THE UNIVERSITY
of EDINBURGH

Dissecting lineage specification in EpiSC and neuromesodermal progenitor cultures

Eleni Pavlina Karagianni

THE UNIVERSITY *of* EDINBURGH

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

Institute for Stem Cell Research, MRC Centre for Regenerative Medicine

School of Biological Sciences, University of Edinburgh

September, 2016

Declaration

I declare that the work presented in this thesis is my own, unless otherwise stated. The work described in this thesis has not been submitted for any other degree or professional qualification.

Eleni Pavlina Karagianni

Abstract

During mouse embryo gastrulation, the pluripotent epiblast gives rise to the three embryonic germ layers, the ectoderm, mesoderm and endoderm. After somitogenesis begins and pluripotency disappears from the epiblast, bipotent neuromesodermal progenitors (NMPs) drive axis elongation, contributing to the formation of the posterior nervous system, as well as the axial and paraxial mesoderm. Early NMPs arise in the E8.5 mouse embryo, in and near the primitive streak, while late NMPs are found in the tail bud (E9.5 - E13.5). NMP regions are characterized by coexpression of *Tbra* (Brachyury) and *Sox2*. Sox1, another neural related transcription factor, has also been detected in NMP regions. Importantly, it has been shown that *Sox1* expression increases as NMPs transit from the primitive streak to the tail bud stages.

Mouse epiblast derived stem cells (EpiSCs) recapitulate the properties of the post-implantation epiblast and therefore serve as a good *in vitro* system for the study of early lineage specification events. EpiSCs express pluripotency factors and early differentiation markers, including *Sox2*, *Sox1 and Tbra*. Based on studies reporting that EpiSC cultures contain distinct subpopulations that have progressed further into lineage specification, I analyzed the properties of the *Tbra* expressing EpiSCs and by dissecting their expression profile, I assess whether these cells are pluripotent or they have progressed further into lineage specification, possibly into an NM fate. I show that EpiSC cultures include a large fraction of Tbra/Sox2 double positive cells; however, Nanog expression was detected in the vast majority of $Tbra^{+}/Sox2^{+}$ EpiSCs suggesting that most of the $Tbra^{+}$ cells are pluripotent rather than bipotent NMPs. Using a previously published Tbra-GFP reporter cell line, I present that Tbra-GFP^{+} cells constitute a dynamic fraction of the culture that has not exited pluripotency (as shown by expression of the pluripotency markers), but have adopted an early primitive streak-like character. Similar to the cells of the posterior epiblast, these EpiSCs are in a reversible state and they retain their ability to undergo neural differentiation. In contrast to the overlap of Tbra and Sox2 positivity in self-renewing EpiSCs, it has been

shown that *Tbra* expression is mutually exclusive with expression of *Sox1-GFP*, that seems to mark a distinct subpopulation with neural-like characteristics.

In vitro NMPs can be generated from EpiSCs upon treatment with Fgf2 and the Gsk-3 antagonist/Wnt agonist CHIRON99021 (FGF/CHI). In these conditions, 80% of the culture becomes $Tbra^{+}/Sox2^{+}$. Given that Sox1 is present in NMP regions *in vivo*, I hypothesized that the NMP cultures could contain $Tbra^{+}Sox1^{+}$ NM bipotent cells. Most importantly, the upregulation of *Sox1* at the tail bud stages drove the hypothesis that *Sox1* expression could mark the transition from an early- to a late-like NMP state *in vitro*. In this study, using a Sox1-GFP reporter cell line, I show that Tbra/Sox2/Sox1-GFP triple positive cells emerge in FGF/CHI treated EpiSCs. Importantly, Sox1-GFP^{+} cells express NMP markers and are enriched in transcripts of *Hox* genes. The expression profile of Sox1-GFP^{+} cells resembles the alteration of *Hox* gene activation that takes place in the caudal progenitor regions during the transition from early NMPs (E8.5) to late NMPs (E9.5-10.5) and hence supports the hypothesis that Sox1-GFP marks NMPs that correspond to the axial progenitors found at tail bud stages. Although the gene activity observed in the Sox1-GFP^{+} subpopulation correlates with the NM developmental potential, these cells exhibit strong neurogenic capacity, while evidence for their ability to give rise to mesoderm differentiation products is still lacking.

Since *Tbra* and *Sox1/Sox2* are not expressed in NMP regions exclusively, but also in mesoderm and neural fated tissues respectively, double rather than single reporter cell lines would be more suitable tools for tracking and isolating bipotent NM progenitors *in vivo* and *in vitro*. Here, I present the CRISPR/Cas9-mediated generation of a reliable Tbra-GFP reporter ES cell line that in contrast to the one published before, contains both endogenous *Tbra* loci intact. By targeting the *Sox2* locus in the Tbra-GFP ES cells, I generated a Tbra-GFP/Sox2-tdTomato double reporter ES cell line, that in the future, could help us to dissect the molecular mechanisms underlying the self-renewal and differentiation of NMPs.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Prof. Valerie Wilson and Dr. Anestis Tsakiridis for their guidance and help throughout my PhD. Their passion and dedication to science has been inspiring for me. I would also like to thank the members of my committee, Dr. Sally Lowell and Prof. Ian Chambers for their suggestions. I am grateful to the Medical Research Council (MRC) for providing my PhD scholarship.

I would like to extend my appreciation to all the past and present members of the Wilson's group for making my time in the lab an invaluable experience. A big thank you to Frederick Wong, for his contribution to my project and for being always available to answer my questions no matter how late at night it was. Special thanks to Yali Huang for performing grafting experiments for me and in this way letting me in the magic world of embryos. I would like to say a big thank you to Julia Watson for her great support and all the fun we had talking about science outside the lab. I am also really grateful to Filip Wymeersch (my great Lord Russell *party* Place flatmate) and Ron Wilkie for their help and advice on embryology. I would also like to thank Claire Cryer and Fiona Rossi for their help with the FACS analysis and Bertrand Vernay for his help with microscopy. A big thanks also to Suresh Kaushik, for his suggestions and his help with gene targeting.

During the last four years, I had the luck to meet great people who supported me during hard times in and out the lab. A special thanks goes to Eleni, "the real one", for becoming my family in Edinburgh. A big thank you to Luca and Jing for being my best friends and a great companion in the lab. I would also like to thank Matina and Yolanda for being my library buddies during the endless days of thesis writing. See you in "Morocco" girls! I cannot find words to express my gratitude to a special friend who made me smile whenever my cells made me feel grumpy and took me to trips around the world when writing about those cells made feel stressed. I have learnt so much from you…*Danke!*

Last but not least, I would like to say a big thank you to my family, for keeping safe for me a small paradise on the other side of Europe. These people and that place have been the source of the strength I needed for the journeys I have done so far...

'Ενα μεγάλο ευχαριστώ στην οικογένεια μου, που μου κρατά φυλαγμένο ένα μικρό παράδεισο στην άλλη άκρη της Ευρώπης. Αυτοί οι άνθρωποι και αυτό το μέρος είναι η πηγή της δύναμης που χρειάστηκα για όσα ταξίδια έχω κάνει μέχρι τώρα. Σας ευχαριστώ...

Abbreviations

A	
A – P	Anterior – posterior
AVE	Anterior visceral endoderm
B	
Blasto	Blasticidin
C	
Cas9	CRISPR-associated protein-9 nuclease
Cas9n	Cas9 nickase
CHI	CHIRON99021
CLE	Caudal lateral epiblast
CNH	Chordoneural hinge
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
D	
DVE	Distal visceral endoderm
E	
EB	Embryoid bodies
EMT	Epithelial-to-mesenchymal transition
EpiSCs	Epiblast-derived stem cells
ES cells	Embryonic stem cells

F	
FACS	Fluorescence-activated cell sorting
FGF/CHI	Fgf2/ CHIRON99021
FRT	Flippase recombinase recognition target
G	
G418	Neomycin/geneticin
gDNA	Genomic DNA
GFP	Green fluorescent protein
GSG spacer	Glycine-serine-glycine spacer
H	
H2B	Histone 2B
HRM assay	High resolution melting assay
HSV	Herpes simplex virus
I	
ICM	Inner cell mass
IRES	Internal ribosome entry site
L	
L1-5	Lateral 1-5
LIF	Leukemia inhibitory factor
M	
M	Mesoderm
MSCV	Murine stem cell virus
N	

N	Neuroectoderm
NLS	Nuclear localisation signal
NMPs	Neuromesodermal progenitors
NSB	Node streak border
	O
ORF	Open reading frame
	P
Pac	Puromycin N-acetyltransferase
PCR	Polymerase chain reaction
P-D	Proximal – distal
PGK	Phosphoglycerate kinase 1
	Q
QRT-PCR	Quantitative real-time PCR
	R
RA	Retinoic acid
RED	dsRed2
	S
SAG	Shh agonist
sgRNA	Single guide RNA
SRR2	Sox2 regulatory region 2
	T
T2A	*Thosea asigna* virus 2A
tdT	Tandem dimmer Tomato

