed) polypeptide E	-	Yes	-	-	-
Klf5 complex						
Anp32b	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	-	No	Down	-	-
Runx1 complex						
CbfB	core binding factor beta	-	Yes	-	-	E12.5 [72]
Complex				-	-	
Rad51	RAD51 homolog (S. cerevisiae)	-	No	-	Lethality [73]	E2.5 [73]
Brca2	breast cancer 2	-	No	Down	Lethality [74]	E7.5 [74]
Transport						
Xpo4	exportin 4	-	Yes	-	-	-
Ranbp2	RAN binding protein 2	-	Yes	-	-	Before E13.5 [75]
Rangap1	RAN GTPase activating protein 1	-	No	-	-	~ E6.5 [76]
Rcc1	regulator of chromosome condensation 1	-	No	-	-	-
Kpnb1	karyopherin (importin) beta 1	-	Yes	Down	-	Before E5.5 [77]
Other						
mCG14769	-	-	-	-	-	-
Set	SET translocation	-	Yes	Down	-	-
Raly	hnRNP-associated with lethal yellow	-	No	-	-	-
Akap8	A kinase (PRKA) anchor protein 8	-	Yes	-	-	-
Lmna	lamin A	-	Yes	Up	-	Postnatal lethality [78]
Dnttip1	deoxynucleotidyltransferase, terminal, interacting protein 1	-	Yes	-	-	-
Ppp2cb	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	-	Yes	-	-	-

Table 2

Protein	Description	Promoter Co-occup. with Sox2 [a]	Gene bound by Sox2 [a]	Expression changes upon Sox2 depletion [a]	ESC depletion phenotype	Developmental phenotype (E of lethality)
Ogt	O-linked N-acetylglucosamine (GlcNAc) transferase	-	No	-	Lethality [79]	~ E5 [79]
Ube2i (UBC9)	ubiquitin-conjugating enzyme E2I	-	No	-	-	Before E7.5 [80]
Tet2	tet oncogene family member 2	-	-	-	-	PN1 [81]
Csnk2b	casein kinase 2, beta polypeptide	-	No	-	Lethality [82]	~ E6.5 [82]
0610010K14Rik	-	-	-	-	-	-
Wdr68	WD repeat domain 68	-	Yes	-	-	-
Neb	nebulin	-	-	Up	-	Postnatal lethality [83]
Cell-division cycle						
Rif1	Rap1 interacting factor 1 homolog (yeast)	-	Yes	Down	Differentiation [15]	-
Cdc2a (CDK1)	cell division cycle 2 homolog A (S. pombe)	-	No	-	Lethality [84]	Before E1.5 [84]
Pcnp	PEST proteolytic signal containing nuclear protein	-	No	-	-	-
Mi-2beta (Chd4)	chromodomain helicase DNA binding protein 4	-	Yes	-	-	-
Mta1	metastasis associated 1	-	Yes	-	Differentiation [37]	-
Rbbp7	retinoblastoma binding protein 7	-	Yes	Up	-	-
Gatad2b	GATA zinc finger domain containing 2B	-	Yes	-	-	-
Mbd3	methyl-CpG binding domain protein 3	-	Yes	-	Increased self-renewal [85]	~ E8 [85]
Gatad2a	GATA zinc finger domain containing 2A	-	No	-	Not detected [86]	~ E10.5 [86]
Hdac2	histone deacetylase 2	-	No	Down	Not detected [37]	Perinatal [87]
Mta3	metastasis associated 3	-	Yes	-	-	-
SWI/SNF						
Brg1 (Smarca4)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	-	No	-	Differentiation[88]	Before E6.5 [88]
Baf155 (Smarcc1)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	-	Yes	Down	Lethality [89]	Before E5.5 [89]
Baf60a (Smarcd1)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1	-	Yes	Down	-	-

Table 2						
Protein	**Description**	**Promoter Co-occup. with Sox2** [a]	**Gene bound by Sox2** [a]	**Expression changes upon Sox2 depletion** [a]	**ESC depletion phenotype**	**Developmental phenotype (E of lethality)**
Ini1 (Smarcb1)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	-	Yes	Up	Lethality [90]	Before E5.5 [90]
N-CoR/SMRT						
Tbl1xr1	transducin (beta)-like 1X-linked receptor 1	-	Yes	-	-	-
Tbl1x	transducin (beta)-like 1 X-linked	-	Yes	-	Not detected [91]	-
Ncor2	nuclear receptor co-repressor 2	-	No	Up	-	Before E16.5 [92]
Hdac3	histone deacetylase 3	-	No	-	-	Before E9.5 [93]
PRC1 complex						
Phc1	polyhomeotic-like 1 (Drosophila)	-	Yes	Down	-	Peri-natal [94]
Ring1B (Rnf2)	ring finger protein 2	Yes	No	-	Differentiation [39, 95]	Before E10.5 [96]
Rybp	RING1 and YY1 binding protein	-	Yes	-	Not detected [97]	Before E7.5 [97]
LSD1 complex						
Aof2 (Lsd1)	amine oxidase (flavin containing) domain 2	-	No	-	-	Before E7.5 [40]
Rcor2	REST corepressor 2	-	Yes	Down	-	-

Table 2. Transcriptional network and phenotype of Sox2-interacting proteins

[a] Criteria for promoter co-occupancy with Sox2, encoding gene bound by Sox2, expression change upon Sox2 depletion in 2TS22C ES cells [4] and references for (genome-wide) chromatin immunoprecipitation data are in the Material and methods.

References

1. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156 (1981).
2. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78, 7634-7638 (1981).
3. Niwa, H. How is pluripotency determined and maintained? Development 134, 635-646 (2007).
4. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol 9, 625-635 (2007).
5. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-1234 (2007).
6. Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y. & Pallavi, B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol 39, 2195-2214 (2007).
7. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17, 126-140 (2003).
8. Zhao, S., Nichols, J., Smith, A.G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol Cell Neurosci 27, 332-342 (2004).
9. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567-582 (2008).
10. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-676 (2006).
11. Zhou, Q., Chipperfield, H., Melton, D.A. & Wong, W.H. A gene regulatory network in mouse embryonic stem cells. Proc Natl Acad Sci U S A 104, 16438-16443 (2007).
12. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-956 (2005).
13. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106-1117 (2008).
14. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132, 1049-1061 (2008).
15. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38, 431-440 (2006).
16. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. Nature 444, 364-368 (2006).
17. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S.H. Cell 132, 1049-1061 (2008).
18. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379-391 (1998).
19. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631-642 (2003).
20. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643-655 (2003).
21. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533-538 (2006).
22. Chen, X., Fang, F., Liou, Y.C. & Ng, H.H. Zfp143 regulates Nanog through modulation of Oct4 binding. Stem Cells 26, 2759-2767 (2008).
23. Remenyi, A. et al. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev 17, 2048-2059 (2003).
24. Jiang, J. et al. A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol 10, 353-360 (2008).
25. Fujikura, J. et al. Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev 16, 784-789 (2002).
26. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell 10, 615-624 (2006).
27. Narita, N., Bielinska, M. & Wilson, D.B. Cardiomyocyte differentiation by GATA-4-deficient embryonic stem cells. Development 124, 3755-3764 (1997).
28. Murakami, A., Shen, H., Ishida, S. & Dickson, C. SOX7 and GATA-4 are competitive activators of Fgf-3 transcription. J Biol Chem 279, 28564-28573 (2004).

29. Speck, N.A. Core binding factor and its role in normal hematopoietic development. Curr Opin Hematol 8, 192-196 (2001).
30. Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G. & Downing, J.R. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell 84, 321-330 (1996).
31. Okumura-Nakanishi, S., Saito, M., Niwa, H. & Ishikawa, F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem 280, 5307-5317 (2005).
32. Marshall, N.F., Peng, J., Xie, Z. & Price, D.H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J Biol Chem 271, 27176-27183 (1996).
33. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77-88 (2007).
34. Tichy, E.D. & Stambrook, P.J. DNA repair in murine embryonic stem cells and differentiated cells. Exp Cell Res 314, 1929-1936 (2008).
35. Lee, S.A., Roques, C., Magwood, A.C., Masson, J.Y. & Baker, M.D. Recovery of deficient homologous recombination in Brca2-depleted mouse cells by wild-type Rad51 expression. DNA Repair (Amst) (2008).
36. Munemasa, Y. et al. Promoter region-specific histone incorporation by the novel histone chaperone ANP32B and DNA-binding factor KLF5. Mol Cell Biol 28, 1171-1181 (2008).
37. Liang, J. et al. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol 10, 731-739 (2008).
38. Kidder, B.L., Palmer, S. & Knott, J.G. SWI/SNF-Brg1 Regulates Self-Renewal and Occupies Core Pluripotency-Related Genes in Embryonic Stem Cells. Stem Cells (2008).
39. Endoh, M. et al. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development 135, 1513-1524 (2008).
40. Wang, J. et al. Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature 446, 882-887 (2007).
41. Ambrosetti DC, B.C., Dailey L. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol. Nov;17(11), 6321-6329 (1997).
42. van den Berg, D.L. et al. Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. Mol Cell Biol 28, 5986-5995 (2008).
43. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol 21, 183-186 (2003).
44. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11, 1475-1489 (1983).
45. Sanchez, C. et al. Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics 6, 820-834 (2007).
46. Sharov, A.A. et al. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics 9, 269 (2008).
47. Zhang, Y. et al. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev 13, 1924-1935 (1999).
48. Mohrmann, L. & Verrijzer, C.P. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim Biophys Acta 1681, 59-73 (2005).
49. McDonel, P., Costello, I. & Hendrich, B. Keeping things quiet: Roles of NuRD and Sin3 co-repressor complexes during mammalian development. Int J Biochem Cell Biol 41, 108-116 (2009).
50. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer 6, 846-856 (2006).
51. Elling, U., Klasen, C., Eisenberger, T., Anlag, K. & Treier, M. Murine inner cell mass-derived lineages depend on Sall4 function. Proc Natl Acad Sci U S A 103, 16319-16324 (2006).
52. Nishinakamura, R. et al. Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development. Development 128, 3105-3115 (2001).
53. Davenport, T.G., Jerome-Majewska, L.A. & Papaioannou, V.E. Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. Development 130, 2263-2273 (2003).
54. Parekh, V. et al. Defective extraembryonic angiogenesis in mice lacking LBP-1a, a member of the grainyhead family of transcription factors. Mol Cell Biol 24, 7113-7129 (2004).

55. Mackler, S. et al. Requirement for the POZ/BTB protein NAC1 in acute but not chronic psychomotor stimulant response. Behav Brain Res 187, 48-55 (2008).
56. Kehler, J. et al. Oct4 is required for primordial germ cell survival. EMBO Rep 5, 1078-1083 (2004).
57. Steingrimsson, E., Tessarollo, L., Reid, S.W., Jenkins, N.A. & Copeland, N.G. The bHLH-Zip transcription factor Tfeb is essential for placental vascularization. Development 125, 4607-4616 (1998).
58. Luo, J. et al. Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. Nature 388, 778-782 (1997).
59. Ray, M.K., Wang, G., Barrish, J., Finegold, M.J. & DeMayo, F.J. Immunohistochemical localization of mouse Clara cell 10-KD protein using antibodies raised against the recombinant protein. J Histochem Cytochem 44, 919-927 (1996).
60. Papadaki, C. et al. Transcriptional repressor erf determines extraembryonic ectoderm differentiation. Mol Cell Biol 27, 5201-5213 (2007).
61. Yamaguchi, Y., Yonemura, S. & Takada, S. Grainyhead-related transcription factor is required for duct maturation in the salivary gland and the kidney of the mouse. Development 133, 4737-4748 (2006).
62. Oka, C. et al. Disruption of the mouse RBP-J kappa gene results in early embryonic death. Development 121, 3291-3301 (1995).
63. Molkentin, J.D., Lin, Q., Duncan, S.A. & Olson, E.N. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes Dev 11, 1061-1072 (1997).
64. Parisi, S. et al. Klf5 is involved in self-renewal of mouse embryonic stem cells. J Cell Sci 121, 2629-2634 (2008).
65. Shindo, T. et al. Kruppel-like zinc-finger transcription factor KLF5/BTEB2 is a target for angiotensin II signaling and an essential regulator of cardiovascular remodeling. Nat Med 8, 856-863 (2002).
66. Marin, M., Karis, A., Visser, P., Grosveld, F. & Philipsen, S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell 89, 619-628 (1997).
67. Niakan, K.K. et al. Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. Mol Genet Metab 88, 261-271 (2006).
68. Li, S., Zhou, D., Lu, M.M. & Morrisey, E.E. Advanced cardiac morphogenesis does not require heart tube fusion. Science 305, 1619-1622 (2004).
69. Dannenberg, J.H. et al. mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. Genes Dev 19, 1581-1595 (2005).
70. Kanellopoulou, C. et al. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19, 489-501 (2005).
71. Hurd, E.A. et al. Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. Mamm Genome 18, 94-104 (2007).
72. Sasaki, K. et al. Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. Proc Natl Acad Sci U S A 93, 12359-12363 (1996).
73. Tsuzuki, T. et al. Targeted disruption of the Rad51 gene leads to lethality in embryonic mice. Proc Natl Acad Sci U S A 93, 6236-6240 (1996).
74. Ohba, H. et al. Sox21 is a repressor of neuronal differentiation and is antagonized by YB-1. Neurosci Lett 358, 157-160 (2004).
75. Dawlaty, M.M. et al. Resolution of sister centromeres requires RanBP2-mediated SUMOylation of topoisomerase IIalpha. Cell 133, 103-115 (2008).
76. DeGregori, J. et al. A murine homolog of the yeast RNA1 gene is required for postimplantation development. Genes Dev 8, 265-276 (1994).
77. Kitamura, K. et al. Mouse Pitx2 deficiency leads to anomalies of the ventral body wall, heart, extra- and periocular mesoderm and right pulmonary isomerism. Development 126, 5749-5758 (1999).
78. Sullivan, T. et al. Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. J Cell Biol 147, 913-920 (1999).
79. Shafi, R. et al. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proc Natl Acad Sci U S A 97, 5735-5739 (2000).
80. Nacerddine, K. et al. The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. Dev Cell 9, 769-779 (2005).
81. Tang, H., Araki, K., Li, Z. & Yamamura, K. Characterization of Ayu17-449 gene expression and resultant kidney pathology in a knockout mouse model. Transgenic Res 17, 599-608 (2008).
82. Buchou, T. et al. Disruption of the regulatory beta subunit of protein kinase CK2 in mice leads to a cell-autonomous defect and early embryonic lethality. Mol Cell Biol 23, 908-915 (2003).
83. Rosenbaum, R.M., Picciano, P., Kress, Y. & Wittner, M. Ultrastructure of in vitro type 2 epithelial cell cysts derived from adult rabbit lung cells. Anat Rec 188, 241-261 (1977).

84. Santamaria, D. et al. Cdk1 is sufficient to drive the mammalian cell cycle. Nature 448, 811-815 (2007).
85. Kaji, K. et al. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol 8, 285-292 (2006).
86. Marino, S. & Nusse, R. Mutants in the mouse NuRD/Mi2 component P66alpha are embryonic lethal. PLoS ONE 2, e519 (2007).
87. Montgomery, R.L. et al. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. Genes Dev 21, 1790-1802 (2007).
88. Bultman, S. et al. A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Mol Cell 6, 1287-1295 (2000).
89. Kim, J.K. et al. Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. Mol Cell Biol 21, 7787-7795 (2001).
90. Guidi, C.J. et al. Disruption of Ini1 leads to peri-implantation lethality and tumorigenesis in mice. Mol Cell Biol 21, 3598-3603 (2001).
91. Perissi, V., Aggarwal, A., Glass, C.K., Rose, D.W. & Rosenfeld, M.G. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. Cell 116, 511-526 (2004).
92. Ogawa, S. et al. A nuclear receptor corepressor transcriptional checkpoint controlling activator protein 1-dependent gene networks required for macrophage activation. Proc Natl Acad Sci U S A 101, 14461-14466 (2004).
93. Bhaskara, S. et al. Deletion of histone deacetylase 3 reveals critical roles in S phase progression and DNA damage control. Mol Cell 30, 61-72 (2008).
94. Takihara, Y. et al. Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. Development 124, 3673-3682 (1997).
95. van der Stoop, P. et al. Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. PLoS ONE 3, e2235 (2008).
96. Voncken, J.W. et al. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. Proc Natl Acad Sci U S A 100, 2468-2473 (2003).
97. Pirity, M.K., Locker, J. & Schreiber-Agus, N. Rybp/DEDAF is required for early postimplantation and for central nervous system development. Mol Cell Biol 25, 7193-7202 (2005).

Chapter

6

Discussion

Discussion

Sox2 is a transcription factor member of the Sox protein family. Sox2 contains a conserved high-mobility group (HMG) domain which binds DNA in a sequence-specific manner, acting as an architectural protein and as a transcriptional regulator to activate or repress target gene expression [1]. Sox2 protein on its own does not possess sufficient affinity for DNA binding and requires for its transcriptional activity recruitment of other protein partners, such as Oct4, which facilitate and stabilize the formation of Sox2 transcription complexes [2, 3]. The action of Sox2 is therefore context-sensitive, and dependent on promoter environment and binding available partners. In this chapter we will discuss first molecular details of Sox2 activity during very early development (**chapters 4 and 5**) and subsequently its role in lung development (**chapter 3**). We will also discuss potential links of inappropriate Sox2 activity to human congenital malformation and cancer, especially those related to lung development.

I. Molecular details of Sox2 in early development

Sox2 plays an essential role during early development and its expression starts at the morula-stage. In the blastocyst stage it is specifically detected within the cells of the ICM [4]. Sox2 is also expressed in ES cells [4], which are derived from the ICM, and have two defining properties, self-renewal and pluripotency [5, 6]. These properties are maintained by a transcription factor network that appears to be dominated by three transcription factors, Sox2, Oct4 and Nanog [7, 8]. Deletion of Sox2 causes a failure of the ICM to develop [4] and depletion of Sox2 in ES cells results in differentiation into trophectoderm-like cells [7]. It has been suggested that the main function of Sox2 in ES cells is to indirectly maintain the expression of Oct4 [7]. Consistent with this, Oct4 depletion has the same ICM phenotype as Sox2 ablation [9].

Sox2 not only maintains the undifferentiated state of ES cells. Together with Oct4, Klf4 and c-Myc, it can reprogram somatic cells into ES cell-like induced pluripotent stem (iPS) cells, a cell type that has a lot of potential to advance regenerative medicine [10, 11].

1. Sox2 interaction partners and their potential role in Sox2 function

As mentioned above, the interaction of Sox2 with different protein partners is an essential element in determining the functional specificity of Sox2. For that reason it is important to identify partner proteins for Sox2 to understand the molecular mechanisms of Sox2 activity. However, despite the critical role of Sox2 in mammalian development, no interacting proteins, besides Oct4 [7], are known. In **chapter 5**, we used a biochemical approach to identify Sox2 interacting proteins in mouse ES cells. FLAG-tagged Sox2 was affinity-purified and its putative interacting proteins were identified by mass spectrometry. In the list of almost 80 putative Sox2 partners, we found ES cell self-renewal transcription factors, lineage-specific transcription factors, several subunits of chromatin remodeling complexes and transport related factors, including the transport receptor, Exportin 4 (**Figure 1**). This long list of potential partners, which needs to be confirmed by independent methods, appears to confirm that Sox2 requires co-factors for

its specificity and function. However, this result also poses new questions, which need to be answered in the future, to really understand the function of Sox2 in detail. A first question to answer is which interactions are direct? Direct interactors of Sox2 are more likely to be important for its function. One way to determine this is using bacterially expressed recombinant proteins. This is, however, difficult to test for the several multi-subunit complexes that bind to Sox2. Sequential immunoprecipitations may also be a useful tool to investigate the subunit composition of the different Sox2-containing complexes. A further purification of Sox2 interactors from ES cells in the presence or absence of Sox2 [7] may be another way to provide an increase in the resolution of the connections in the Sox2 interaction network.

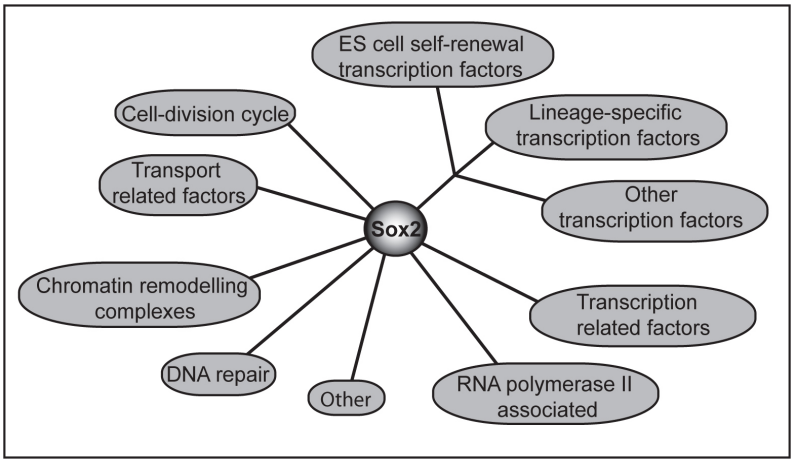


Figure 1. Depiction of the putative Sox2 protein interaction network in mouse ES cells. Classification of the newly identified Sox2 interacting proteins by their function, according to the results obtained in **chapter 5**.

Another question to be answered is how the formation and breakage of the different interactions is regulated. It is unlikely that all 80 Sox2 interactors are binding Sox2 at the same time. A possible way for Sox2 to discriminate different interactions is by post-translational modification, a subject discussed in the next section. Alternatively, once Sox2 binds to DNA, the promoter context, including histone modifications and other transcription factors, becomes the dominant feature in deciding which Sox2 interactions are maintained and which not.

This leads to the question of the functional relevance of the Sox2 interactions. The genome-wide binding sites of Sox2 in ES cells were determined in several studies [12, 13]. **Chapter 5** describes the identification as Sox2 interactors of several transcription factors that co-localize with Sox2 on many promoters, which may indicate their function as Sox2 transcriptional co-factors. Whether their promoter binding depends on Sox2 can be determined using ES cells that can be rapidly depleted for Sox2 [7]. The functional relevance of their promoter presence can be determined by the effect on gene expression of their depletion by RNAi or gene targeted deletion.

2. Nuclear import of Sox proteins

In **chapter 4**, we describe a novel nuclear import pathway for the Sox family of transcription factors. Transcription factors are produced in the cytoplasm and imported into the nucleus to regulate gene expression. A family of transport receptors called importins mediates this active process [14]. The HMG domain contains two highly conserved nuclear localization signals (NLS), which are responsible for the nuclear localization of the Sox proteins [3, 15]. We identified Exp4 as an interaction partner of Sox2 in mouse ES cells in chapter 5 and showed that it acts as a nuclear import receptor for Sox2 and SRY in chapter 4. To date, Exp4 has an established function as a nuclear export receptor for the translation initiation factor eIF-5A and the transcription factor Smad3 [16, 17]. Thus Exp4 is an example of a nuclear transport receptor carrying distinct cargoes into different directions, in Figure 2 we propose a model for the bi-directional nuclear transport mediated by Exportin 4.

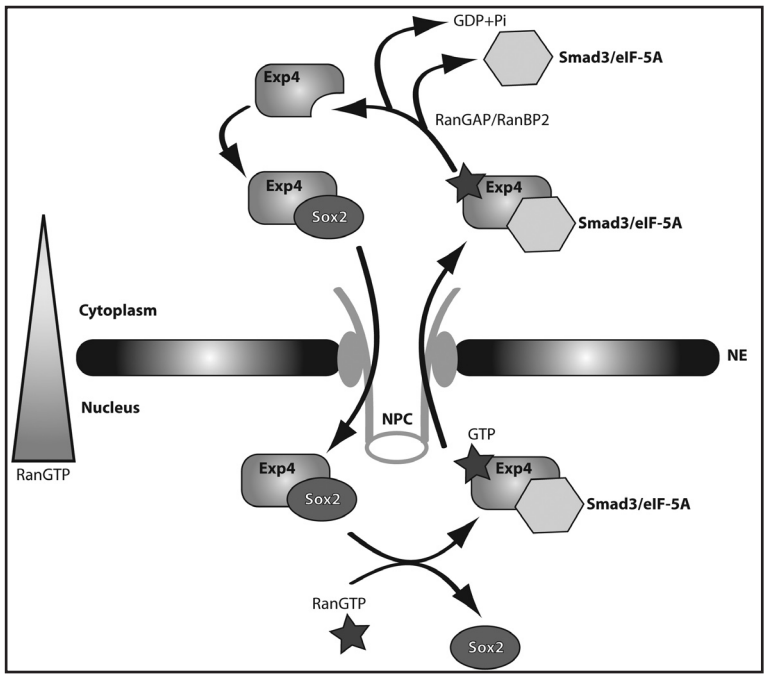


Figure 2. Schematic model for the bi-directional nuclear transport mediated by Exp4. Exportin 4 acts as a Sox2 importin, binding Sox2 at low RanGTP-levels in the cytoplasm, mediating translocation through the nuclear pore complexes (NPC) and releasing Sox2 upon RanGTP-binding in the nucleus (**chapter 4**). Exportin 4 acts as an exportin for Smad3 and eIF-5A [23, 24], Exportin 4 binds cargo together with RanGTP in the nucleus and releases its cargo upon GTP-hydrolysis in the cytoplasm (NE, nuclear envelope; NPC, nuclear pore complex).

The Exp4-Sox protein recognition requires residues in the HMG box domain that are completely conserved throughout the Sox protein family, which suggests that Exp4 operates as a nuclear import receptor for all 20 Sox proteins. We showed this to be true for SRY, the prototypic Sox family member [18], that is efficiently imported into the cell nucleus by binding to Exp4. To prove our hypothesis, it will be necessary to test the other Sox family members. Developmental defects associated with mutations in Sox transcription factors may be caused not only by impairment in their biochemical properties as transcription factors, but also by a deficient nuclear import or export [19]. For example, mutations causing defective nuclear localization of SRY and SOX9 have been associated with human sex reversal syndrome. Sex reversal syndromes are a set of human genetic diseases whereby the gonadal sexuality is not the same as the chromosome sexuality [20]. The different clinical types identified are 46, XY female and 46, XX male [20]. The first type is a form of pure gonadal dysgenesis, also known as Swyer's syndrome, and is associated with defects in SRY or SOX9. Clinical mutations cluster in the region encoding the HMG box, predominantly affecting the ability of SRY or SOX9 to bind to and bend DNA, or to be transported into the nucleus. Subsequently, the (partial) inactivation of SRY or SOX9 by these mutations leads to a failure of testicular differentiation and a female somatic phenotype [21-23].

The SRY N-NLS sex reversal mutants used in chapter 4 have in addition to an impaired nuclear localization, also decreased DNA binding and transcriptional activity. A sex-reversal causing mutation in the SRY N-NLS (M64I) was shown to impair SRY nuclear localization, but not DNA binding or transcriptional activity [24]. It will be interesting to analyze whether this mutant SRY protein is still able to bind and be imported by Exp4. A defect in the nuclear import of this mutant will suggests that the sex reversal phenotype caused by this mutation is solely due to the defective import of the mutant protein by import factors, such as Exp4.

A recent study proposed that Impβ alone can import Sox2 in undifferentiated ES cells and that Impβ in conjunction with Impα3 or Impα5 import Sox2 in neural progenitors [25]. However, we found no evidence for any Impβ/α combination stimulating Sox2-import. Instead, we found that Imp9 and the Impβ/7 heterodimer mediate nuclear import of Sox2 in parallel to Exp4 and these three import pathways share the same Sox2 import signals. The import of Sox2 by multiple import pathways makes it difficult to predict the biological and cellular impact of Exp4 depletion. Exp4-mediated import may be rate-limiting for certain cargoes in certain tissues depending on the presence of alternative import routes. Also, the role of Exp4 as an exportin complicates the molecular dissection of the role of Exp4 in mammalian development.

The identification of multiple nuclear import pathways for Sox2 makes it unlikely that Sox2 nuclear localization is regulated by the presence of specific import receptors. Regulation of Sox2 localization might be achieved by modifying Sox2, a strategy that does not affect other cargoes that depend on the same import pathway [26, 27].

3. Post-translational modifications of Sox proteins

Post-translational modification of cargo proteins, such as sumoylation, acetylation or phosphorylation, may be important in the regulation of their transport between the cytoplasm and nucleus by inducing conformational changes that will expose or mask signals that are important in the recognition by the transport receptors [28]. In the case of eIF-5A, Exp4 recognizes a complex export signal that includes a hypusine post-translational modification of eIF-5A [17]. For Smad3, Exp4 preferentially associates with its unphosphorylated form [16]. We have shown that Exp4 is able to efficiently import recombinant Sox2 protein, produced in bacteria, and therefore probably free of post-translational modifications. However, as in the case of the other two known cargos for Exp4, it is still possible that the affinity of Sox2 for Exp4 is modulated by post-translational modification of Sox2. Interestingly, in the late blastocyst stage Sox2 is nuclear only in the cells of the ICM, while in trophectoderm cells it is exclusively cytoplasmic [4]. The analysis of the post-translational modifications of the Sox2 protein present in the cytoplasm of trophectoderm cells, together with the mutation of potential post-translational modifications sites could provide insight to how nuclear accumulation or exclusion of Sox2 is regulated.

Post-translational modifications are a mechanism to modulate not only nucleocytoplasmic transport, but also the general activity of Sox proteins. Many Sox proteins feature putative consensus recognition sites for sumoylation, acetylation or phosphorylation. However, the actual occurrence of such modifications and their function *in vivo* are not well studied [29]. It would be interesting to study whether these modifications alter the affinity of Sox proteins for Exp4. So far, the only described post-translational modification for Sox2 is the SUMO modification (at K247) [30]. Sumoylation is a reversible modification by SUMO (small-ubiquitin-related modifier), through a series of enzymatic reactions. The small polypeptide SUMO (11 kDa) is covalently attached to a specific lysine (K) residue in the protein substrate [30]. Many proteins involved in transcriptional regulation can be sumoylated, which might produce some conformational changes or create new interaction surfaces for protein-protein interaction. These changes can have variable substrate specific consequences for protein activity, localization, stability, DNA-binding efficiency and protein interactions [31]. There are several examples of the importance of sumoylation as a mechanism to regulate Sox proteins. A study in Xenopus embryos showed that the activity of SoxE during development is regulated by SUMO, as sumoylated SoxE promotes inner ear development while unmodified SoxE promotes neural crest development [32]. Another example, is that mutation of the SUMO acceptor lysine in human SOX3 increased its activity, which suggests that sumoylation inhibits transcriptional activation by SOX3 [31] . In the case of Sox2, it has been described that in mouse ES cells Sox2 is mostly present in the unmodified form [30]. In **chapter 5**, we identified Ubc9 (E2 SUMO-conjugating enzyme) and Ranbp2 (SUMO-ligase) as putative Sox2 partner proteins. Ranbp2 is located at the cytoplasmic rim of the NPC [33], and might sumoylate Sox2 when the protein is transported in and out of the nucleus through the NPC. A recent study suggests that sumoylation regulates the interaction of Sox2 to the

target DNA [30]. SUMO modification of Sox2 at K247 negatively affected Sox2-mediated gene expression of the *Fgf4* gene by reducing the DNA-binding ability of Sox2 [30]. The SUMO modification of Sox2 occurs outside the HMG domain, which suggests that sumoylation of Sox2 causes a conformational change that affects DNA binding and/or protein-protein interactions [30].

A recent study suggested that regulation of the SRY subcellular distribution is also controlled by post-translational modifications [34]. SRY interacts with the histone acetyltransferase (HAT), p300, which mediates acetylation of SRY (at K136) and increases its nuclear accumulation. Conversely, specific deacetylation of K136 by the histone deacetylase, HDAC3, or site-specific mutation of this residue, which abolishes acetylation, induced a partial cytoplasmic localization [35]. The acetylated SRY motif is close to the C-terminal NLS and is conserved across species and within the SoxB, C and G groups [35]. Interestingly, in **chapter 5**, we found that Sox2 is also able to bind to the HAT, CREB-binding protein (p300/CBP) and to the histone deacetylase, HDAC3. This may suggest that acetylation by p300/CBP may also affect Sox2 cellular localization.

Acetylation/deacetylation modification is also implicated in the control of gene expression. Acetylation of transcription factors can induce transcriptional activation by increasing their DNA binding capacity. In contrast, transcription factor deacetylation is generally associated with the suppression of gene transcription, depending on the promoter context [36].

Phosphorylation is another common modification that regulates nuclear import of proteins, especially when extracellular stimuli require quick responses. Several Sox proteins have putative phosphorylation sites for different types of kinases [34]. SRY and SOX9 undergo phosphorylation by protein kinase A, which increases their DNA-binding affinity and the transcriptional activity. Additionally, it has been shown that SOX9 phosphorylation allows its nuclear translocation in the developing male gonad [37, 38]. In **chapter 5**, we found protein kinases, Akap8 and Csnk2b, as putative interactors of Sox2, which might suggest that Sox2 can be also phosphorylated.