TK	Thymidine kinase
	U
UTR	Untranslated region
	V
VE	Visceral endoderm

Table of contents

List of Figures

List of Tables

Chapter 1: General Introduction

1.1. Pluripotency and lineage segregation during pre-implantation mouse embryo development

Mouse embryo development initiates with the generation of a single totipotent cell, the zygote and unfolds through progressive restriction of the developmental potential and cell differentiation. Following three symmetric divisions (E2.5), the zygote gives rise to 8 blastomeres, all of which are totipotent, exhibiting the capacity to generate all of the embryonic and extraembryonic tissues (Tarkowski and Wróblewska, 1967). At this stage, the blastomeres develop intercellular junctions and flatten against one another, forming a rough sphere, the morula (E3.0). Asymmetrical cell divisions at the morula stage (16-32 cells) generates two visibly distinct subpopulations: polarized cells in the outer surface of the morula, that will become allocated to the trophectoderm and apolar inner cells, that will form the inner cell mass (ICM) (reviewed in Saiz and Plusa, 2013). Hippo signalling plays a key role in determining ICM versus trophectoderm identity. In outer blastomeres of the morula, the Hippo pathway is inactive. As a result, the Yes-associated protein (YAP) is present in the nucleus promoting the expression of trophectoderm specific genes, including the *Caudal type homeobox 2* (*Cdx2*). In inner cells however, Hippo signalling is active, restricting YAP in the cytoplasm. In this way, expression of trophectoderm genes is replaced by transcriptional activation of genes such as the *Octamer-Binding Protein 4* (*Oct4*), that is essential in defining the ICM (Nishioka *et al.*, 2009; Ralston *et al.*, 2010).

Around the 32-cell stage, the mouse embryo cavitates and the morula develops into an early blastocyst. ICM cells derived from early blastocysts are capable of forming trophectoderm, whereas cells isolated from fully expanded blastocysts have lost their plasticity and are unable to contribute to this lineage (reviewed in Rossant *et al.*, 2003). Around E4.0, the embryo hatches out of the zona pellucida and the ICM segregates into two lineages: the pre-implantation epiblast and a layer of cells that overlies the

epiblast, the primitive endoderm (figure 1.1). Contribution of ICM cells of E3.5 embryos to both primitive endoderm and epiblast lineages is rare and cells exhibit strong bias towards one or other lineage. At that stage, the ICM exhibits a mosaic "salt and pepper" pattern, consisting of cells expressing either *Nanog* or the *GATA-binding protein 6* (*Gata6*), committed to become epiblast and primitive endoderm, respectively (Chazaud *et al.*, 2006).

The trophectoderm gives rise to the extraembryonic ectoderm and the ectoplacental cone, both of whom form the placenta. The primitive endoderm differentiates into the parietal endoderm, that contacts the maternal tissue, and the visceral endoderm, that remains in contact with the embryo and gives rise to the endoderm of the visceral yolk sac. The pre-implantation epiblast retains pluripotency and constitutes the origin of the somatic tissues, the germ cell lineage and the extra-embryonic mesoderm of the embryo (reviewed in Arnold and Robertson, 2009).

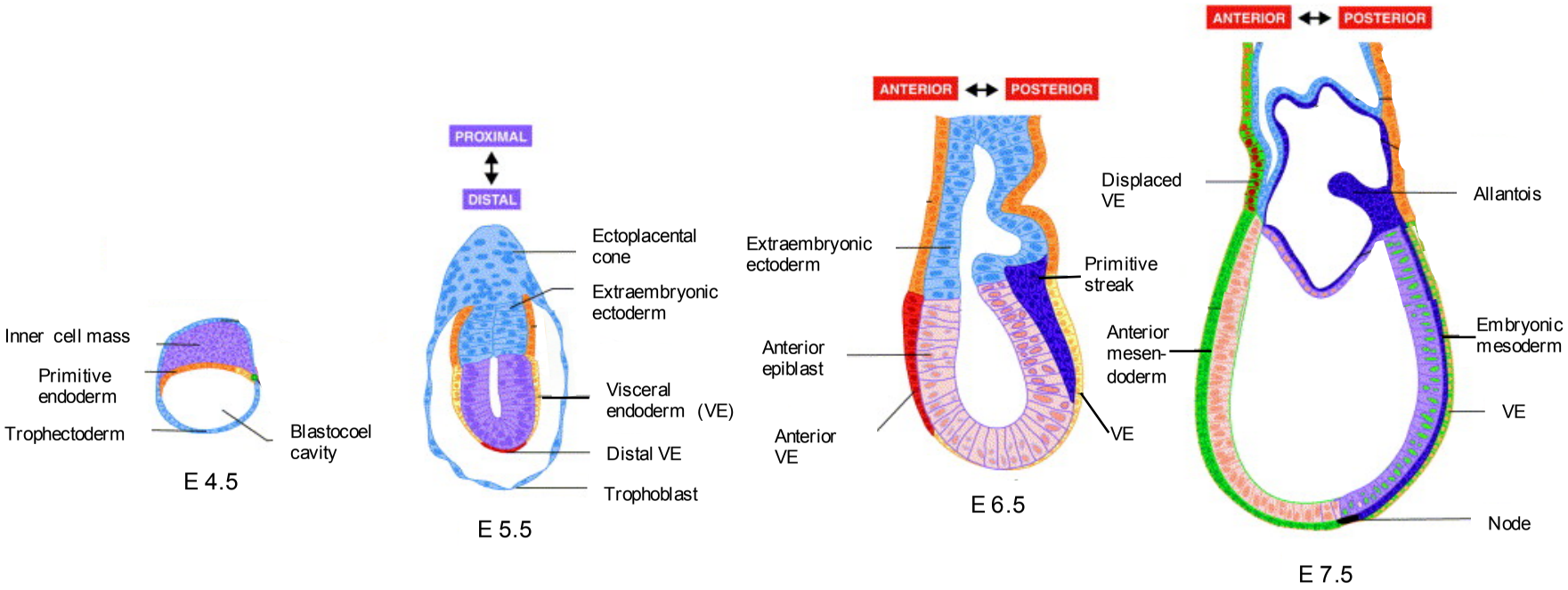


Figure 1.1: Mouse embryo development from E4.5 to E7.5. Development of embryonic and extra-embryonic tissues from the pre-implantation blastocyst (E4.5) to the early bud (E7.5) gastrula. Modified from Lu *et al.* (2001).

1.2. Pluripotency and lineage specification during mouse gastrulation

Following implantation (around E5.0), the mouse embryo goes through dramatic morphological changes, elongating along the proximal-distal (P-D) axis and forming the egg cylinder (figure 1.1). Starting at E5, Nodal signalling, that initially is present all over the epiblast, induces expression of the Wnt and Nodal inhibitors *Dickkopf-related protein 1* (*Dkk1*), *Left-right determination factor 1* (*Lefty1*) and *Cerberus 1* (*Cer1*), in the distal visceral endoderm (DVE). Thus, the DVE acts as the first specialised signalling centre that creates a gradient of Nodal activity and in this way, it establishes P-D polarity in the egg cylinder. Canonical Wnt signalling is involved in the specification of DVE and is required for maintenance of the P-D axis, whereas Bone morphogenetic proteins (Bmp) signalling from the extraemryonic ectoderm inhibits the formation of the DVE. At E6, the DVE migrates to the prospective anterior side of the embryo, contributing to the formation of the anterior visceral endoderm (AVE) (figure 1.1). The emergence of the AVE establishes the anterior-posterior (A-P) axis and generates an A-P gradient of Nodal and Wnt signals in the epiblast (reviewed in Arnold and Robertson, 2009; Arkell and Tam, 2012).

At around E6.5, the primitive streak appears opposite to the AVE, in the posterior part of the epiblast, marking the initiation of gastrulation that will give rise to the three embryonic germ layers, mesoderm, endoderm and ectoderm (figure 1.1) During gastrulation, the primitive streak extends to the distal part of the embryo, while posterior epiblast cells undergo epithelial-to-mesenchymal transition (EMT) as they ingress through the primitive streak and migrate to form nascent mesoderm and definitive endoderm. Nodal, Wnt and Bmp signalling play a key role in regulating the initiation of gastrulation. *Nodal*$^{-/-}$ mice are unable to form primitive streak and consequently fail to gastrulate (Camus *et al.*, 2006), while *Cer1*$^{-/-}$; *Lefty1*$^{-/-}$ (Nodal repressors) double mutant mice display expansion or ectopic primitive streak (Perea-Gomez *et al.*, 2002). Similarly, *Wnt3*$^{-/-}$ embryos arrest at the egg cylinder stage and fail to form a primitive streak or generate mesoderm and endoderm (Liu *et al.*, 1999). Crosstalk between Nodal signalling in the epiblast, Bmp4 signalling in the

extraembryonic ectoderm and Wnt3 in the posterior epiblast creates a positive feedback loop that increases Nodal signalling in the proximal posterior epiblast and initiates the formation of the primitive streak (Ben-Haim *et al.*, 2006). This explains the observation that *Bmp4*$^{-/-}$ mice fail to express the early primitive streak marker *Brachyury* (*Tbra*) in the proximal epiblast and showed little or no mesoderm differentiation (Winnier *et al.*, 1995).

The time and site of ingression through the streak determines the fate of distinct cell lineages. At first, epiblast cells migrate through the streak in the most posterior part of the epiblast and in response to Bmp4 signalling from the extraembryonic ectoderm, they give rise to extraembryonic mesoderm, including the mesodermal layer of the chorion, the visceral yolk sac mesoderm and blood islands. Cells that slightly later migrate through the intermediate and anterior levels of the streak form the lateral plate, paraxial and cardiac mesoderm, while cells that migrate through the extreme anterior tip of the primitive streak give rise to the prechordal plate, the notochord and the node, as well as the definitive endoderm lineage (Kinder *et al.*, 1999; Kinder *et al.*, 2001). Endoderm cells, characterised by flattened morphology and expression of the marker *Forkhead box protein A2* (*Foxa2*), seem to move through mesoderm cells, that appear rounder and Tbra positive, until they reach the superficial layer of the forming endoderm (Burtscher and Lickert, 2009). Tbra^{+} presumptive mesoderm and Foxa2^{+} presumptive axial mesoderm and endoderm cell populations are already specified in the epiblast, before they ingress through the primitive streak (Burtscher and Lickert, 2009). Although the two germ layers are mostly generated by these two distinct cell types, clonal analysis has revealed the existence of common bipotent progenitors that contribute to a smaller extent to mesoderm and endoderm formation (Lawson *et al.*, 1991; Tzouanacou *et al.*, 2009). The *Fibroblast growth factor 8* (*Fgf8*) and the *Fibroblast growth factor receptor 1* (*Fgfr1*), which at that stage are expressed in the primitive streak, play a crucial role in regulating the EMT and the migration of cells away from the primitive streak. In *Fgf8*$^{-/-}$ embryos, cells ingress at the streak but fail to migrate, therefore mesoderm and endoderm are not formed (Sun *et al.*, 1999). Fgfr1 has been shown to regulate EMT (Ciruna and Rossant, 2001); this is consistent with earlier studies that reported accumulation of cells in the posterior streak and reduction

in the paraxial mesoderm formation upon homozygous deletion of *Fgfr1* (Deng *et al.*, 1994; Yamaguchi *et al.*, 1994).

Cells that do not ingress through the primitive streak but remain in the epiblast form the ectoderm, which gives rise to the surface ectoderm and the neural tissues. Retrospective clonal analysis studies (see section 1.3.1) have shown that the anterior and posterior nervous system have a common origin but become clonally separate by late gastrulation (Mathis and Nicolas, 2000; Forlani *et al.*, 2003; Cajal *et al.*, 2012). Grafting experiments in E7.5 embryos have demonstrated that the anterior regions of the epiblast give rise to the neuroectoderm of the prosencephalon and mesencephalon, while cells in the distal-lateral regions flanking the rostral end of the primitive streak form the rhombencephalon (Tam, 1989). The neuroectoderm of the spinal cord originates in the ectoderm overlying the node region and lateral to the anterior and the middle regions of the primitive streak (named caudal later epiblast -CLE) (Tam, 1989).

In summary, interaction between embryonic and extraembryonic tissues create gradients of different signalling pathways in the epiblast, which consequently dictate the fate of epibast cells. This regionalization of cell fate is reflected by expression of early lineage specification markers, such as the neural marker *Orthodenticle homeobox 2* (*Otx2*), in the anterior epiblast and the mesodermal markers *Caudal type homeobox 2* (*Cdx2*) and *Eomesodermin* (*Eomes*) in the posterior epiblast. However, this regionalisation is not representative of the developmental potential of the epiblast cells. When anterior and distal parts of the E8.0 epiblast were grafted beneath the testis capsule, teratocarcinomas were generated from both sides, showing that pluripotency is present in multiple locations in the post-implantation epiblast (Beddington, 1983). In addition, it has been shown that it is possible to establish epiblast stem cell (EpiSC) lines from multiple parts of the epiblast (Brons *et al.*, 2007; Tesar *et al.*, 2007; Chenoweth *et al.*, 2010; Osorno *et al.*, 2012). Hence, although epiblast cells have the potential to form all three germ layers, their fate is determined by their position on the post-implantation epiblast. Pluripotency disappears from the epiblast at the beginning

of somitogenesis, which coincides with extinction of the pluripotency factor Nanog, but before Oct4 becomes undetectable (Osorno *et al.*, 2012).

1.3. Neuromesodermal progenitors (NMPs) in axis elongation

1.3.1. Identification of NMPs in the mouse embryo

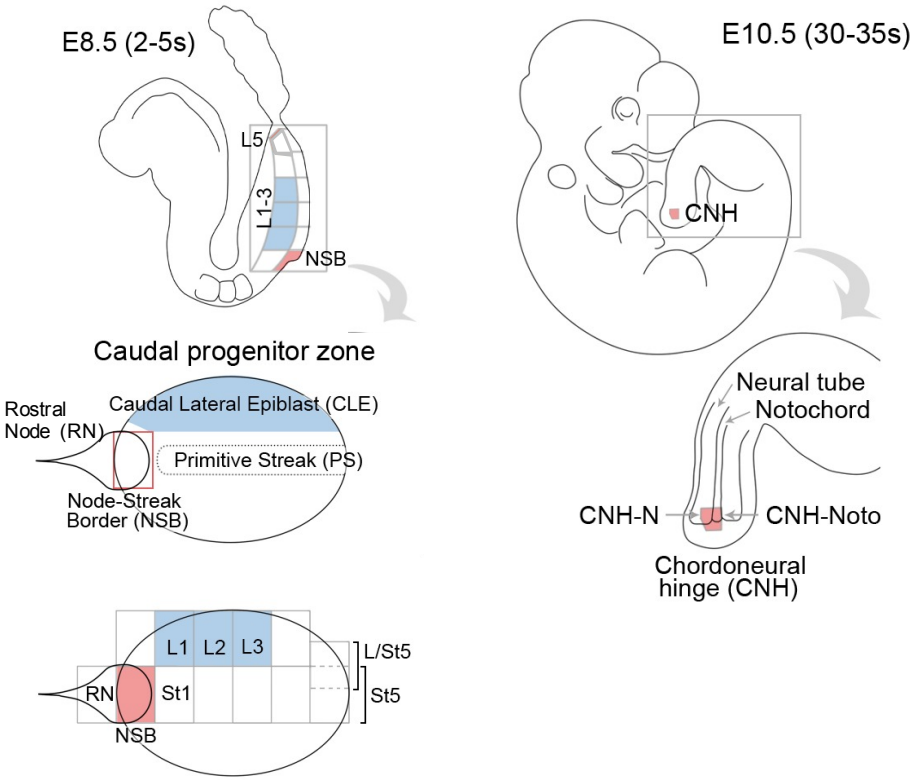
Fate mapping experiments and retrospective clonal analysis studies have indicated the existence of axial stem cells in the primitive streak and the tail bud of the mouse embryo.

The clonal analysis studies presented in this section are based on a *laacZ* reporter construct that lacks β-galactosidase (β-gal) activity due to an internal duplication of the *lacZ* gene. Upon recombination, that happens at random, the open reading frame of the *laacZ* gene is restored, leading to expression of β-gal and labelling in this way, of individual cells and their descendants (Bonnerot *et al.*, 1987). When *LaacZ* was placed under the control of a myotome-specific promoter, the α subunit of the acetylcholine receptor (αAChR), several long clones (contribution > 7 somites) were observed, which had variable anterior limits but extended as far as the posterior end of the myotome on both left and right sides. The characteristics of the clones suggest that there is a pool of self-renewing cells that persists during axis formation, while the bilateral location of the clones and the detection of some of them in very anterior somites suggests that this pool of stem-cell like progenitors must originate in the primitive streak, before the formation of paraxial mesoderm (i.e. E7.5 and/or the anterior primitive streak at E8.5) (Smith *et al.*, 1994; Nicolas *et al.*, 1996). When the *LaacZ* gene was placed under the control of the *neuron specific enolase* (*NSE*) promoter, the $LacZ^{+}$ clones that emerged in the transgenic mouse embryos were extended to the caudal end of the neural tube. Similar to the clonal analysis in the myotome, the characteristics of the clones support the notion that stem cell like

progenitors contribute to the development of the caudal spinal cord (Mathis and Nicolas, 2000).