All these modifications are not mutually exclusive; various types of post-translational modifications can coexist on a target protein at different residues, modulating the balance between nuclear import and export, transcriptional activity and protein-protein interactions [34].

II. Role of Sox2 in human pathology during later embryonic development

Sox2 also plays a role later in development in the development of several organs such as eye development [39], inner ear formation [40], development of taste buds [41], development of the trachea and esophagus [42] and lung development [43] . Here we will discuss the role of Sox2 in human pathologies with an emphasis on those related to lung development.

1. Human congenital anomalies associated with Sox2 mutations

Heterozygous *SOX2* mutations have been associated with human congenital malformations; a summary of the molecular and clinical features associated with some of these *SOX2*

mutations is shown in **Table 1**. We will describe in more detail the malformations most frequently associated with Sox2 mutations.

Heterozygous, *de novo,* loss-of-function *SOX2* mutations have been reported in patients with severe bilateral eye malformations [44]. Normal dosage of the *SOX2* gene is critical for human eye development [39]. Mutations in *SOX2* account for approximately 10% of the cases of anophthalmia (an absence of eye structures) and microphthalmia (an abnormally small eye) [44]. Both are rare disorders that may occur in isolation or as part of a syndrome, as is the case in one-third of the cases [45]. Patients with *SOX2* anophthalmia syndrome usually present extra-ocular abnormalities, including: neurocognitive impairments, seizures, brain malformation, specific motor abnormalities, mild facial dysmorphism and postnatal growth failure. In males, urogenital anomalies are reported, most commonly cryptorchidism, hypospadias and micropenis [46].

The anophthalmia-esophageal-genital (AEG) syndrome is particularly interesting, because it suggests a direct role of SOX2 in the development of the trachea and oesophagus in humans. This hypothesis was supported by a study in mouse showing the dose-dependent roles for Sox2 in the patterning and differentiation of the anterior foregut endoderm [42]. AEG syndrome is an association of anophthalmia/microphthalmia, oesophageal atresia (EA) with or without tracheo-oesophageal fistula (TEF), in which the trachea and esophagus fail to separate, and urogenital anomalies [47]. *SOX2* anophthalmia syndrome and AEG syndrome can be caused by similar heterozygous loss-of-function mutations in *SOX2*, since the genital and eye phenotypes are indistinguishable in both conditions [47].

The Matthew-Wood syndrome is a rare association of microphthalmia/anophthalmia, pulmonary agenesis, and diaphragmatic defect, also known by the acronym "PMD" (Pulmonary agenesis, Microphthalmia, Diaphragmatic defect) [48]. It has been reported to be caused by truncating mutations in the retinol-binding protein receptor gene STRA6 [49]. Mutations in STRA6 cause several malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation [50]. The similarities in the phenotype suggest a possible connection in the functions of SOX2 and STRA6.

SOX2 is also necessary for the normal development and function of the hypothalamo-pituitary axis. Mice heterozygous for a targeted disruption of Sox2 did not have eye defects, but showed abnormal anterior pituitary development with reduced levels of growth hormone, luteinizing hormone, and thyroid-stimulating hormone. Some patients with *SOX2* heterozygous mutations presented, in addition to eye defects, also pituitary abnormalities (anterior pituitary hypoplasia and hypogonadotropic hypogonadism) [51].

Genetic screening for other candidate genes and the identification of genetic factors influencing *SOX2* expression may help to identify other factors that play a role in the occurrence of these severe forms of eye malformation and associated syndromes [44].

Mutation	Nuclear localization	DNA binding	Transactivation	Clinical phenotype
c.60insG	Impaired	Loss	Loss	Bilateral anophthalmia, HH, learning difficulties, esophageal atresia, diplegia
c.70del20	Impaired	Loss	Loss	Left anophthalmia, right microphthalmia, HH, learning difficulties
c.387delC	Unaffected	Unaffected	Impaired	Left microphthalmia, right coloboma, HH, learning difficulties, diplegia
Y160X	Unaffected	Unaffected	Impaired	Bilateral microphthalmia, HH, learning difficulties, quadriparesis
Q177X	Unaffected	Unaffected	Impaired	Bilateral anophthalmia, HH, learning difficulties
c.479delA	Unaffected	Unaffected	Impaired	Bilateral anophthalmia, HH, GH deficiency, learning difficulties, sensorineural deafness
G130A	Unaffected	Unaffected	Unaffected	Bilateral optic nerve hypoplasia, normal growth
A191T	Unaffected	Unaffected	Unaffected	Bilateral optic nerve hypoplasia; GH deficiencie

Table 1. Summary of molecular and clinical features associated with some *SOX2* mutations in humans. HH, hypogonadotropic hypogonadism; GH, growth hormone (adapted from [51]).

2. Role of Sox2 in lung development

During lung development Sox2 has a dynamic expression pattern in the airway epithelium. It is highly expressed in non-branching proximal regions and absent from branching distal regions, which suggests that downregulation of Sox2 at the branching sites is crucial for correct lung development. To test this hypothesis, we generated transgenic mice that ectopically express *Sox2* in the distal airway epithelium using a doxycycline inducible system (SPC-rtTA*pTRE-Tight-myc-Sox2). The SPC promoter drives expression of the rtTA gene in epithelial cells of the embryonic lung, and the rtTA protein can activate the Tet-On promoter under the influence of doxycycline. **Chapter 3** describes the study of the phenotype of these mice. We could show that the spatial regulation of Sox2 expression during lung development is critical for the normal branching morphogenesis and the correct differentiation of the pulmonary epithelium. The transgenic lungs presented a normal lung lobulation but the subsequent branching was hampered, leading to a

severe reduction in the number of airways and to enlarged, cyst-like airspaces. This may suggest that formation of the main bronchi and branching morphogenesis are controlled by different molecular mechanisms, with Sox2 overexpression only affecting the latter one. Alternatively, at E9.5, when the lung bud starts to form, the level or distribution of transgenic Sox2 is not sufficient to prevent branching, or the epithelial cells are already programmed to enter the branching process at the time the transgene is switched on.

Our results confirmed that Sox2 downregulation at branching sites is required for the epithelial cells to be able to respond normally to branch-inducing signals, since continuous Sox2 expression blocks the branching capacity of the airway epithelium. This was supported by the analysis of Fgf10 and Bmp4 signaling in E12.5 lung embryos, when the branching defect starts. The Fgf10 branch-inducing signal is normally active at the distal *epithelium but the* transgenic lungs are no longer responsive. The transgenic lungs become responsive to the autocrine *Bmp4* branch-inhibiting signal, which is normally not active at the distal epithelium (see model, **Figure 3**). Further work is needed to establish the precise mechanisms by which Sox2 overexpression affects both signaling pathways.

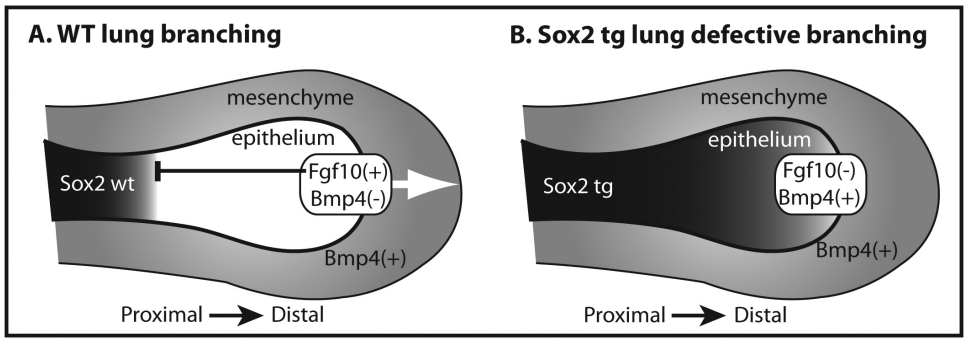


Figure 3. Model for the branching defect in Sox2 overexpressing lungs. Figure shows a developing lung bud during branching morphogenesis. Sox2 expression is depicted in black. (**A**) In control lungs Sox2 is expressed in non-branching proximal regions and absent from branching distal regions. The active *Fgf10* signaling induces branching of the distal epithelium (white arrow) and represses Sox2 expression [42]. (**B**) In the transgenic lungs, Sox2 is also expressed in the distal airway epithelium which is now no longer responding to Fgf10 signaling and becomes responsive to autocrine Bmp4 signaling, which together blocks the branching capacity of the distal epithelium. In both situations, the paracrine Bmp4 signaling is normally active in the mesenchyme. See **chapter 3** for further details.

Epithelial cell differentiation was also affected in the transgenic lungs, although it was not completely disrupted, since epithelial cells were still able to initiate differentiation into several lineages along the proximo-distal axis of the fetal lung. Sox2 overexpression prematurely induces cells into a differentiation program, leading to an increase in the number of pulmonary neuroendocrine cells (PNECs) and (pre-) basal cells,

which are both committed progenitor cells. In addition, cells with a gastric/intestinal expression pattern were also found in the transgenic lungs.

Interestingly, the morphology of Sox2 overexpressing lungs resembles the phenotype of human congenital cystic adenomatoid malformation (CCAM), a rare, non-hereditary abnormality of lung development. CCAMs are dysplastic lung tumors characterized by the overgrowth of the terminal respiratory bronchioles. The cystic lesions are usually benign, unilateral and only in one lobe, but can extend to a whole lung, and a neonatal pneumonectomy may then be necessary [52] . CCAM is classified into five pathological types (0-IV), depending on cyst size and structure [53, 54]. The pathogenesis is unknown, CCAM develops during the pseudoglandular period and it has been suggested that the different types originate as a result of a localized arrest in the bronchial tree branching morphogenesis, at different time points during lung development [52]. Type 0 (acinar dysplasia) and type I (multiple large cysts or a single dominate cyst) are the more similar to the phenotype of our transgenic mice. Preliminary data showed that the Sox2 protein was expressed in a higher level in the cyst-like structures of the human CCAM lungs compared with normal lungs (**Figure 4**). This observation was corroborated by published microarray expression data from CCAM samples and the relative controls, in which it was found that Sox2 expression was up-regulated (6-fold) in the postnatal CCAM compared with the control [52]. This may suggest that Sox2 expression is involved in the pathogenesis of CCAM in humans. It is, however, unclear if there is a causal relationship between *SOX2* misexpression and the CCAM lesions. The observed Sox2 overexpression could also be a consequence of the presence of immature bronchopulmonary tissue in the dysplastic lung, a tissue in which Sox2 is expressed during normal lung development.

Another interesting question would be whether the overexpression of Sox2 is specific for CCAM or a general characteristic of also other congenital lung malformations. To answer this question, we will have to analyze Sox2 expression in congenital lung malformations, such as congenital diaphragmatic hernia (CDH), congenital lobar emphysema (CLE) or bronchopulmonary sequestration (BPS). CDH refers to a defect in the diaphragma that allows abdominal organs to be present in the chest cavity, which results in pulmonary hypoplasia [55]. CLE refers to an overexpansion of a pulmonary lobe and the resultant compression of the remaining ipsilateral lung [56]. BPS consists of a non-functioning mass of lung tissue that lacks normal communication with the tracheobronchial tree and that receives arterial blood supply from the systemic circulation [57]. These three congenital malformations all show immature hypoplastic or dysplastic lung tissue where Sox2 expression might be detected.

To complete our study of the role of Sox2 during lung development, it would be necessary to complement our overexpression study with a loss-of-function approach. *Sox2* can be conditionally deleted in the lung epithelium by crossing the available conditional Sox2 knock-out mice [39] with SPC-Cre mice. Deletion of Sox2 in lung epithelium may not result in a severe phenotype, due to functional redundancy with other Sox family members present in the lung epithelium. For example, Sox9 is also expressed in the respiratory epithelium and in the proximal mesenchyme throughout lung morphogenesis [58]. However, Sox9-deletion in the respiratory epithelium resulted in completely normal lungs [58].

It would be especially interesting to study the effect of Sox2-depletion on the formation of PNECs. If Sox2 is required for the differentiation of these cells, PNECs will be absent in the mutant mice. Sox17 overexpression in the lung epithelium also induces an increase in the number of PNECs [59], which might suggest an overlap in Sox2 and Sox17 function.

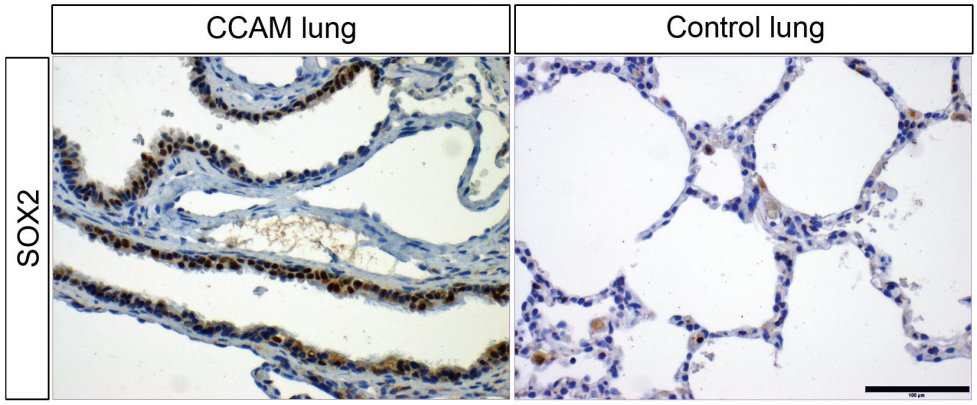


Figure 4. Sections of CCAM affected and control human lungs were stained with Sox2. Sox2 was detected in epithelial cells surrounding the cysts of the CCAM affected lung (bronchiolar, type II). Sox2 expression was not detected in the alveoli of control lungs. Bar, 100 µm.

To better understand the molecular circuit of Sox2 in the lung, identification of the downstream targets of Sox2 in the embryonic lung epithelium would be of great help. To date, analysis of the function of Sox2 has focused on early embryonic development, as mentioned in **chapter 5**, and there is a lot of information available on the putative target genes of Sox2 in mouse and human ES cells [12, 13, 60]. We used this information to identify putative downstream targets of Sox2 in lung epithelium. We searched for genes that are mis-regulated in an opposite direction in Sox2-depleted ES cells [7], as compared to Sox2-overexpressing lung epithelium, as described in **chapter 3** (**Table 2**). We identified around 20 genes that are down-regulated in ES cells after Sox2 depletion and up-regulated in the lung after Sox2 overexpression, and around 10 genes that are regulated in the opposite way. From these genes, more than half have been reported as putative target genes for Sox2 in ES cells [12], which suggest that they maybe directly regulated by Sox2. The rest of the genes might be indirectly regulated by Sox2.

3. Association of Sox2 with cancer

In the list of genes in **Table 2**, there are several putative direct targets of Sox2: *Cdh2*[62], *Mycl1*[63], *Sox21*[64], *Ptn*[65], *Mad2l1*[66] and *Xrrc5*[67], which are expressed in lung tumors. Lung cancer is classified into two major categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC accounts for about 20% of all lung cancers, is a carcinoma of neuroendocrine origin and is a highly aggressive form of lung cancer [64].