The insertion of *laacZ* into the ubiquitously expressed *Rosa26* locus allowed for a retrospective clonal analysis that is not restricted to a particular lineage (Tzouanacou *et al.*, 2009). This study demonstrated the existence of bipotent cells, the descendants of which can contribute to the development of both neuroectodermal (N) and mesodermal (M) tissues. These cells are known as neuromesodermal progenitors (NMPs). The frequency of the clones contributing to the N and M lineages in E8.5 and E10.5 embryos suggested a continued presence of common progenitors for these tissues. Importantly, all of the clones in E8.5 embryos and some of the clones in E10.5 embryos were initiated at variable rostral levels, but extended as far as the posterior neuropore, exhibiting characteristics that fit the profile of a stem-cell population at the posterior region of the embryo.

Although these studies indicate that axis elongation unfolds through the self-renewal and specification of stem-cell like precursors, they do not provide information about the localisation of these axial progenitors. The precise embryonic regions where axial stem cells reside during axis elongation were revealed by transplantation studies (figure 1.2). Homotopic grafts of rostral node gave rise to cells that colonised the notochord, whereas grafts of the primitive streak fifths 1 to 4 (St1-4) gave rise to somites and tail bud mesoderm (TBM). NM potential was exhibited by grafts at the border between the node and the primitive streak (NSB), that contributed to the formation of the ventral neural tube, somites, and gave rise to sporadic cells in the caudal end of the notochord. In addition, the chordoneural hinge (CNH) and TBM were colonised (Cambray and Wilson, 2007). Given that fate mapping experiments have shown that the ventral layer of the NSB exclusively forms notochord (Wilson and Beddington, 1996; Cambray and Wilson, 2002), the NM bipotent cells must reside in the dorsal layer of the NSB (figure 1.2).



Stage	Region	Fate	NM potency
E8.5	RN	Notochord	NO
	Dorsal NSB	**Neural tube, paraxial and tail bud mesoderm, notochord, CNH**	**YES**
	Ventral NSB	Notochord	NO
	St1	Paraxial and tail bud mesoderm	NO
	St2-4	Paraxial, tail bud and intermediate mesoderm	NO
	St5	Intermediate, lateral/ventral and tail bud mesoderm	NO
	L1	**Neural tube, paraxial and tail bud mesoderm, notochord, CNH**	**YES**
	L2-3	**Paraxial and tail bud mesoderm, low contribution to lateral/ventral mesoderm**	**YES**
	L4	Not extensively studied *	
	L/St5	Lateral/ventral and tail bud mesoderm	NO
E10.5	**CNH**	**Neural tube, paraxial and tail bud mesoderm, notochord**	**YES**

Figure 1.2: Location of NMPs in the mouse embryo. In E8.5 embryos (2-5s), NMPs are located in the NSB and the L1-3 of the CLE. At the tail bud stages, here represented by an E10.5 embryo, NMPs are found in the CNH. The table summarises the fate and the neuromesodermal (NM) potential of regions in the E8.5 caudal progenitor zone and the E10.5 tail bud. NSB, node-streak border; CLE, caudal lateral epiblast; L1-5, lateral fifths of the lateral epiblast, with L1 corresponding to most rostral adjacent to the node; L/St5, the name indicates that due to the small size of the L5 region, it is possible that part of the St5 region was included in the L5 grafts; St1, rostral 1/5 of the streak; St5, caudal 1/5 pf the streak; CNH, chordoneural hinge. The CNH is composed of two parts: the neural part (termed CNH-N) and the notochordal part (termed CNH-Noto). * Cambray and Wilson (2007) have analysed the fate of L4 in 1 embryo only, in which homotopically grafted cells exhibited contribution to lateral mesoderm, neural tube, CNH and some contribution to paraxial mesoderm; the fate and potency of this region needs to be investigated further. Image modified from Wymeersch *et al*., (2016).

The above studies demonstrate that the node- streak border contributes to the formation of the chordoneural hinge and the tail bud mesoderm. Grafting of chordoneural hinge regions derived from E10.5-12.5 embryos in E8.5 node- streak border resulted to contribution to the axis, including neuroectoderm and mesodermal tissues. In addition, descendants of the grafted cells populated the tail bud mesoderm and the chordoneural hinge. On the contrary, the vast majority of heterochronic grafts of tail bud mesoderm did not incorporate correctly and did not populate neither the tail bud mesoderm nor the chordoneural hinge. Crucially, the latter was shown to contribute cells to the axis and repopulate the chordoneural hinge without apparent loss of potency through serial passages on host embryos, over three generations (Cambray and Wilson, 2002). Altogether, these transplantation studies suggest that the chordoneural hinge in E10.5-12.5 embryos is the equivalent region of the earlier node- streak border and that both regions contain self-renewing NM progenitors that give rise to cells in the neural tube, somites and notochord, contributing in this way to axis formation.

The region flanking the primitive streak is called caudal later epiblast (CLE), while the fifths of this region are named L1-5, with L1 closest to the node (figure 1.2). Homotopic grafts at the L1 region generated descendants in the neural tube, the paraxial and tail bud mesoderm, as well as in the chordoneural hinge, indicating that the L1 region contains putative NMPs. In contrast to that, L2-3 grafts contributed to mesoderm, including the tail bud mesoderm, but not the neural tube or the chordoneural hinge, suggesting that cells in this region are mesoderm-fated. However, when cells from L2-3 were grafted in the node-streak border, their predicted fate was altered and they populated the neural tube, in addition to mesodermal tissues, demonstrating that L2-3 contain NM potent cells (figure 1.2). Interestingly, the fate of L1-3 grafts was affected by the localisation of the cells in the rostral/caudal end of the NSB, suggesting that NMPs do not carry cell-intrinsic lineage bias, but their differentiation into mesoderm or neuroectoderm depends on environmental cues (Wymeersch *et al.*, 2016).

Transplantation of node-streak border, caudal lateral epiblast (L1-3) and chordoneural hinge under the kidney capsule resulted to the generation of small tissue masses that contained mesoderm and neuroectoderm derivatives but lacked embryonal carcinoma cells. In accordance with previous studies showing that pluripotency disappears at the beginning of somitogenesis (Osorno *et al.*, 2012), the inability of these grafts to cause teratocarcinoma formation confirms that cells at these regions are not pluripotent, while the composition of the resulting tissue masses verifies the restriction of the potential of these cells in the neural and mesodermal lineages (Wymeersch *et al.*, 2016).

In summary, NMPs are bipotent, stem cell-like axial progenitors that contribute to the formation of neural tube, somites and notochord. In E8.5 embryos, NM fated progenitors reside in the node-streak border (NSB) and the caudal lateral epiblast (CLE) region L1, while the CLE regions L2-3 contain NM potent cells; both cell populations are considered NMPs. In tail bud stages (E10.5-E12.5), NM progenitors are located in the chordoneural hinge (CNH) (figure 1.2).

1.3.2. Identification of NMPs in other model organisms

Single cell labelling-based clonal analysis in chick embryos (Hamburger and Hamilton stage 4 - HH4/ approximately equivalent of mouse E7) revealed that the region between the Hensen's node (equivalent of the mouse node) and the anterior primitive streak contains cells that contribute to more than one tissue (somite and notochord, or notochord and ventral neural tube) (Selleck and Stern, 1991). Neural and somitic contribution have also been observed upon labelling of 1-3 cells, found close to the node in the CLE of HH6 chick embryos (headfold stage) (Brown and Storey, 2000). In addition, it has been shown that the node contributes to the CNH of the chick embryo, which can be serially passaged to younger host embryos, and as a result repopulates the CNH, but also contributes to the neural tube, the axial and paraxial mesoderm of the host embryo (McGrew *et al.*, 2008). Hence, it seems that axial NM bipotent progenitors are found also in the chick, in areas similar to those in the mouse embryo (reviewed in Wilson *et al.*, 2009).