Table 2				
Gene symbol	Gene name	Target of Sox2[a]	Sox2 depletion in ES cells [b]	Sox2 overexpression in the lung [c]
Dach2	*dachshund 2 (Drosophila)*	No	Down	Up
Dner	*delta/notch-like EGF-related receptor*	No	Down	Up
Fgf13	*fibroblast growth factor 13*	No	Down	Up
Mycl1	*v-myc myelocytomatosis viral oncogene*	Yes	Down	Up
Oas1a	*2'-5' oligoadenylate synthetase 1A*	Yes	Down	Up
Pfn2	*profilin 2*	No	Down	Up
Serpini1	*serine (or cysteine) peptidase inhibitor, clade I, member 1*	Yes	Down	Up
Slc6a15	*solute carrier family 6 (neurotransmitter*	Yes	Down	Up
Sox21	*SRY-box containing gene 21*	Yes	Down	Up
Uchl1	*ubiquitin carboxy-terminal hydrolase L1*	No	Down	Up
Ptn	*pleiotrophin*	Yes	Down	Up
Mad2l1	*MAD2 (mitotic arrest deficient, homolog)-like*	Yes	Down	Up
Mia1	*melanoma inhibitory activity 1*	Yes	Down	Up
Ddx58	*DEAD (Asp-Glu-Ala-Asp) box polypeptide 58*	Yes	Down	Up
Wdr40b	*WD repeat domain 40B*	No	Down	Up
Bnc2	*basonuclin 2*	Yes	Down	Up
Xrcc5	*X-ray repair complementing defective repair in Chinese hamster cells 5*	Yes	Down	Up
Brunol4	*bruno-like 4, RNA binding protein (Drosophila)*	No	Down	Up
Cdh2	*cadherin 2*	Yes	Down	Up
Frk	*fyn-related kinase*	No	Down	Up
Ms4a6b	*membrane-spanning 4-domains, subfamily A, member 6B*	No	Down	Up
Abca3	*ATP-binding cassette, sub-family A (ABC1), member 3*	No	Up	Down
Car12	*carbonic anyhydrase 12*	No	Up	Down
Gtf2ird2	*GTF2I repeat domain containing 2*	No	Up	Down
Ldlr	*low density lipoprotein receptor*	Yes	Up	Down
Klf7	*Kruppel-like factor 7 (ubiquitous)*	Yes	Up	Down
Qk	*quaking*	Yes	Up	Down
Hkdc1	*hexokinase domain containing 1*	Yes	Up	Down
Tbx1	*T-box 1*	No	Up	Down
Ankrd12	*ankyrin repeat domain 12*	No	Up	Down
Pecam1	*platelet/endothelial cell adhesion molecule 1*	Yes	Up	Down
Kcns3	*rectifier, subfamily S, member 3*	Yes	Up	Down

[a] Genes bound by Sox2 were assigned according to the detection of Sox2 at their promoter by ChIP-seq data showing an association score > 0.1 [12].

[b] Genes regulated by Sox2 in mouse ES cells were derived from Sharov et al. [61].

[c] Genes regulated by Sox2 in lung were derived from **chapter 3**, genes that were at least 1.5-fold up or down regulated after Sox2 overexpression.

Dense-core neurosecretory granules and other neuroendocrine markers are characteristic features of SCLC. Sox2 and Sox21 have been described as immunogenic tumor antigens in patients with neuroendocrine SCLC [64]. This suggests a potential link between the increase in the number of PNECs present in the Sox2 overexpressing lungs (**chapter 3**) and neuroendocrine SCLC. Interestingly, lungs with Sox2 overexpression show a hyperplastic epithelium, similar to a human condition known as bronchiolization of alveoli (BOA), a premalignant lesion that can lead to lung carcinomas [68]. SOX2 overexpression and/or gene-amplification are associated with other cancers [69], such as prostate cancers, gastric carcinogenesis and in basal cell-like breast carcinomas [64, 70]. This is not surprising, since it has been shown that many proteins important during normal development are also implicated in carcinogenesis [69]. Furthermore, several types of poorly differentiated and aggressive human tumors show an ES-like gene expression signature, with overexpression of genes normally enriched in ES cells, including Sox2, Nanog and Oct4 [71]. This would explain the 'stem cell' phenotypic properties of many tumors [71]. For example, Oct4 is normally only expressed in the germ line and its expression is also detected in germ-cell tumors and some somatic tumors. In a recent study, *Oct4* was ectopically expressed in adult mice using a similar doxycycline-dependent expression system as described in **chapter 3**. Expression of *Oct4* blocked progenitor-cell differentiation and causes dysplasia in epithelial tissues [72]. However, little is known about the potential oncogenic properties of Sox2 and, so far there are no specific mutations in SOX2 associated with human cancer or mouse models to demonstrate a direct causal effect of SOX2 in tumorigenesis. The search for possible cancer-causing mutations in SOX2 and the use of the transgenic Sox2 overexpression mice generated in **chapter 3** could clarify if there is a direct causal link between *SOX2* gene expression and cancer. The identification of Sox2 interacting partners and post-translational modifications regulating Sox2 activity may provide potential applications in relation to human cancer.

Concluding remarks

Sox2 is an essential transcription factor for multiple stages of mammalian development and has recently also been used to make ES cell-like induced pluripotent stem (iPS) cells for potential future regenerative therapies. Sox2 mutations cause human congenital malformations and its aberrant expression may be involved in other human pathologies, including cancer. It is therefore important to investigate how Sox2 functions to better understand the consequences of its misregulation. In this thesis, we have investigated how Sox2 and other Sox proteins get to the nucleus of the cell, their place of activity. By isolating putative Sox2 interacting partners, we have provided testable hypotheses on how Sox2 performs its role as a transcriptional regulator and in which regulatory pathways Sox2 is involved. On the level of the organism, we have shown that Sox2 needs to be tightly regulated for lung development to proceed properly and provided ideas for the role of Sox2 in congenital lung malformations. Although these results have increased our knowledge on Sox2, much is still to be clarified. For example, Sox2 expression appears to

correlate with progenitor identity and to counteract differentiation in different organs. How Sox2 performs this cell-specifying activity and whether it is relevant for human pathologies involving *Sox2* misexpression or deletion is currently unclear. Investigating Sox2 conditional knock-out and overexpressing mouse lines for relevant stages of organ development, in combination with a molecular analysis of the misregulated pathways in such mutants, may be able to give us answers on these questions and teach us more about this fascinating transcription factor.

References

1. Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y. & Pallavi, B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol 39, 2195-214 (2007).
2. Kamachi, Y., Uchikawa, M. & Kondoh, H. Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet 16, 182-7 (2000).
3. Wilson, M. & Koopman, P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Genet Dev 12, 441-6 (2002).
4. Avilion, A. A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17, 126-40 (2003).
5. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-6 (1981).
6. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78, 7634-8 (1981).
7. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol 9, 625-35 (2007).
8. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-4 (2007).
9. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372-6 (2000).
10. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567-82 (2008).
11. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-76 (2006).
12. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106-17 (2008).
13. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. Cell 132, 1049-61 (2008).
14. Gorlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol 15, 607-60 (1999).
15. Sudbeck, P. & Scherer, G. Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. J Biol Chem 272, 27848-52 (1997).
16. Kurisaki, A. et al. The mechanism of nuclear export of Smad3 involves exportin 4 and Ran. Mol Cell Biol 26, 1318-32 (2006).
17. Lipowsky, G. et al. Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. Embo J 19, 4362-71 (2000).
18. Sinclair, A. H. et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240-4 (1990).
19. Smith JM, K. P. The ins and outs of transcriptional control: nucleocytoplasmic shuttling in development and disease. Trends Genet. Jan;20(1), 4-8. (2004).
20. Wang, T., Liu, J. H., Yang, J., Chen, J. & Ye, Z. Q. 46, XX male sex reversal syndrome: a case report and review of the genetic basis. Andrologia 41, 59-62 (2009).
21. Berta, P. et al. Genetic evidence equating SRY and the testis-determining factor. Nature 348, 448-50 (1990).
22. Foster, J. W. & Graves, J. A. An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. Proc Natl Acad Sci U S A 91, 1927-31 (1994).

23. Poulat, F. et al. Nuclear localization of the testis determining gene product SRY. J Cell Biol 128, 737-48 (1995).
24. Li, B. et al. SRY-directed DNA bending and human sex reversal: reassessment of a clinical mutation uncovers a global coupling between the HMG box and its tail. J Mol Biol 360, 310-28 (2006).
25. Yasuhara, N. et al. Triggering neural differentiation of ES cells by subtype switching of importin-alpha. Nat Cell Biol 9, 72-9 (2007).
26. Kaffman, A., Rank, N. M., O'Neill, E. M., Huang, L. S. & O'Shea, E. K. The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus. Nature 396, 482-6 (1998).
27. Kaffman, A., Rank, N. M. & O'Shea, E. K. Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. Genes Dev 12, 2673-83 (1998).
28. Gasiorowski, J. Z. & Dean, D. A. Mechanisms of nuclear transport and interventions. Adv Drug Deliv Rev 55, 703-16 (2003).
29. Wegner, M. Secrets to a healthy Sox life: lessons for melanocytes. Pigment Cell Res 18, 74-85 (2005).
30. Tsuruzoe, S. et al. Inhibition of DNA binding of Sox2 by the SUMO conjugation. Biochem Biophys Res Commun 351, 920-6 (2006).
31. Gill, G. SUMO changes Sox for developmental diversity. Mol Cell 20, 495-6 (2005).
32. Taylor, K. M. & Labonne, C. SoxE factors function equivalently during neural crest and inner ear development and their activity is regulated by SUMOylation. Dev Cell 9, 593-603 (2005).
33. Girdwood, D. W., Tatham, M. H. & Hay, R. T. SUMO and transcriptional regulation. Semin Cell Dev Biol 15, 201-10 (2004).
34. Sim, H., Argentaro, A. & Harley, V. R. Boys, girls and shuttling of SRY and SOX9. Trends Endocrinol Metab 19, 213-22 (2008).
35. Thevenet, L. et al. Regulation of human SRY subcellular distribution by its acetylation/deacetylation. Embo J 23, 3336-45 (2004).
36. Nusinzon, I. & Horvath, C. M. Histone deacetylases as transcriptional activators? Role reversal in inducible gene regulation. Sci STKE 2005, re11 (2005).
37. Malki, S. et al. Prostaglandin D2 induces nuclear import of the sex-determining factor SOX9 via its cAMP-PKA phosphorylation. Embo J 24, 1798-809 (2005).
38. Huang, W., Zhou, X., Lefebvre, V. & de Crombrugghe, B. Phosphorylation of SOX9 by cyclic AMP-dependent protein kinase A enhances SOX9's ability to transactivate a Col2a1 chondrocyte-specific enhancer. Mol Cell Biol 20, 4149-58 (2000).
39. Taranova, O. V. et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev 20, 1187-202 (2006).
40. Kiernan, A. E. et al. Sox2 is required for sensory organ development in the mammalian inner ear. Nature 434, 1031-5 (2005).
41. Okubo, T., Pevny, L. H. & Hogan, B. L. Sox2 is required for development of taste bud sensory cells. Genes Dev 20, 2654-9 (2006).
42. Que, J. et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 134, 2521-31 (2007).
43. Gontan, C. et al. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. Dev Biol 317, 296-309 (2008).
44. Zhou, J. et al. Identification of novel mutations and sequence variants in the SOX2 and CHX10 genes in patients with anophthalmia/microphthalmia. Mol Vis 14, 583-92 (2008).
45. Verma, A. S. & Fitzpatrick, D. R. Anophthalmia and microphthalmia. Orphanet J Rare Dis 2, 47 (2007).
46. Ragge, N. K. et al. SOX2 anophthalmia syndrome. Am J Med Genet A 135, 1-7; discussion 8 (2005).
47. Williamson, K. A. et al. Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. Hum Mol Genet 15, 1413-22 (2006).
48. Li, L. & Wei, J. A newborn with anophthalmia and pulmonary hypoplasia (the Matthew-Wood syndrome). Am J Med Genet A 140, 1564-6 (2006).
49. Golzio, C. et al. Matthew-Wood syndrome is caused by truncating mutations in the retinol-binding protein receptor gene STRA6. Am J Hum Genet 80, 1179-87 (2007).
50. Pasutto, F. et al. Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. Am J Hum Genet 80, 550-60 (2007).
51. Kelberman, D. et al. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. J Clin Invest 116, 2442-55 (2006).
52. Wagner, A. J. et al. Genetic analysis of congenital cystic adenomatoid malformation reveals a novel pulmonary gene: fatty acid binding protein-7 (brain type). Pediatr Res 64, 11-6 (2008).

53. Stocker, J. T., Madewell, J. E. & Drake, R. M. Congenital cystic adenomatoid malformation of the lung. Classification and morphologic spectrum. Hum Pathol 8, 155-71 (1977).
54. Wilson, R. D. et al. Cystic adenomatoid malformation of the lung: review of genetics, prenatal diagnosis, and in utero treatment. Am J Med Genet A 140, 151-5 (2006).
55. Fauza, D. O. & Wilson, J. M. Congenital diaphragmatic hernia and associated anomalies: their incidence, identification, and impact on prognosis. J Pediatr Surg 29, 1113-7 (1994).
56. Kravitz, R. M. Congenital malformations of the lung. Pediatr Clin North Am 41, 453-72 (1994).
57. Vieira, J. et al. Bronchopulmonary sequestration--a 12-year experience. Rev Port Pneumol 12, 489-501 (2006).
58. Perl, A. K., Kist, R., Shan, Z., Scherer, G. & Whitsett, J. A. Normal lung development and function after Sox9 inactivation in the respiratory epithelium. Genesis 41, 23-32 (2005).
59. Park, K. S., Wells, J. M., Zorn, A. M., Wert, S. E. & Whitsett, J. A. Sox17 influences the differentiation of respiratory epithelial cells. Dev Biol 294, 192-202 (2006).
60. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-56 (2005).
61. Sharov, A. A. et al. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics 9, 269 (2008).
62. Agathanggelou, A. et al. Identification of novel gene expression targets for the Ras association domain family 1 (RASSF1A) tumor suppressor gene in non-small cell lung cancer and neuroblastoma. Cancer Res 63, 5344-51 (2003).
63. Spinola, M. et al. Ethnic differences in frequencies of gene polymorphisms in the MYCL1 region and modulation of lung cancer patients' survival. Lung Cancer 55, 271-7 (2007).
64. Gure, A. O. et al. Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer. Proc Natl Acad Sci U S A 97, 4198-203 (2000).
65. Jager, R. et al. Serum levels of the angiogenic factor pleiotrophin in relation to disease stage in lung cancer patients. Br J Cancer 86, 858-63 (2002).
66. Sotillo, R. et al. Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. Cancer Cell 11, 9-23 (2007).
67. Lee, M. N. et al. Epigenetic inactivation of the chromosomal stability control genes BRCA1, BRCA2, and XRCC5 in non-small cell lung cancer. Clin Cancer Res 13, 832-8 (2007).
68. Jensen-Taubman, S. M., Steinberg, S. M. & Linnoila, R. I. Bronchiolization of the alveoli in lung cancer: pathology, patterns of differentiation and oncogene expression. Int J Cancer 75, 489-96 (1998).
69. Dong, C., Wilhelm, D. & Koopman, P. Sox genes and cancer. Cytogenet Genome Res 105, 442-7 (2004).
70. Chen, Y. et al. The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. J Biol Chem 283, 17969-78 (2008).
71. Ben-Porath, I. et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 40, 499-507 (2008).
72. Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell 121, 465-77 (2005).

Chapter 7

Summary

Summary

Sox2 is a transcription factor essential for multiple steps in vertebrate development. Recently, Sox2 was also shown to be one of a set of four transcription factors that can reprogram somatic cells into ES cell-like iPS cells. A number of aspects of Sox2 functioning have become increasingly well understood in recent years. On a molecular level, the identification of Sox2 genomic binding sites in ES cells, in combination with the effect of Sox2 depletion on ES cell gene expression, allows for the prediction of Sox2 target genes in ES cells. On a level of the organism, tissue-specific knock-out of Sox2 in mice revealed its function in the development of several organs. However, much is still unclear about how Sox2 does its job. This thesis describes the research on three diverse aspects of Sox2 that are still poorly understood, its function in lung development, its nuclear import and its interaction partners in ES cells.

During lung development Sox2 has a dynamic expression pattern in the airway epithelium. It is highly expressed in non-branching proximal regions and absent from branching distal regions, which suggests that downregulation of Sox2 expression at the branching sites is crucial for correct lung development. In **chapter 3**, we tested this hypothesis by generating transgenic mice that ectopically express Sox2 in the distal airway epithelium in a doxycycline inducible manner. Using these mice, we show that the spatial regulation of Sox2 expression during lung development is critical for the normal branching morphogenesis and the correct differentiation of the pulmonary epithelium. Sox2 transgenic lungs presented an abnormal branching morphogenesis, leading to a severe reduction in the number of airways and causing the formation of enlarged, cyst-like airspaces. In addition, ectopic expression of Sox2 leads to an increase in the number of pulmonary neuroendocrine cells (PNECs) and (pre-) basal cells, which are both committed progenitor cells. Our overexpression study could be complemented by a loss-of-function approach to further understand the role of Sox2 during lung development.

Sox2 is a transcription factor with a low DNA binding affinity, which can be increased by interaction partners. Interaction partners can therefore affect Sox2 function, but besides Oct4, there are no Sox2 interacting proteins known in ES cells. In **chapter 5** we used a protocol based on FLAG-tag affinity to purify Sox2 from ES cells. Mass spectrometry analysis revealed a long list of potential Sox2 binding proteins, including ES cell transcription factors, lineage-specific transcription factors and chromatin-modifying complexes. Recent studies showed that several of the Sox2 interacting transcription factors co-localise with Sox2 on promoters. This suggests that they may assist Sox2 in promoter binding or help regulating the near-by gene. The identification of both activating and repressing Sox2 co-factors may suggest that Sox2 can perform both transcriptional activation and repression. Future research will have to reveal how important these Sox2 interacting factors are for gene regulation by Sox2.

In **chapter 4** we investigated the interaction of Sox2 with Exportin 4 (Exp4), one of the most prominent interacting partners of Sox2. So far, Exp4 was only known as a nuclear export receptor. Using *in vitro* import assays on permeabilized HeLa and ES cells, we showed that Exp4 acts as a nuclear import receptor for Sox2. As the Sox2 import signals are conserved within the Sox protein family, this import pathway may also be relevant

for other Sox proteins. Indeed, we found that Exp4 could also mediate import for SRY, the archetypical Sox protein, which is involved in male sex determination. Interestingly, Exp4-mediated import is disrupted by several mutations in SRY that cause sex reversal in humans. Recently, a study suggested that several importin α (Impα) isoforms could import Sox2. We tested all the Impα isoforms and also other members of the importin ß (Impß) superfamily for their Sox2 import capacity. We found that none of the Impα isoforms could import Sox2, while of the Impß superfamily only Imp9 and the Impß/7 combination showed Sox2 import capacity. In conclusion, we show that Sox2 can use three separate nuclear import pathways, mediated by Exp4, Impß/7 and Imp9. This makes it unlikely that Sox2 import is regulated at the level of the import receptor, as previous literature suggested.

Samenvatting

Sox2 is een transcriptiefactor die essentieel is voor verschillende stappen in de ontwikkeling van vertebraten. Recentelijk is ook aangetoond dat Sox2 één van een set van vier transcriptiefactoren is die somatische cellen kan reprogrammeren naar ES cel-achtige iPS cellen. Een aantal aspecten van het functioneren van Sox2 is duidelijker geworden in recente jaren. Op een moleculair niveau kunnen Sox2-gereguleerde genen voorspeld worden in ES cellen door de identificatie van Sox2 bindings plaatsen op het ES cell genoom, in combinatie met het effect van Sox2 depletie op genexpressie. Op het niveau van het organisme, heeft tissue-specifieke deletie van het Sox2 gen aangetoond dat Sox2 een functie heeft in de ontwikkeling van vele organen. Er is echter ook nog veel onduidelijk over hoe Sox2 zijn werk doet. Dit proefschrift beschrijft onderzoek naar drie uiteenlopende aspecten van Sox2 die nog slecht begrepen zijn; zijn functie in longontwikkeling, zijn import naar de cel kern en zijn interactiepartners in ES cellen.

Sox2 heeft een dynamisch expressiepatroon in het epithelium van de luchtwegen gedurende de longontwikkeling. Het komt hoog tot expressie in de niet-vertakte proximale delen en is niet aanwezig in de vertakkende distale delen, hetgeen suggereert dat het verlagen van de Sox2 expressie op de vertakkingspunten essentieel is voor een correcte longontwikkeling. In **hoofdstuk 3** hebben wij deze hypothese getest door transgene muizen te maken waarbij Sox2 tot expressie gebracht kan worden in het distale luchtwegepithelium met behulp van doxycycline. Met deze muizen konden we aantonen dat het specifieke expressiepatroon van Sox2 gedurende longontwikkeling belangrijk is een normale vertakkingsmorphologie en de correcte differentiatie van het longepithelium. Sox2 transgene longen hebben een abnormaal vertakkingspatroon hetgeen leidde tot een sterke afname van het aantal luchtwegen en de formatie van vergrote cysteachtige luchtzakken. Ectopische expressie van Sox2 leidde ook tot een groter aantal neuroendocriene cellen en basale cellen in de longen, die beiden gespecialiseerde progenitor cellen zijn. Om de rol van Sox2 in longontwikkeling better te begrijpen zou onze overexpressie studie kunnen worden gecomplementeerd met een aanpak waarbij het Sox2 gen specifiek gedeleteerd wordt.

Sox2 is een transcriptiefactor met een lage affiniteit voor DNA, die kan worden verhoogd met de hulp van interactiepartners. Interactiepartners kunnen daarmee een invloed hebben op het functioneren van Sox2, maar behalve Oct4 zijn er geen Sox2 interactiepartners bekend in ES cellen. In **hoofdstuk 5** gebruikten we een protocol gebaseerd op affiniteit voor de FLAG-tag om Sox2 op te zuiveren uit ES cellen. Analyse met de massa-spectrometer gaf ons een lange lijst met potentiële Sox2-bindende eitwitten, waaronder ES cell transcriptiefactoren, weefsel-specifieke transcriptiefactoren en chromatine-modificerende eiwitten. Recente studies hebben laten zien dat een aantal van deze Sox2-bindende eiwitten samen met Sox2 binden aan promotoren van genen. Dit suggereert dat ze Sox2 kunnen helpen bij het binden van promotoren en de regulatie van de bijbehorende genen. De identificatie zowel activerende als represserende Sox2 co-factoren suggereert dat Sox2 zowel een transcriptionele activator als repressor kan zijn. Onderzoek in de toekomst zal moeten uitwijzen hoe belangrijk deze Sox2-interacterende factoren zijn voor genregulatie door Sox2.

In **hoofdstuk 4** onderzochten we de interactie van Sox2 met Exportin 4 (Exp4), één van de meest prominente interactiepartners van Sox2. Tot dusverre was Exp4 alleen bekend als een factor die export uit de celkern verricht. Met behulp van *in vitro* import assays op gepermeabiliseerde HeLa cellen en ES cellen konden we laten zien dat Exp4 een import-receptor voor Sox2 is. Aangezien de import-signaal sequenties van Sox2 geconserveerd zijn in de gehele familie van Sox eiwitten, zou deze import route ook relevant kunnen zijn voor andere Sox eiwitten. We konden inderdaad laten zien dat Exp4 ook SRY, het eerst ontdekte Sox eiwit en essentieel voor de geslachtsontwikkeling bij mannen, kan transporten naar de kern. Het is interessant te zien dat Exp4-gemedieerde import wordt verstoord door een aantal mutaties in SRY die de mannelijke geslachtsontwikkeling verstoren. In een recente studie wordt het gesuggereerd dat Sox2 door verschillende importin α (Impα) receptoren zou kunnen worden geimporteerd naar de kern. Wij hebben alle Impα receptoren en ook receptoren van de importin ß (Impß) familie getest. We vonden dat geen van de Impα receptoren Sox2 kon importeren terwijl van de Impß familie alleen Imp9 en de Impß/7 combinatie Sox2-import activiteit bezitten. We concluderen dat Sox2 gebruik kan maken van drie aparte import routes, verzorgd door Exp4, Impß/7 en Imp9. Dit maakt het onwaarschijnlijk dat Sox2 import gereguleerd wordt op het niveau van de import receptor, zoals eerder in de literatuur gesuggereerd was.

Curriculum vitae

Name: Maria Cristina Gontan Pardo
Born: 30 September 1976, Madrid, Spain

Education

1994-1998	B.Sc. Biological Sciences. Complutense University of Madrid, Spain
1998-2001	M.Sc. Genetics. Complutense University of Madrid, Spain
2001-2002	Post-graduate program in Molecular Biology and Biochemistry, Complutense University of Madrid, Spain
2002-2003	M.Sc. Biotechnology. Center for Molecular Biology, UAM-CSIC, Spain

Research and professional experience

2009- present	Postdoctoral fellow at the Department of Reproduction and Development, Erasmus MC, Rotterdam, the Netherlands (Dr. J. Gribnau)
2007	Visiting scientist at the Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany (Prof.dr. D. Görlich)
2003-2008	PhD student at the Departments of Pediatric Surgery and Cell Biology, Erasmus MC, Rotterdam, the Netherlands (Prof.dr. D. Tibboel, Prof.dr. F.G. Grosveld and Dr. R.J. Rottier)
2002-2003	Graduate student at the Department of Molecular Endocrinology, Instituto de Investigaciones Biomedicas Alberto Sols, UAM-CSIC, Spain (Dr. M.J. Obregon)
1999-2000	Undergraduate student at the Laboratory of General Botany and Nature Management. Erasmus/Socrates fellowship. Vrije University Brussels, Belgium (Dr. L. Triest)

List of publications

Gontan, C.*, Güttler, T.*, Engelen, E., Demmers, J., Fornerod, M., Grosveld, F., Tibboel, D., Dirk Görlich, Raymond A. Poot, Robbert Rottier.
'Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors' *Journal of Cell Biology, in press,* **2009**.

Gontan, C., de Munck, A., Vermeij, M., Grosveld, F., Tibboel, D. and Rottier, R. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation.
Developmental Biology, 1; 317(1): 296-309, **2008**.

(* equal author contribution)

Erasmus MC
Universitair Medisch Centrum Rotterdam

PhD Portfolio Summary

Summary of PhD training and teaching activities

Name PhD student: Cristina Gontan Pardo Erasmus MC Department: Pediatric Surgery Research School: Postgraduate School Molecular Medicine	PhD period: August 2003 – May 2008 Promotor(s): Prof.dr. D. Tibboel Prof.dr. F.G. Grosveld Supervisor: Dr. R.J. Rottier

1. PhD training

	Year
General academic skills	
- Radiation Protection Course (level 5b). Rotterdam	2005
- Laboratory animal science (Artikel 9). Utrecht	2004
In-depth courses	
- In Vivo Imaging "From Molecule to Organism". Rotterdam	2007
- Analysis of microarray gene expression data. Rotterdam	2007
- Cells into Organ Summer School. Basel, Switzerland	2006
- From Development to Disease. Rotterdam	2005
- Transgenesis and Gene Targeting. Leiden	2005
- Signal Transduction and control of cellular proliferation. Leiden	2004
Presentations	
- 18th MGC-Symposium, September, Rotterdam	2008
- Winter School of the International Graduiertenkolleg "Transcriptional Control In Developmental Processes". Kleinwalsertal, Austria	2005, 2007,2008
International conferences	
- International Symposium "Stem Cells, Development and Regulation". Amsterdam (poster)	2008
- Society for Developmental Biology 67th Annual Meeting. University of Philadelphia, Pennsylvania, USA (poster)	2008
Seminars and workshops	
- Dutch Stem Cell Meeting. Groningen, The Netherlands (poster)	2008
- EuroSTELLS, Exploring Chromatin in Stem Cells Workshop. Montpellier, France (poster)	2007
- British Societies for Cell and Developmental Biology (BSDB). York, UK	2006
- 3rd Dutch Chromatin Community meeting. Rolduc,The Netherlands	2005

Acknowledgements

It has been more than five years since I started my PhD and the thesis that is now in your hands would not have been possible without the help of many people.

I would like to start by thanking my two promoters Prof. Dick Tibboel and Prof. Frank Grosveld for giving me the opportunity to study and work for my PhD thesis at the Erasmus MC in Rotterdam. I also would like to express my gratitude to Robbert Rottier, my copromoter, for supporting me as a PhD student in his group, I have learned a lot during these years.

During my PhD period I had very fruitful collaborations. From the Department of Cell Biology, I especially want to thank Raymond Poot without whom chapters 4 and 5 of this thesis would not have been possible. Raymond, thank you for teaching me biochemistry, one of the most useful tools I learned during my PhD. Your brilliant ideas and scientific rigor are an example for me.

I am thankful to Maarten Fornerod, from the NKI in Amsterdam, for his decisive advice and reagents at the beginning of the Exportin 4 story and for being in my thesis committee.

I acknowledge Dirk Görlich for having me in his lab for three months and for all his intelligent suggestions. Thomas, it was a real pleasure and a lot of fun working with you, thanks for teaching me so many things about protein purification. Finally the cows came home! Laura, I appreciated very much your advice with the import assays and your friendship. Thanks to all the members of Dirk's lab for making my stay in Göttingen so pleasant.

From the Josephine Nefkens Institute, I would like to acknowledge Marcel Vermeij for *sharing his expertise* in immunohistochemistry and Ronald de Krijger for his helpful advice with the histological sections. My thanks also goes to Frank van der Panne for taking the time to explain me how to make optimal use of Photoshop to make the best possible figures.

I want to express my gratitude to Joost Gribnau for giving me the opportunity to continue my scientific career in Rotterdam and being in my reading committee. Joost, thank you for giving me the time I needed to finish my thesis. Sjaak Philipsen and Dies Meijer, thank you for being in my reading committee.

I would like to thank the members (past and present) of lab 10.34, all the girls from Robbert's lab and people from Bob's group. A special mention goes to my dear friend and colleague Prapapan for her quiet understanding; I really miss you in the office. Pascal, I had a lot of fun with your hilarious way of complaining and taking care of our fish. Luna, I am very honored that you are my Paranimf, I wish you all the best with your PhD! I also would like to thank Ton de Wit for his cloning advice at the beginning of my PhD and for sharing all his holiday stories and pictures.

I really enjoyed my time in such an international department, not only on the scientific aspects but also on a personal level. I was very lucky to have many good colleagues and friends and I would like to thank them all for their help and support during my thesis. Unfortunately, scientists do not stay very long in the same place and many of them are already gone. The positive thing is that you can visit your friends all around the world. I did visit some of you already but I promise I will visit the rest once my thesis is finished and I have more time.

Helen, thank you for your friendship, you are a wonderful person (our night with Robbie was one of the most magical moments I had in the Netherlands). I really miss you and I am very happy that you are coming to my thesis defense. Athina, good luck with your thesis, I am very, very, pleased that you are going to stay in The Netherlands. Suzanne, I hope you can also make it for my defense; otherwise we will go to see you soon in Freiburg. Sanja, I enjoyed our coffee breaks and conversations, I really appreciate that you delayed your departure from Holland to be at my thesis defense. Diana, many people went away but some new people came and I am really happy that you are one of them.

Thanks to all the people in the 10th floor for creating such a nice working atmosphere, Jeffrey, Noori, Ekim, Ana, Susana, Gina, Frank, Daniel and many others….I always enjoyed our chats in the corridors.

Nothing would have been the same with out the Spanish "family" in Rotterdam: Javi (my big brother), Maria, Empar (always so creative), Begona, Laura, Jennifer (not really Spanish but very integrated), Elena, David, Pablo, Oscar and Lourdes. I really had a lot of fun with all of you and I miss our late lunches. I wish you could all be here sharing my promotion day with me but I am happy we are still in contact over distance. I miss you a lot!

Living for four years with two of my Paranimfen was one of the best things that happened to me in Rotterdam. Muchachas, Hanne and Magda, you really made me feel at home and you were always there in the good and not so good moments. It was an honor to be your paranimf and I am really pleased that you are doing the same job for me. Barbara, you were not so long in the house, but I have to say that it was very good fun when you were there. Mi Rute, so many years since Brussels…we could not imagine we will be still in the Low Lands, obrigada por tu amistad y apoyo. Sonia, obrigada for all the coffee breaks on Fridays when I was writing my thesis at home.

I also want to mention my friends from Spain that even in the distance contributed to this thesis by always supporting me. Barbara, mucho mas que una amiga, gracias por venir a mi tesis con la familia Gontan. Mayte gracias por estar siempre ahi y dejarme participar de tu familia. David, gracias por ser mi amigo y consejero. Laura, gracias por entenderme y compartir nuestras neuras. Pili, cuando te vienes a vivir a Holanda? Nada me haria mas feliz. Belen, Elena, Alicia, Maria, Ana, Jacobo, Adolfo, Amaya, Pedro,…y

todos los que no puedo menciaonar por falta de espacio, GRACIAS!

Finally, I would like to thank my family, Gracias a mis padres y hermanas (Laura, Esther y Elena) porque siempre me han apoyado y han confiado en mi. Siento vivir tan lejos, pero ya sabeis que a pesar de la distancia os quiero mucho. Teneis que venir mas veces a Holanda!. I also would like to remember Ria, I know you wanted to be here today and we really miss you, thank you for being such a fascinating schoonmoeder.

Raymond, you make me happy.