Fate mapping in the tail bud of *Xenopus* suggested that multipotent progenitors, that contribute to notochord, neural tube, myotome and other tissues, reside in the dorsal part of the CNH. However, the study was performed labelling a group of three cells, therefore we cannot exclude the possibility that the formation of multiple tissues resulted from individual lineage restricted progenitors (Davis and Kirschner, 2000)

In zebrafish, the ventral component of the shield (equivalent of the mouse node) gives rise to the tail somites, while the dorsal part contributes to the tail somites, spinal cord and notochord; however, fate mapping of single cells failed to detect multipotent progenitors in this region (Kanki and Ho, 1997). A small fraction of multipotent cells was detected in the shield at earlier gastrulation stages (Kimmel *et al.*, 1990; Melby *et al.*, 1996). Consistent with Kanki and Ho (1997), a recent study demonstrated that single cells transplanted homotopically into the ventral margin of a shield stage embryo most often populate posterior somites (although a minority of cells contributing to neural tissues or neural tube and somites was also observed); however,

when Wnt signalling inhibition was induced to transplanted cells before or after gastrulation, the number of somite clones was decreased, while the number of spinal cord clones increased, indicating the presence of progenitors with neural/mesodermal potential (Martin and Kimelman, 2012) (Wnt signalling is described in section 1.3.4).

The work presented in this dissertation involved the study of mouse *in vitro* NMPs, hence in the next sections, I will focus mainly on describing the gene expression profile and the signalling pathways that have been characterized in the NMP regions of the mouse embryo.

1.3.3. Markers of NMPs

Expression analysis of genes known to be involved in axis elongation demonstrated that *Wnt3a*, *Fgf8*, *Tbra*, *Cdx2*, *Even-skipped homeobox 1* (*Evx1*) and *Nkx1.2* (also known as *Sax1*) are expressed in the NSB, CLE (L1-3) and CNH. However, these genes are not expressed in NMP regions exclusively, but also in the primitive streak, the emerging mesoderm, the posterior CLE and the tail bud mesoderm (Cambray and Wilson, 2007; unpublished data). To identify markers, but also signalling pathways driving the induction and maintenance of NMPs, our group has performed microarray analysis in regions dissected from the posterior progenitor zone of E8.5 to E13.5 mouse embryos. However, no unique markers could be identified (Wymeersch *et al*., manuscript in preparation).

In several vertebrates, including zebrafish, chick and human, regions known to contain NM axial progenitors are characterised by co-expression of the mesodermal marker *Tbra* and the neural marker *SRY-box2* (*Sox2*) (Martin and Kimelman, 2012; Olivera-Martinez *et al.*, 2012; Tsakiridis *et al.*, 2014). Expression analysis of the two transcription factors in the mouse embryo demonstrated that $Tbra^{+}Sox2^{+}$ cells are first detected in the E7.5 epiblast, close to the NSB. During axis elongation, $Tbra^{+}Sox2^{+}$ cells were found in all NMP regions, with the number of double positive cells reaching

a peak at E9.5 and decreasing thereafter, until completion of somitogenesis (around E13.5) when they disappeared (Wymeersch *et al.*, 2016). As mentioned earlier, in CLE, L1 region contains NM fated cells, while L2-3 harbours NM potent cells. Interestingly, at E8.5, the numbers of $Tbra^{+}Sox2^{+}$ cells were highest in L1, declining towards L3. Importantly, high expression levels of *Tbra* and *Sox2* were detected in regions fated to mesodermal and neural differentiation respectively, while $Tbra^{+}Sox2^{+}$ cells were characterized by low-to-medium expression levels of both genes. However, the NMP regions were not the only areas where the two proteins were found to be co-localised: a smaller number of $Tbra^{+}Sox2^{+}$ cells was also found at the E8.5 midline primitive streak, the region caudal to the CNH at E10.5, the hindgut and notochord (Wymeersch *et al.*, 2016). In summary, NMPs cannot be identified by expression of a single gene. Although not all $Tbra^{+}Sox2^{+}$ cells are NMPs, co-expression of *Tbra* and *Sox2* seems to overlap extensively with NMP identity.

Brachyury (Tbra)

Tbra, a T-box transcription factor, is found at the primitive streak throughout gastrulation, while after the closure of posterior neuropore (E10.0), its expression continues in the tail bud, until somitogenesis is complete (E13.5). Tbra is present also in the node, and its derivative the notochord, where it persists until E14.5. As cells migrate through the primitive streak forming the early mesoderm, they express *Tbra* transiently, whereas in adult tissues, Tbra is absent. (Wilkinson *et al.*, 1990; Kispert and Herrmann, 1994; Cambray and Wilson, 2007; Wymeersch *et al.*, 2016).

Tbra is a direct target of Wnt3a signalling (see paragraph 1.3.4) (Yamaguchi *et al.*, 1999). Mice heterozygous for *Tbra* have short, and often slightly kinked, tails, whereas homozygous *Tbra* mutant embryos die around mid-gestation (E10.5). Loss of function of both *Tbra* alleles results to axis truncation, kinked neural tube, absence of notochord and thickening of the primitive streak; it also leads to disrupted formation of allantois, which is probably the physiological cause of embryonic death (Beddington *et al.*, 1992). Moreover, *Tbra/Tbra* mutants develop ectopic neural tissue in the expense of

paraxial mesoderm (Yamaguchi *et al.*, 1999). Analysis of chimeras composed of wild-type and *Tbra/Tbra* mutant embryonic stem (ES) cells revealed that during gastrulation, the mutant cells progressively accumulate in the primitive streak, leading, later on, to their incorporation into the tail bud. These observations suggested that Tbra plays a key role in the migration of cells through the primitive streak, possibly affecting the cell-to-cell adhesions (Beddington *et al.*, 1992; Wilson *et al.*, 1995). Collectively, these studies suggest that *Tbra* expression is important for axis elongation, but is not essential for anterior mesoderm formation.

Interestingly, the role of Tbra in cell migration and mesoderm formation seems to be dose dependent. Study of chimeras containing transgenic cells that express *Tbra* at different levels demonstrated that cells that ectopically express high levels of *Tbra* exit the primitive streak prematurely. In contrast, primitive steak was shown to accommodate cells with lower Tbra levels for longer. Based on that, it was suggested that the Tbra levels affect the balance between the maintenance of axial progenitors ($Tbra^{low}$) and mesoderm formation ($Tbra^{high}$) (Wilson and Beddington, 1997). This hypothesis is consistent with a recent study demonstrating that the regions where the axial progenitors NMPs reside contain $Tbra^{low}$ cells, whereas $Tbra^{high}$ cells were found in mesoderm fated regions (Wymeersch *et al.*, 2016).

Sox1 and Sox2

Sox1, *Sox2* and *Sox3* comprise the subfamily group B1 of *Sox* genes (Pevny and Lovell-Badge, 1997). SoxB1 factors are expressed in proliferating neural progenitors which they keep in an undifferentiated state, throughout embryonic development as well as adulthood, while they become downregulated when the progenitors exit mitosis to terminally differentiate (Pevny *et al.*, 1998; Bylund *et al.*, 2003; Graham *et al.*, 2003). Sox2 is also part of the gene network that regulates pluripotency (Masui *et al.*, 2007; Gagliardi *et al.*, 2013).

Expression of *Sox2* initiates at the preimplantation stage, when *Sox2* RNA is first detected in some cells of E2.5 morulas and later on, in the ICM of the blastocyst. Following implantation, expression expands throughout the epiblast, but by E7.0–7.5, it becomes restricted to the prospective neuroectoderm in the anterior part. By E9.5, *Sox2* is expressed throughout the brain, neural tube, gut endoderm, as well as in germ cells (Wood and Episkopou, 1999; Avilion *et al.*, 2003). *Sox2* expression is regulated by different enhancers, the N1 and N2 enhancers, in the posterior and anterior neural plate, respectively (Kamachi *et al.*, 2009). The synergistic action of Fgf and Wnt signalling was shown to activate the N1 enhancer in the CLE (Takemoto *et al.*, 2006), where axial progenitors reside; however, deletion of the enhancer sequence did not cause axis defects (Takemoto *et al.*, 2011). *Sox1* expression initiates at E7.5, in the anterior half of the post-implantation epiblast and thereafter, it is present in the developing brain, the spinal cord and the lens (Pevny *et al.*, 1998; Cajal *et al.*, 2012). Expression of *Sox1* has also been detected in NMP regions and in contrast to *Sox2*, it seems to be upregulated at the tail bud stages (Cambray and Wilson, 2007; Wymeersch *et al.*, 2016). The expression profile of *Sox1* in NMP regions is discussed in detail in chapter 4.

Hox genes

As the body axis elongates, the homeobox-containing *Hox* gene family regulates the anteroposterior identity of the axial tissues. Mammals have 39 *Hox* genes, arranged into four clusters; the genes of each cluster are positioned into a linear order and become sequentially activated from 3' to 5'. 3' *Hox* genes are activated in early structures which develop an anterior identity, whereas 5' genes start being progressively expressed in later structures which acquire a more posterior character. The correspondence between the physical order of the *Hox* genes on the genome and their spatiotemporal expression during embryogenesis is called "spatiotemporal colinearity" (figure 1.3) (reviewed in Mallo *et al.*, 2010).