Cristina

Appendix

Color Figures

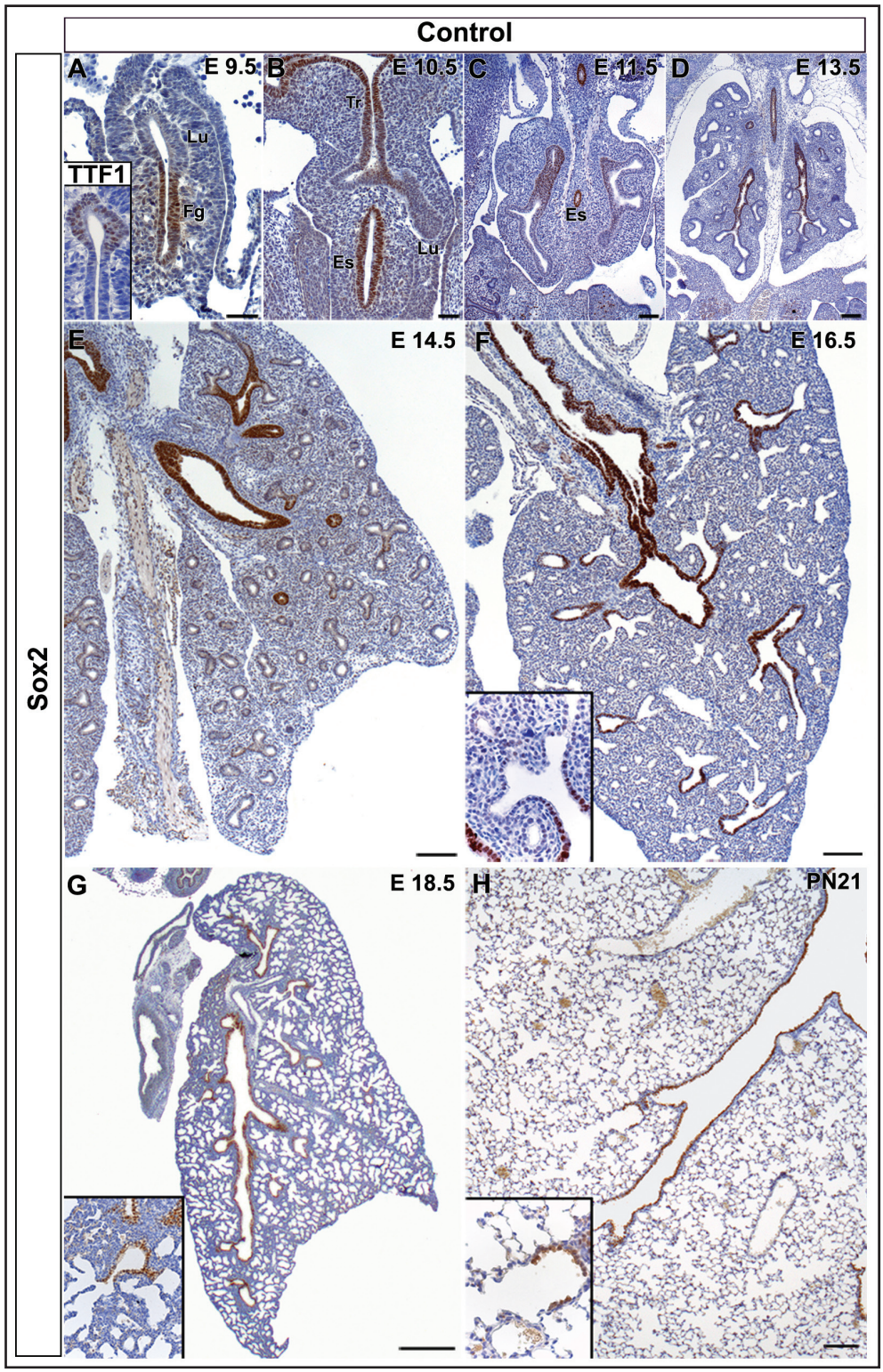

Control
Sox2
A
E 9.5
TTF1
Lu
Fg
B
E 10.5
Tr
Es
Lu
C
E 11.5
Es
D
E 13.5
E
E 14.5
F
E 16.5
G
E 18.5
H
PN21

Fig. 1 page 69

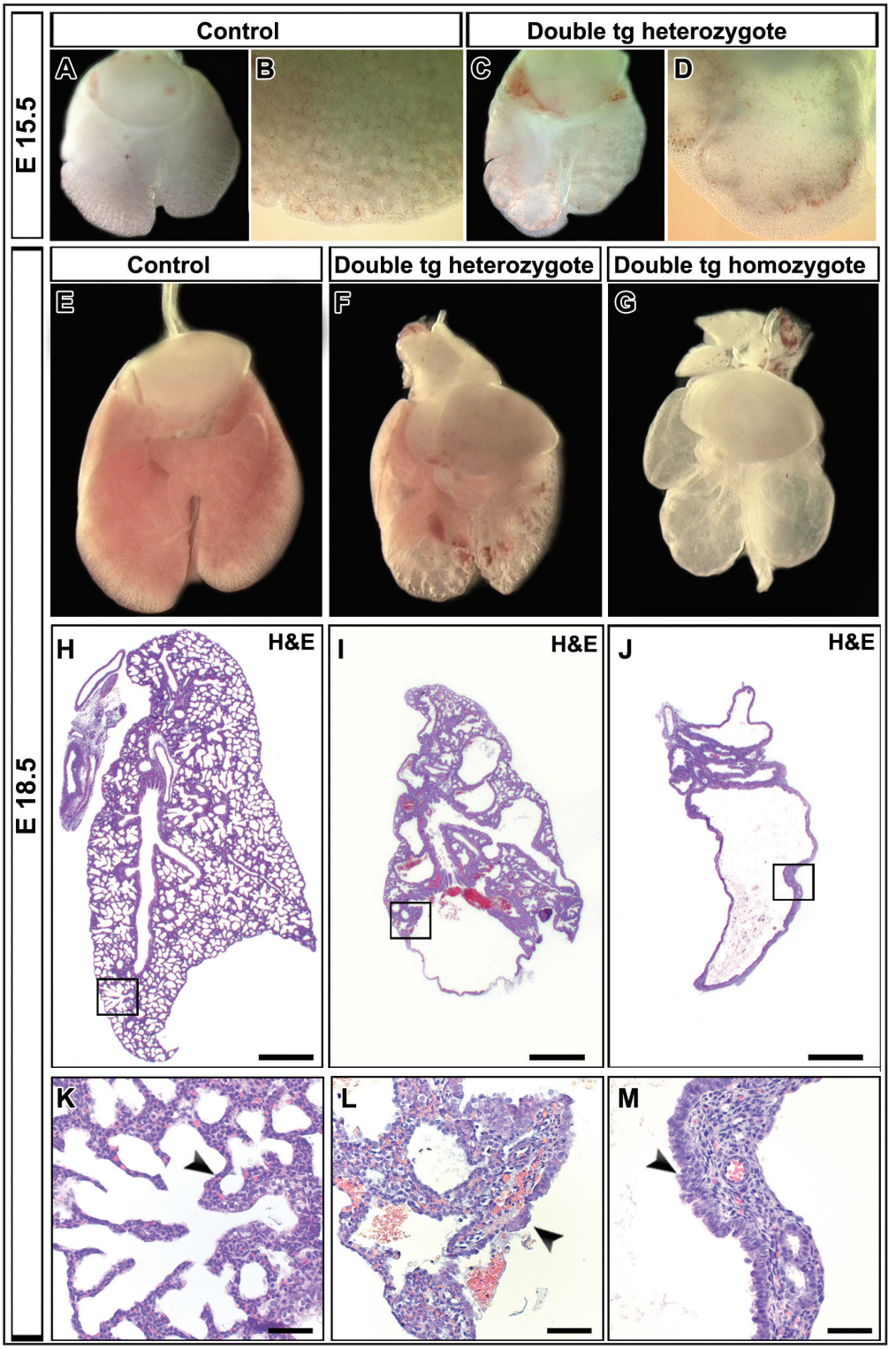

E 15.5
Control
Double tg heterozygote
A
B
C
D
E 18.5
Control
Double tg heterozygote
Double tg homozygote
E
F
G
H
H&E
I
H&E
J
H&E
K
L
M

Fig. 2 page 70

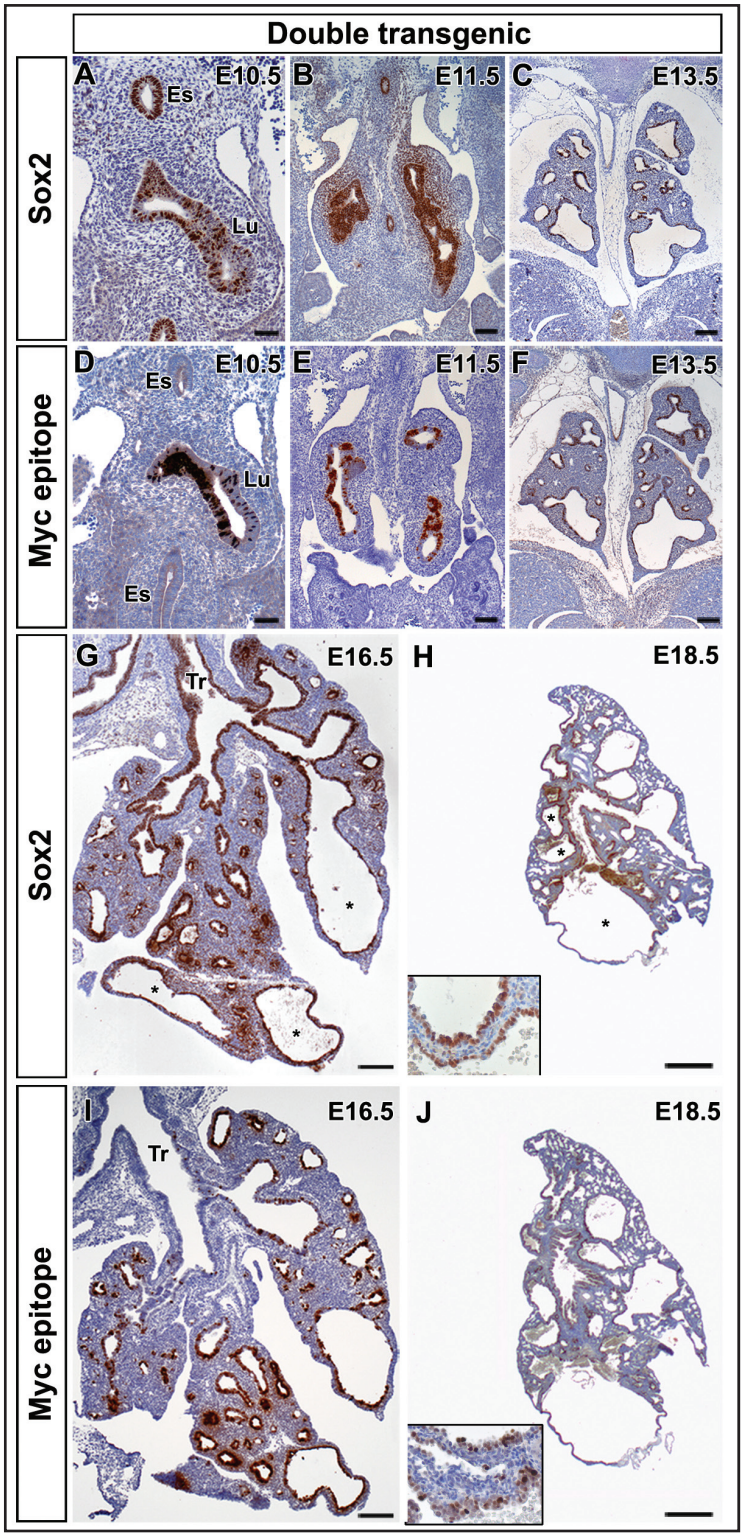

Double transgenic
Sox2
Myc epitope
Sox2
Myc epitope
A
E10.5
Es
Lu
B
E11.5
C
E13.5
D
E10.5
Es
Lu
Es
E
E11.5
F
E13.5
G
E16.5
Tr
H
E18.5
I
E16.5
Tr
J
E18.5

Fig. 3 page 71

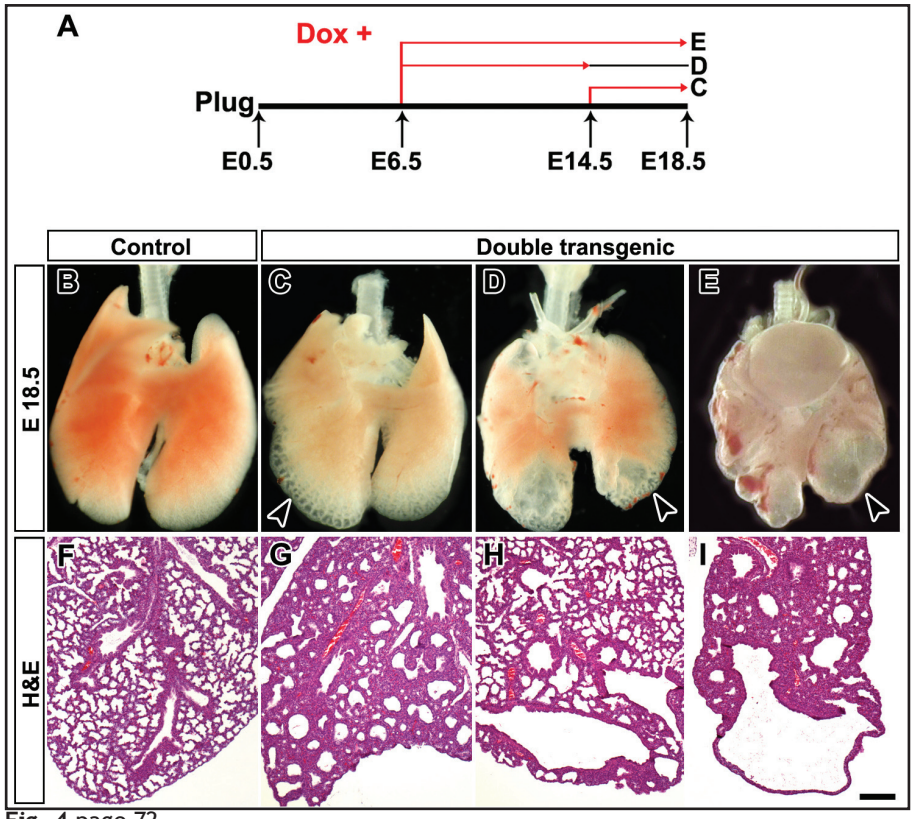

A
Dox +
E
D
C
Plug
E0.5
E6.5
E14.5
E18.5
Control
Double transgenic
E 18.5
H&E
B
C
D
E
F
G
H
I

Fig. 4 page 72

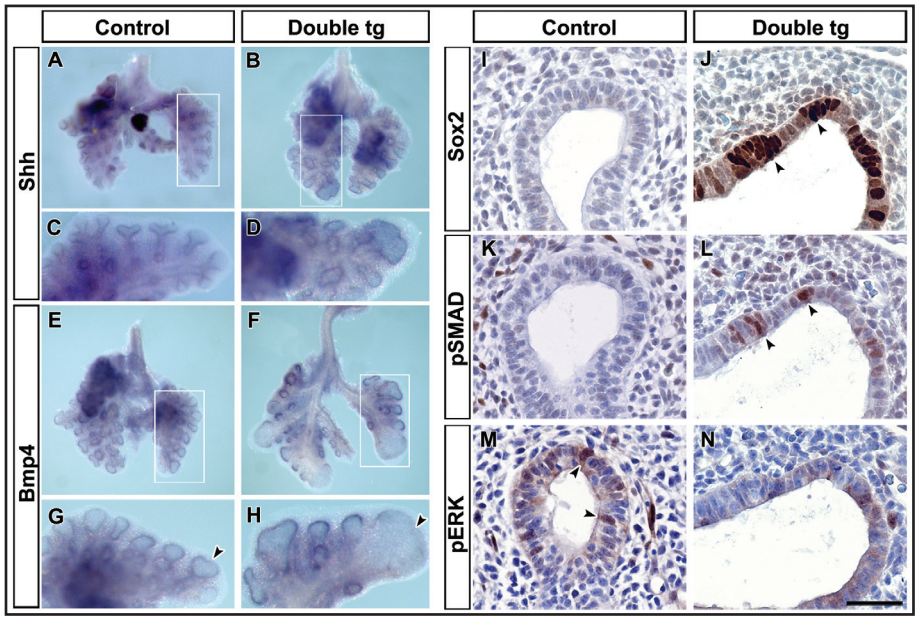

Control
Double tg
Control
Double tg
Shh
Bmp4
Sox2
pSMAD
pERK
A
B
C
D
E
F
G
H
I
J
K
L
M
N

Fig. 5 page 72

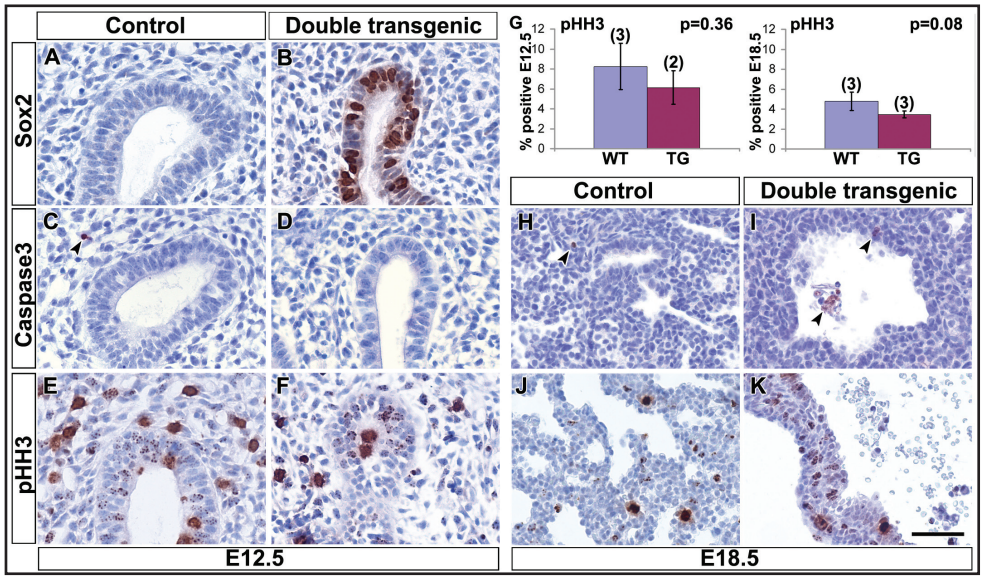

Control
Double transgenic
Sox2
Caspase3
pHH3
A
B
C
D
E
F
G
pHH3
p=0.36
% positive E12.5
(3)
(2)
WT
TG
pHH3
p=0.08
% positive E18.5
(3)
(3)
WT
TG
Control
Double transgenic
H
I
J
K
E12.5
E18.5

Fig. 6 page 73

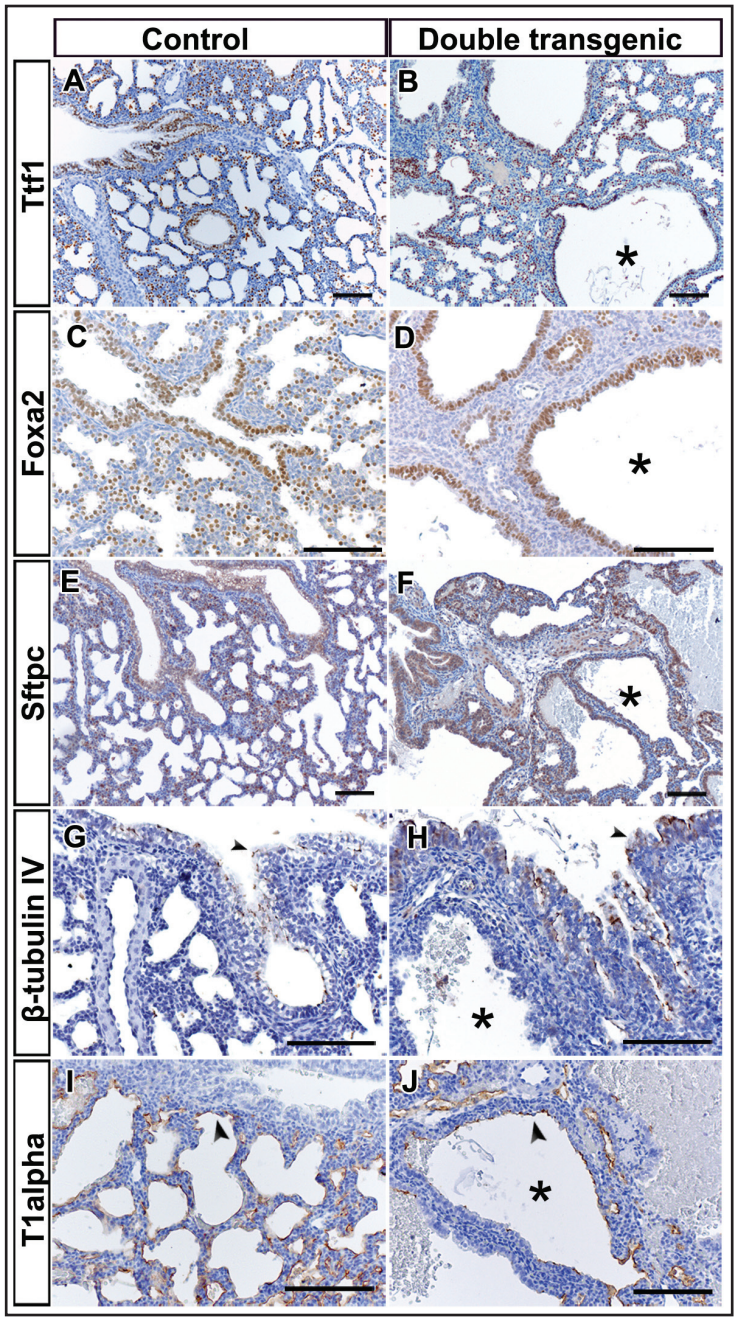

Control
Double transgenic
Ttf1
Foxa2
Sftpc
β-tubulin IV
T1alpha
A
B
C
D
E
F
G
H
I
J

Fig. 7 page 74

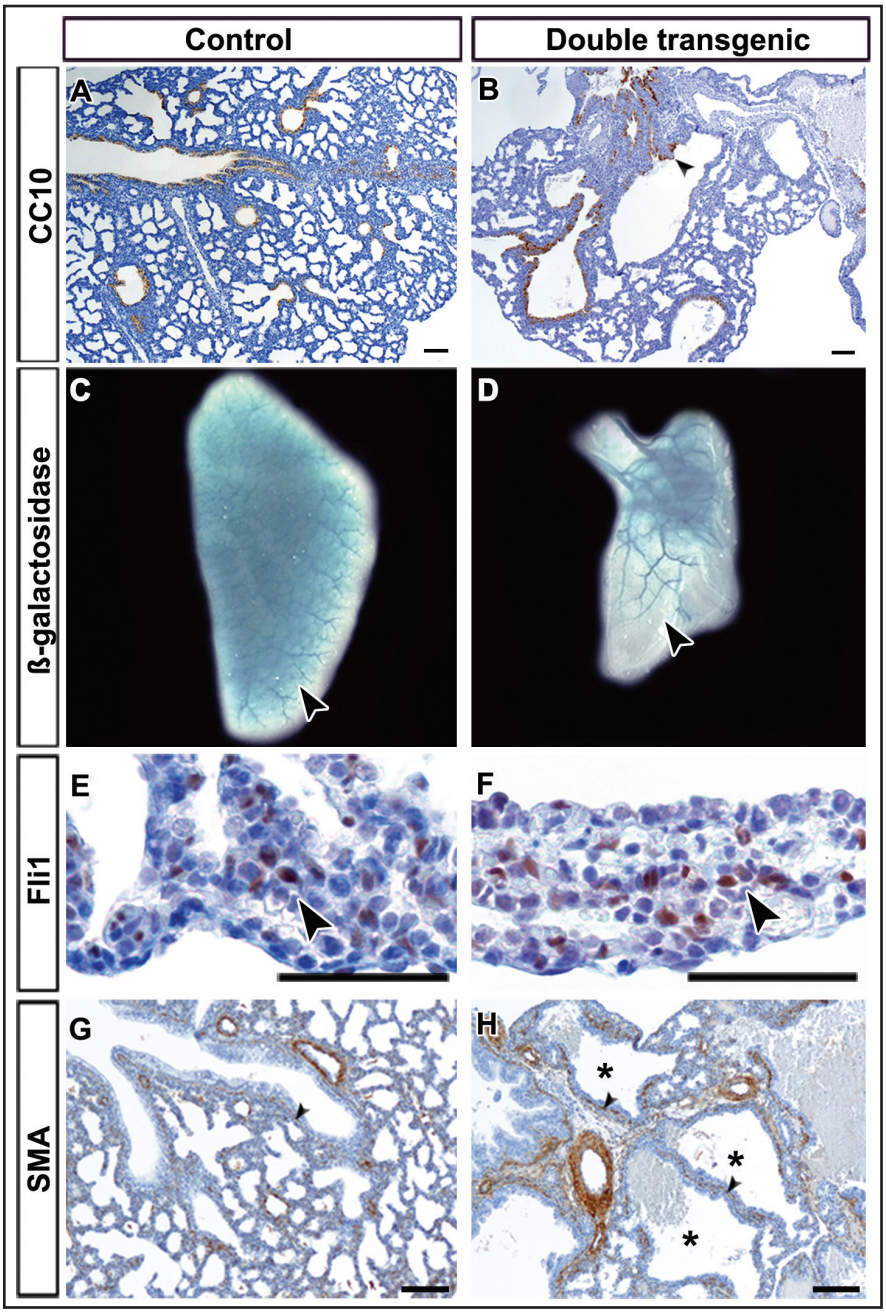

Control
Double transgenic
CC10
ß-galactosidase
Fli1
SMA
A
B
C
D
E
F
G
H

Fig. 8 page 75

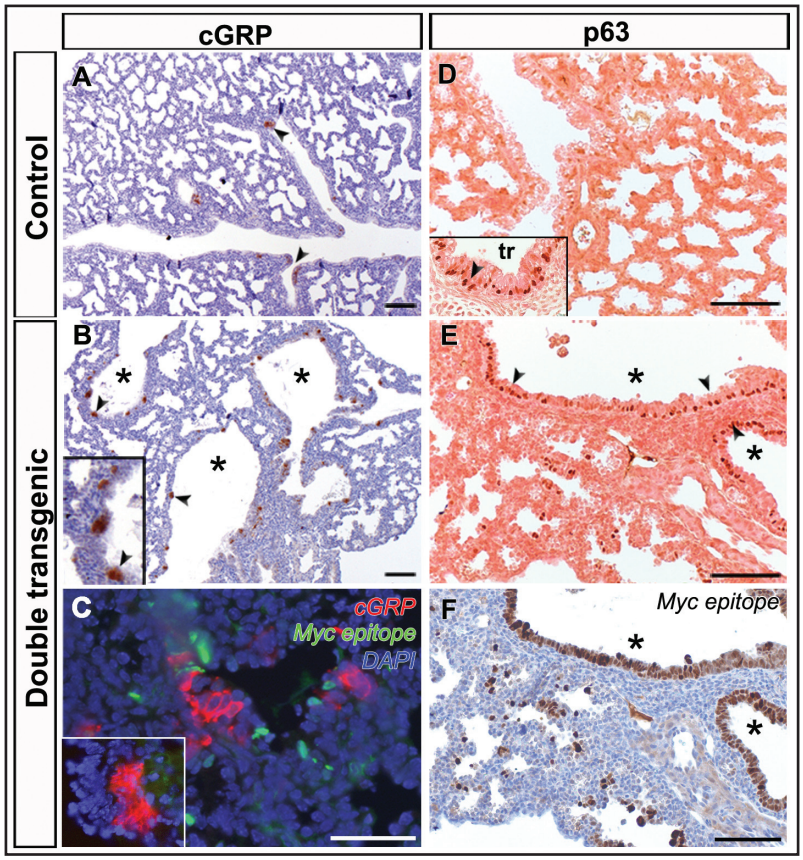

cGRP
p63
Control
Double transgenic
A
D
tr
B
E
C
cGRP
Myc epitope
DAPI
F
Myc epitope

Fig. 9 page 76

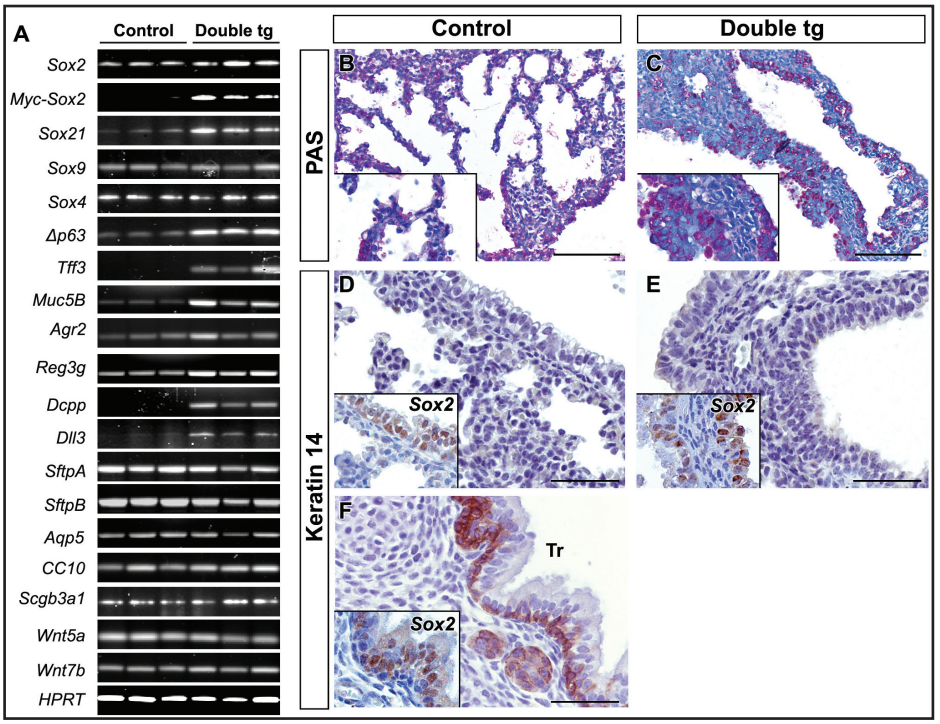


Fig. 10 page 77

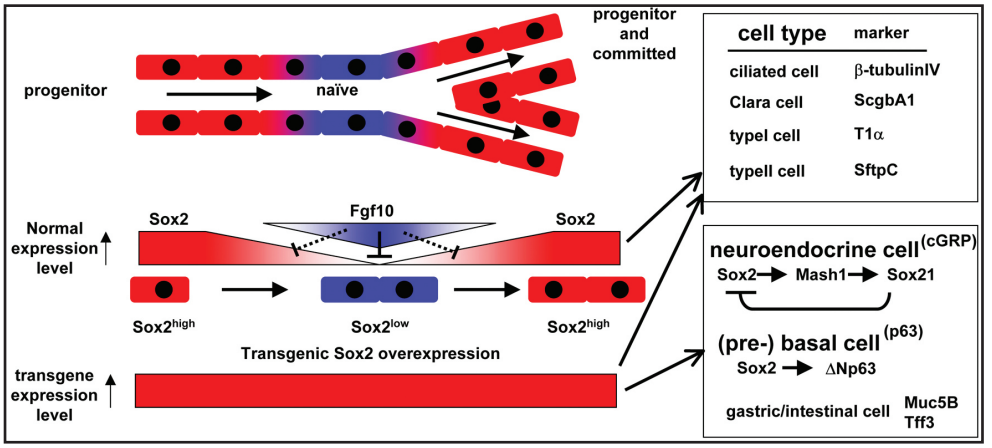


Fig. 11 page 78

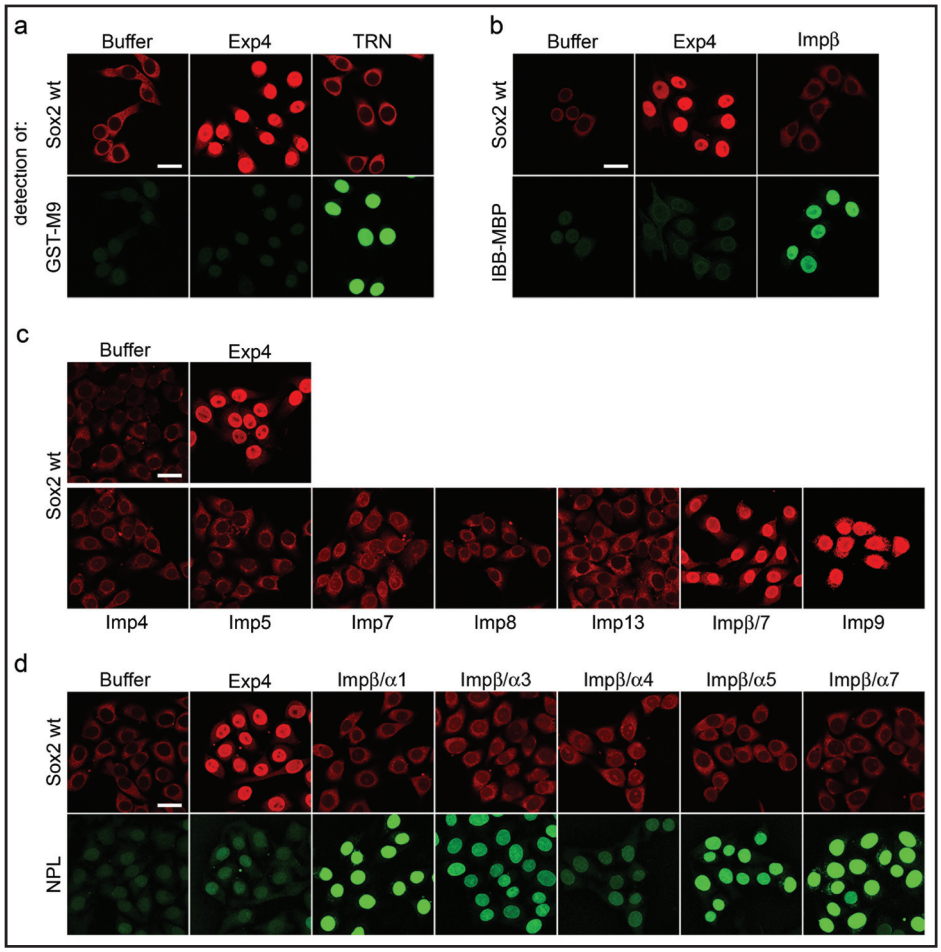

a
Buffer
Exp4
TRN
detection of:
Sox2 wt
GST-M9
b
Buffer
Exp4
Impβ
Sox2 wt
IBB-MBP
c
Buffer
Exp4
Sox2 wt
Imp4
Imp5
Imp7
Imp8
Imp13
Impβ/7
Imp9
d
Buffer
Exp4
Impβ/α1
Impβ/α3
Impβ/α4
Impβ/α5
Impβ/α7
Sox2 wt
NPL