Hox genes start being expressed at late streak stages, when the primitive streak is almost fully extended. They are first activated in the posterior (caudal) part of the primitive streak, and thereafter, they expand rostrally, they meet the node and continue spreading until they reach their definitive rostral boundaries in axial and paraxial structures (figure 1.3). Although expression of the individual genes initiates in a temporal sequence in the primitive streak, the anterior limit of expression of individual *Hox* genes is not inherited by cells passing out of the progenitor region, but instead spreads further anteriorly under the influence of signals such as Wnt, Fgf and retinoic acid (RA) (Forlani *et al.*, 2003; Deschamps and van Nes, 2005). Loss of function mutations in *Hox* genes result in transformation of the rostrocaudal identity of the domain where the gene is expressed (reviewed in Mallo *et al.*, 2010). For example, loss of function mutations in all *Hox10* genes transformed the lumbar and sacral vertebrae, where *Hox10* is normally expressed, into thoracic (Wellik and Capecchi, 2003). Interestingly, in *Hoxb13*$^{-/-}$ embryos, the somites exhibit increased size and the tail contains two extra posterior vertebrae, whereas precocious expression of *Hox13* genes under the *Cdx2* promoter (which is expressed in the posterior progenitor zone from gastrulation to E12.5) causes an axis truncation (Young *et al.*, 2009). Collectively, these studies indicate that while Hox genes in general establish an anteroposterior axial address for groups of vertebrae, *Hox13* genes are involved in the cessation of axis elongation.

Microarray analysis of dissected NMP regions demonstrated collinear activation of *Hox* genes in NMPs (Wymeersch *et al*., manuscript in preparation). More specifically, genes from paralogous groups 1-3 were active at E7.5, genes from groups 4-9 showed minimal activation at E8.5 and a burst of activity between E8.5-9.5, and genes from groups 10-13 were activated at E9.5 and showed a slight further upregulation at E10.5. Interestingly, most *Hox* genes, including members from all paralogous groups, become upregulated at E9.5. This change coincides with an increase in NMP numbers reaching a maximum at E9.5, suggesting that the signals regulating NMP numbers are also intimately linked with the mechanisms of anteroposterior axial patterning (Wymeersch *et al.*, 2016).

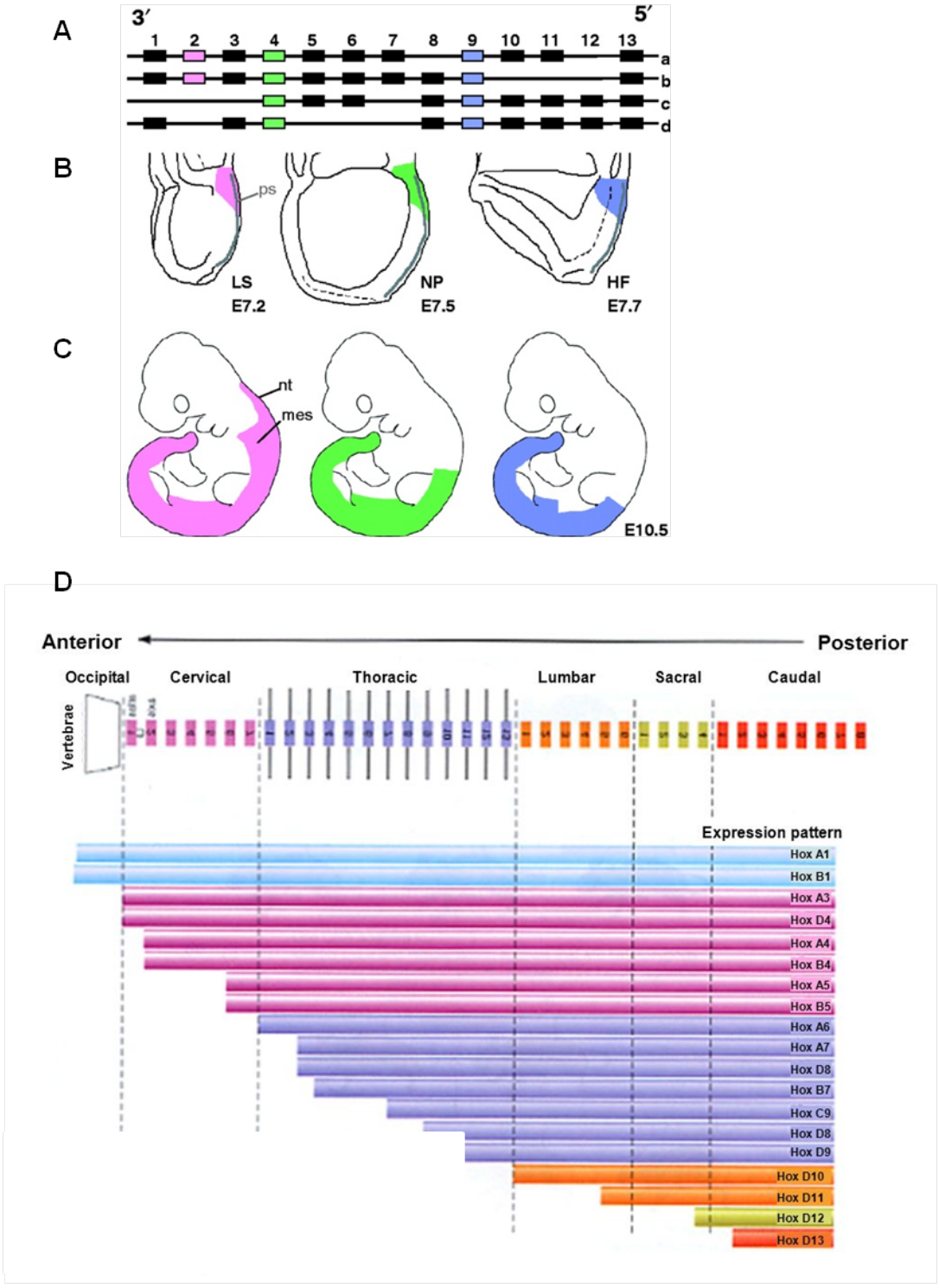


Figure 1.3: *Hox* gene expression in the mouse embryo

Figure 1.3: *Hox* gene expression in the mouse embryo. A) The four clusters of *Hox* genes. Genes with the same number (1 to 13) are called paralogs. The paralogous groups of 2, 4 and 9 are shown in color. **B)** *Hox*2 genes begin to be expressed earlier, at E7.2, while *Hox4* and *Hox*9 paralogs progressively later, at E7.5 and E7.7 respectively, in the posterior part of the primitive streak of the mouse embryo. **C)** At E10.5, the 3' genes extend to more rostral domains than the 5' genes. For each gene, the expression boundary is more anterior in the nervous system than in the mesoderm. Modified from Deschamps and van Nes, (2005). **D)** Hox expression domains along the mouse vertebral column. Image from http://slideplayer.com/slide/8985461/ (accessed 09/03/2017). ps, primitive streak; ls, late streak; np, neural plate; hf, head fold; mes, mesoderm; nt, neural tube.

1.3.4. Signalling pathways involved in axis elongation

A number of signalling pathways, including the Notch, Bmp, RA, Wnt and Fgf pathways are involved in the regulation of axis elongation, affecting the self-renewal or differentiation of NMPs (reviewed in Wilson *et al.*, 2009). Here, I will introduce the Wnt and Fgf signals, as stimulation of these pathways allows us to differentiate pluripotent stem cells to *in vitro* NMPs (see section 1.4.3) which constitute the focus of this work.

Wnt signalling

The Wnt proteins form a large family of 19 secreted glycoproteins. From those proteins, Wnt3a, Wnt5a and Wnt5b are the only ones that are expressed in the posterior progenitor zone from the primitive streak to the tail bud stages (Takada *et al.*, 1994). Amongst the three genes, Wnt5b is the only one that does not cause development abnormalities when homozygous loss of function mutations occur (reviewed in van Amerongen and Berns, 2006), therefore here, I will focus mainly on presenting the role of Wnt3a and Wnt5a in axis elongation.

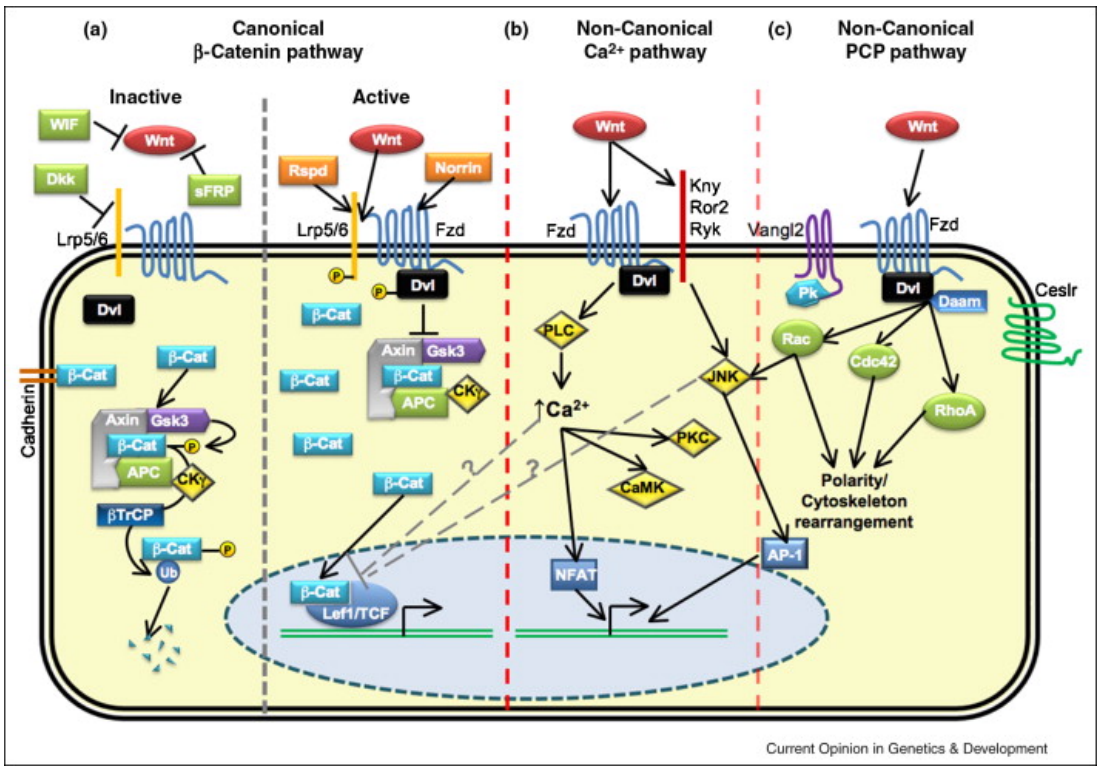


Figure 1.4: Schematic diagram of canonical and non canonical Wnt signalling pathways. A) Key components of the canonical Wnt/ β-catenin signalling pathway. **B)** Key components of the non canonical Wnt/ Ca^{2+} pathway **C)** Key components of the non canonical Wnt/ PCP signalling pathway. Image from Fedon *et al.*, (2012).

Wnt proteins mediate signal transduction through several different pathways that are divided into canonical and non canonical pathways (reviewed in Komiya and Habas, 2008) (figure 1.4). The canonical pathway is β-catenin dependent. In the absence of Wnt signalling, or upon binding of inhibitors such as Dkk1, a complex of proteins, including the Glycogen synthase kinase 3 (Gsk-3), phosphorylates the cytoplasmic β-catenin, leading it in this way to proteasomal degradation. Binding of Wnt to its receptor composed of Frizzled (Fz) and Low density lipoprotein receptor-related protein (LRP5/6) triggers a series of events that disrupts the phosphorylation complex. As a consequence, β-catenin becomes stabilised and accumulates in the cytoplasm. The stabilized β-catenin translocates into the nucleus, where it interacts with transcription factors such as proteins of the Lymphoid enhancer binding factor /T cell-specific transcription factor (Lef/Tcf) family, regulating in this way, transcriptional

activation of target genes (reviewed in Komiya and Habas, 2008). In the mouse embryo, active Wnt/β-catenin signalling has been detected in the primitive streak, its adjacent posterior epiblast, the mesodermal wings, the node and the emergent notochord, during gastrulation and early somitogenesis; at later stages, it has also been detected in the tail bud (Ferrer-Vaquer *et al.,* 2010).

Wnt3a stimulates the canonical Wnt pathway. In *Wnt3*a null mice, most of which die before E12.5, the somites posterior to the forelimb are absent, the notochord is disrupted and an ectopic neural tube develops in the expense of paraxial mesoderm (Yoshikawa *et al.*, 1997). Overexpression of *Wnt3a* in the epiblast, under the enhancer of *Cdx2*, led to severe axis truncations too; however, the formation of neural tissues was prevented, while mesoderm differentiation had taken place (Jurberg *et al.*, 2014). Interestingly, the role of Wnt3a seems to be gene dosage dependent, with increasing levels of *Wnt3a* activity being necessary for the formation of more posterior somites (Greco *et al.*, 1996). Lineage tracing revealed that conditional deletion of Wnt3a or b-catenin in *Tbra* expressing tissues redirects mesoderm fated cells towards neural differentiation (Garriock *et al.*, 2015). A transplantation study shed light on the role of Wnt/β-catenin signalling specifically in the fate choices of NMPs (Wymeersch *et al.*, 2016). As shown in figure 1.2, the caudal lateral regions L1-3 are NM potent, whereas the rostral streak parts St1-3 give rise to paraxial mesoderm and tail bud mesoderm. Wymeersch *et al.* (2016) presented that transplantation of L1-3 to St1-3 contributed exclusively to paraxial mesoderm and tail bud mesoderm; however, upon conditional deletion of *β-catenin* in donor embryos, the potency of NMPs was altered and grafted cells that contributed to the caudal regions of the host embryos produced only neurectoderm, demonstrating that Wnt/β-catenin signalling is required for the mesodermal fate of NMPs and that in the absence of it, NMPs undergo neural differentiation (Wymeersch *et al.*, 2016). Although these studies suggest that Wnt3a/ β-catenin signalling prevents axial progenitors from going into neural differentiation, the synergistic action of Fgf and Wnt signaling has been shown to induce the formation of posterior neural plate through activation of the *Sox2* enhancer N1 in the CLE (Takemoto *et al*., 2006). In addition, stabilization of β-catenin in the $Tbra^{+}$ cell

compartment led to impaired somitogenesis, but did not affect the progression of axial progenitors into neural differentiation (Jurberg *et al.*, 2014).

It has been suggested that Wnt/ β-catenin signalling is also involved in the maintenance of NMPs. As explained in section 1.3.3, the number of double positive cells increases from E8.5 to E9.5 and thereafter it decreases until the end of axis elongation when they disappear (Wymeersch *et al.*, 2016). Interestingly, Wnt3a expression is also gradually lost in the tail bud during axis elongation (Cambray and Wilson, 2007). Wymeersch *et al.* (2016) reported that upon conditional deletion of *β-catenin* the number of $Tbra^{+}Sox2^{+}$ cells in E9.5 embryos was significantly reduced compared to wild type embryos; in addition, the number of $Tbra^{+}Sox2^{+}$ cells in *β-catenin* deleted E9.5 embryos was not significantly different from the wild type E8.5 embryos, suggesting a failure of NMPs to expand (Wymeersch *et al.*, 2016). Collectively, these data indicate that Wnt/ β-catenin signalling plays a key role in the maintenance and the fate decisions of NMPs.

The group of non canonical pathways is β-catenin independent, including the Wnt/Ca^{2+} pathway and the Planar Cell Polarity (PCP) pathway (reviewed in Komiya and Habas, 2008). Activation of the non-canonical Wnt/Ca^{2+} pathway leads to an increase in the intracellular calcium levels, possibly through the activation of Phospholipase C (PLC). Intracellular calcium leads to activation of Ca^{2+}/Calmodulin-dependent (CamKII) protein kinase, Protein kinase C (PKC) and the transcription factor Nuclear factor of activated T cells (NFAT). Additionally, Fzd receptors in association with Knypel (Kny), Ror2 or Ryk receptors can also activate c-Jun N-terminal kinase (JNK), promoting in this way expression of target genes through activation of Activator protein-1 (AP-1). In the PCP pathway, binding of Wnt to its receptor Frizzled (Fz), activates Disheveled, leading to subsequent activation of the downstream effectors JNK and Rho family GTPases. As a result, cytoskeleton rearrangements occur, while asymmetric distribution of Fzd, Celsr (a homolog of flamingo), Prickle (Pk) and Van Gogh/Strabismus (Vangl) leads to the polarization of the cell. The PCP pathway regulates the convergence-extension movements of axial and paraxial mesoderm,

playing in this way a key role in anterior-posterior axis elongation. Disruptions of this signalling pathway result in a shortened and widened A-P axis, and in neural tube defects (Hamblet *et al.*, 2002; Lu *et al.*, 2004; Ybot-Gonzalez *et al.*, 2007; Etheridge *et al.*, 2008; Song *et al.*, 2010).