Fig. 2 page 87

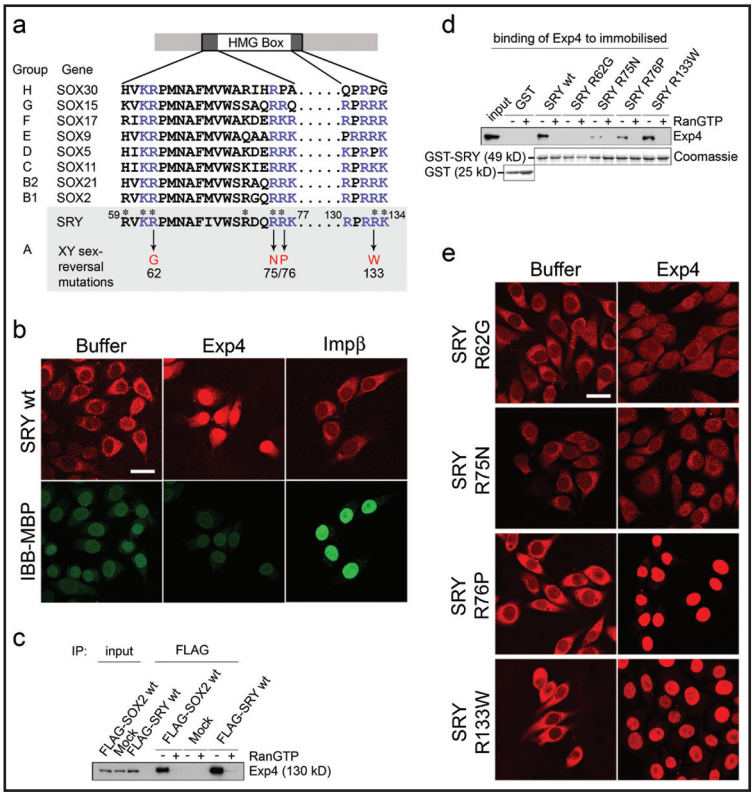

a
HMG Box
Group Gene
H SOX30 HVKRPMNAFMVWARIHRPA.....QPRPG
G SOX15 KVKRPMNAFMVWSSAQRRQ.....RPRRK
F SOX17 RIRRPMNAFMVWAKDERKR.....RPRRR
E SOX9 HVKRPMNAFMVWAQAARRK.....PRRRK
D SOX5 HIKRPMNAFMVWAKDERRK.....KPRPK
C SOX11 HIKRPMNAFMVWSKIERRK.....RPRKK
B2 SOX21 HVKRPMNAFMVWSRAQRRK.....RPRRK
B1 SOX2 RVKRPMNAFMVWSRGQRRK.....RPRRK
SRY RVKRPMNAFIVWSRDQRRK.....RPRRK
A XY sex-reversal mutations
G 62
NP 75/76
W 133
b
Buffer
Exp4
Impβ
SRY wt
IBB-MBP
c
IP: input FLAG
FLAG-SOX2 wt
Mock
FLAG-SRY wt
FLAG-SOX2 wt
Mock
FLAG-SRY wt
RanGTP
Exp4 (130 kD)
d
binding of Exp4 to immobilised
input
GST
SRY wt
SRY R62G
SRY R75N
SRY R76P
SRY R133W
RanGTP
Exp4
GST-SRY (49 kD)
GST (25 kD)
Coomassie
e
Buffer
Exp4
SRY R62G
SRY R75N
SRY R76P
SRY R133W

Fig. 3 page 88

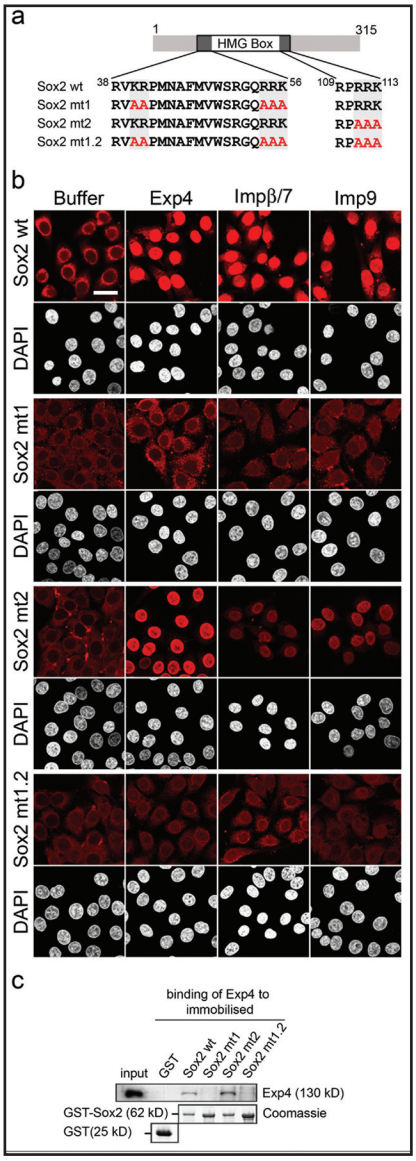

a
1
315
HMG Box
Sox2 wt 38 RVKRPMNAFMVWSRGQRRK 56 108 RPRRK 113
Sox2 mt1 RVAAPMNAFMVWSRGQAAA RPRRK
Sox2 mt2 RVKRPMNAFMVWSRGQRRK RPAAA
Sox2 mt1.2 RVAAPMNAFMVWSRGQAAA RPAAA
b
Buffer
Exp4
Impβ/7
Imp9
Sox2 wt
DAPI
Sox2 mt1
DAPI
Sox2 mt2
DAPI
Sox2 mt1.2
DAPI
c
binding of Exp4 to immobilised
input
GST
Sox2 wt
Sox2 mt1
Sox2 mt2
Sox2 mt1.2
Exp4 (130 kD)
GST-Sox2 (62 kD)
Coomassie
GST(25 kD)

Fig. 4 page 90

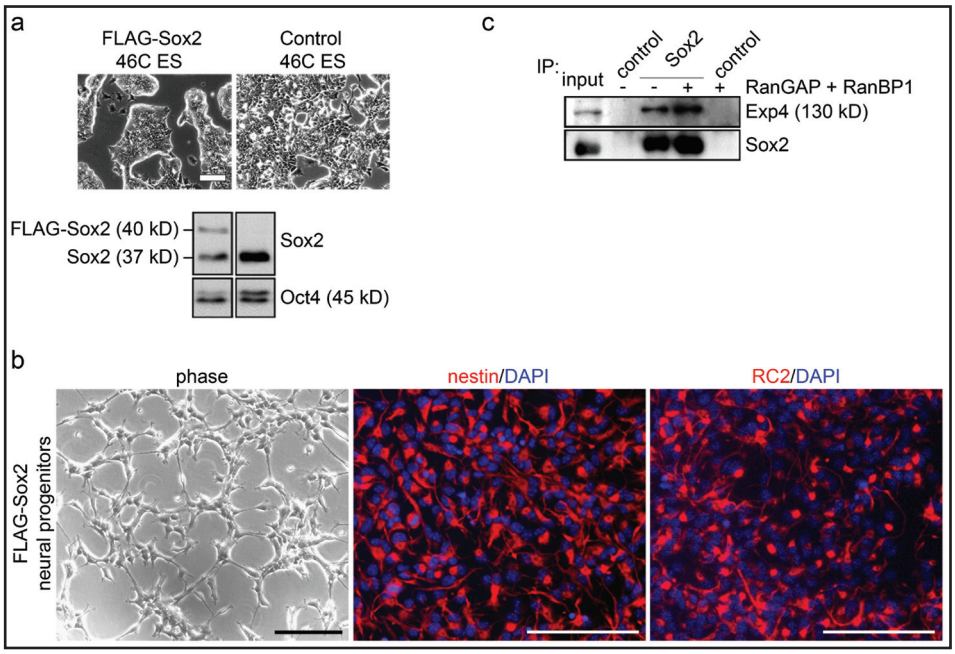


Fig. S1. page 96

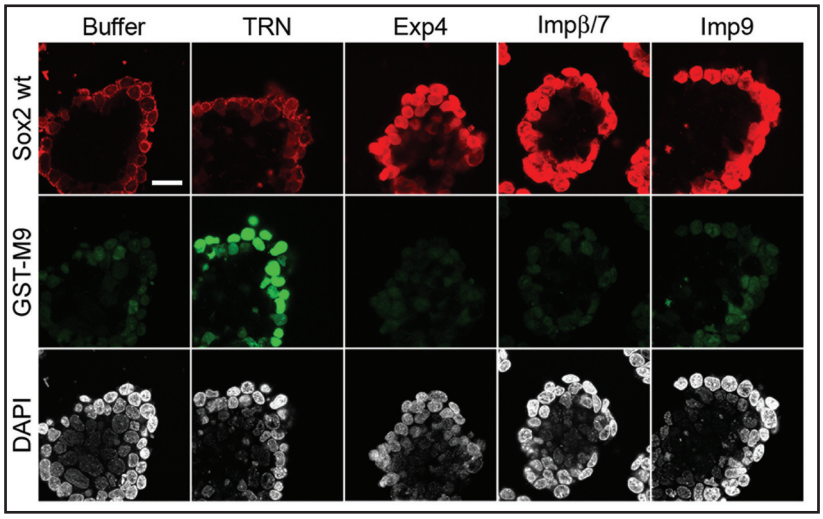


Fig. S3. page 98

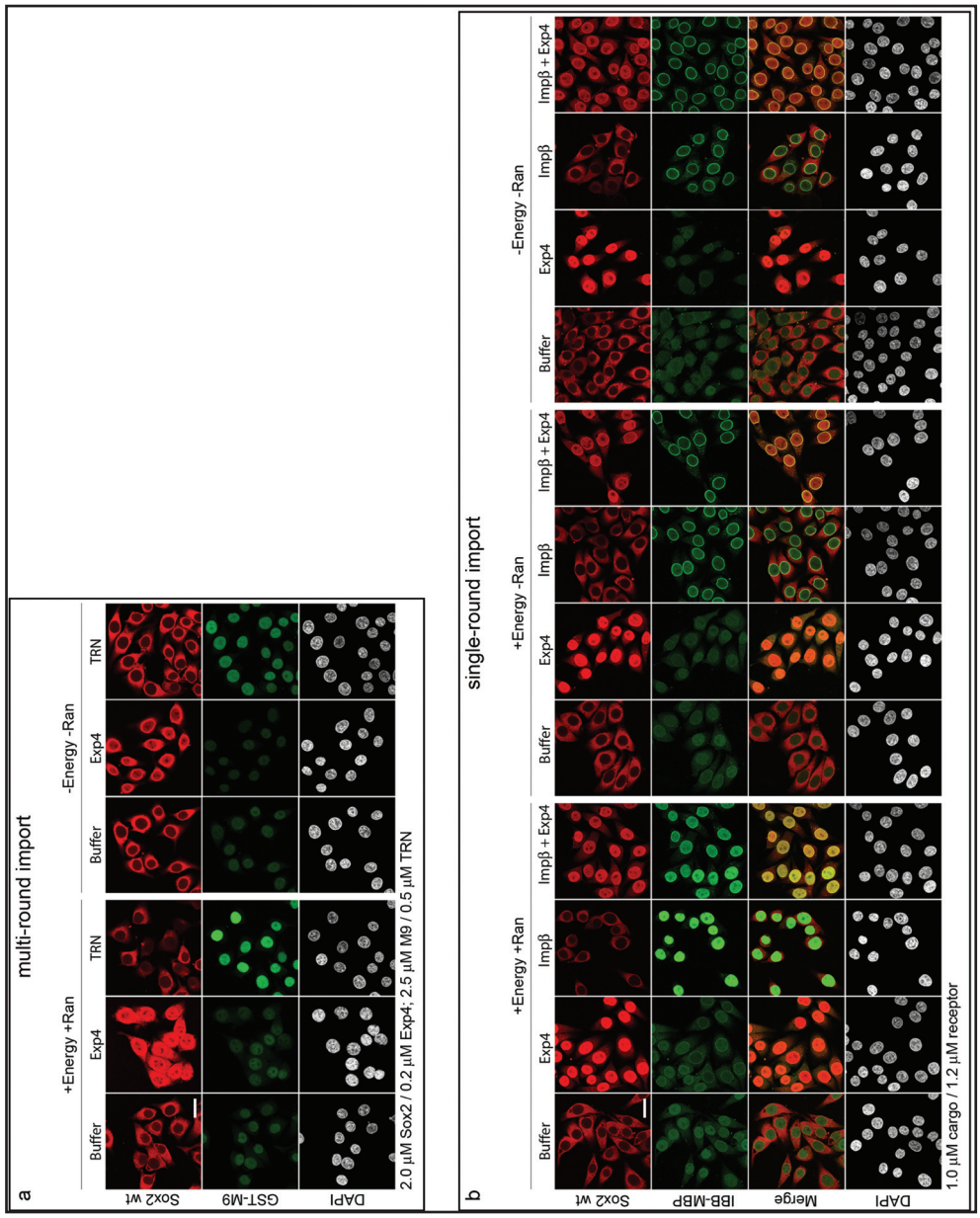


Fig. S2. page 97

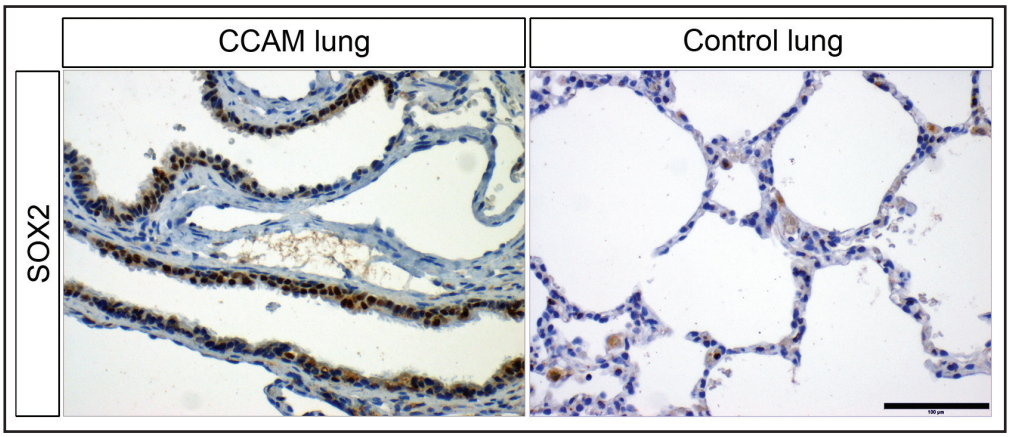


Fig. 4. page 137