Wnt5a stimulates the PCP and the Ca^{2+} non canonical Wnt pathways (Slusarski *et al.*, 1997; Gao *et al.*, 2011). In *Wnt5a* null mice the primitive streak appears shortened and the body axis is characterised by absence of caudal vertebrae and reduction in size of those that have formed (Yamaguchi *et al.*, 1999); further shortening of the body axis was observed upon loss of both Wnt5a and Wnt11 (Andre *et al.*, 2015). It has been shown that in the absence of Wnt5a, ingression of epiblast cells through the primitive streak for mesoderm formation and subsequent paraxial mesoderm specification is not disturbed. In contrast to *Wnt3a* null mice, *Wnt5a* mutants contain a tail bud, but by E10.5 the generation of new paraxial mesoderm is prematurely terminated. Since Wnt5a is not expressed in the forming somites themselves, these data suggest that Wnt5a regulates the self-renewal of axial progenitors (Yamaguchi *et al.*, 1999).

Fgf signalling

Most Fgfs trigger intracellular signalling pathways by binding and activating Fgf receptors (Fgfr), a subfamily of cell surface receptor tyrosine kinases (RTKs). Fgf signal transduction can then proceed through three different intracellular cascades that involve the downstream mediators Ras/Mapk (Erk), PI3K/Akt1 or Plcγ1. An overview of these pathways is presented in figure 1.5.

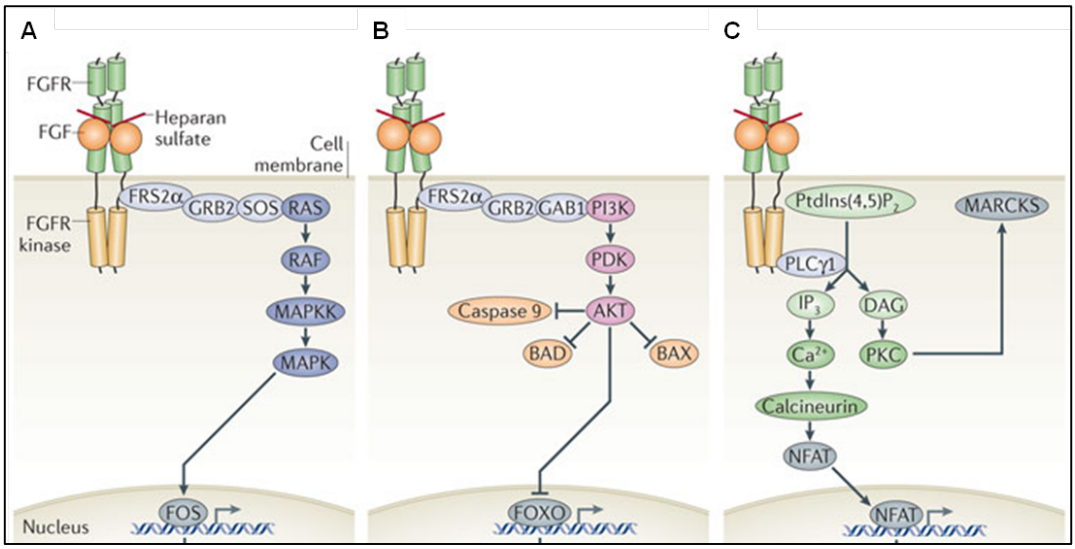


Figure 1.5: Fgf signalling pathways. Binding of Fgfs to their receptors (Fgfrs) induces Fgfr dimerization and initiate distinct signalling pathways that generate diverse cellular responses. **A)** Key components of Ras/Mapk pathway. **B)** Key components of PI3K/Akt pathway; **C)** Key components of Plcγ1 pathways. Image from Goetz and Mohammadi (2013).

The Fgf family includes 23 proteins. Fgf4 is expressed in the primitive streak and the CLE during gastrulation and early somitogenesis, while at E10.5 it is detected in the tail bud (Niswander and Martin, 1992; Boulet and Capecchi, 2012). Homozygous loss of function mutations of *Fgf4 l*ead to embryonic lethality around implantation, due to due to a lack of primitive endoderm (Feldman *et al.*, 1995; Kang *et al.*, 2013). *Fgf8* is another member of the Fgf family that is expressed in the caudal progenitor regions during axis elongation, as well as the presomitic mesoderm, but not in the somites. In the caudal progenitor region of the mouse embryo, Fgf8 expression is maintained by Wnt3a (Aulehla *et al.*, 2003), while in chick, Fgf8 induces Wnt8c (the orthologue of mouse Wnt8a) in the forming neural axis (Olivera-Martinez *et al.*, 2012). In both model organisms, Fgf8 and RA inhibit each other mutually, while RA further inhibits Wnt3a and Wnt8c (Iulianella *et al.*, 1999; Shum *et al.*, 1999; Diez del Corral and Storey, 2004; Olivera-Martinez *et al.*, 2012).

As mentioned earlier, in *Fgf8*$^{-/-}$ embryos, cells ingress at the streak but fail to migrate, therefore mesoderm and endoderm are not formed (Sun *et al.*, 1999). Conditional inactivation of *Fgf4* or *Fgf8* in the posterior regions of the embryo (by using HoxB1-IRES-Cre) did not result in any visible defect in axis development; however, conditional deletion of both genes gave rise to vertebral malformations and axis truncations, suggesting that Fgf4 and Fgf8 are functionally redundant. More specifically, in double mutant mice, the neural tube was shortened and the presomitic mesoderm was affected increasingly in more caudal regions, with somite formation being blocked after the production of about 15–20 somites. These abnormalities seemed to result from a failure to produce sufficient paraxial mesoderm, rather than inability of precursor cells to migrate, suggesting overall that Fgf4 and Fg8 are necessary for the maintenance of axial progenitors in the caudal embryonic regions (Boulet and Capecchi, 2012). When *Fgf4* and *Fgf8* were both inactivated specifically in the presomitic mesoderm before the onset of somitogenesis (using Tbra-cre), the presomitic mesoderm was differentiated prematurely, suggesting a role of Fgf signalling in the suppression of presomitic mesoderm differentiation (Naiche *et al.*, 2011).

Collectively, these studies indicate that Fgf signalling is important for the maintenance of the posterior progenitor zone, as well as the morphogenesis and patterning of mesoderm during axis elongation. Interestingly, Fgf signalling has been shown to induce neural differentiation in amphibians and in chick (reviewed in Stern, 2005), including the activation of *Sox2* enhancer N1 in the CLE (Takemoto *et al.*, 2006). It has been suggested that the dual role of Fgf in mesoderm and neuroectoderm formation may be determined by the levels of activation of this signalling pathway; more specifically, high levels of or prolonged exposure to Fgf signalling may be responsible for inducing mesoderm differentiation, whereas low levels may direct cells towards a neural fate (reviewed in Wilson *et al.*, 2009).

In summary, accumulating evidence suggests that a combination of Fgf and Wnt signalling promotes the self-renewal of NMPs, while RA induces differentiation in part by inhibiting Fgf8 expression as cells move out of the CLE and the primitive streak in chick and mouse embryos. The differential exposure of axial progenitors to Wnt and Fgf signals might affect their fate decisions; according to this model, cells that ingress through the primitive streak are exposed to high levels of Fgf8 and Wnt3a for longer, therefore they give rise to mesoderm, whereas cells that migrate from the NSB/CLE to form the neuroepithelium might be exposed to lower levels of Fgf for shorter time periods and hence they switch to the neural lineage (reviewed in Wilson *et al.*, 2009).

1.4. Modelling the mouse embryonic development *in vitro*

1.4.1. Capturing the mouse ground state pluripotency *in vitro*: mouse ES cells

Based on its ability to give rise to all embryonic tissues, the pre-implantation epiblast is considered the developmental ground state (Gardner, 1998). Isolation of pre-

implantation epiblast cells and expansion under conditions that promote their proliferation gives rise to ES cells (Evans and Kaufman, 1981; Martin, 1981). Hence, ES cells constitute an *in vitro* model of ground state pluripotency (reviewed in Nichols and Smith, 2009). Stem cells are defined by the capacity to generate daughter cells with the exact same properties (self-renewal) and daughter cells specified for differentiation. ES cells exhibit unlimited self-renewal capacity, while they retain their multilineage developmental potency; specifically, it has been shown that when inserted into the pre-implantation epiblast, ES cells contribute to the development of all embryonic lineages, including the germline (Bradley *et al.*, 1984). Although ES cells can be derived from preimplantation embryos of different stages, it has been shown that their transcriptional profile resembles the E4.5 epiblast (Boroviak *et al.*, 2014). The gene network that regulates self-renewal of ES cells includes the three core pluripotency factors, *Oct4*, *Sox2* and *Nanog*, but also the *Kruppel-like factors 2* and *4* (*Klf2* and *Klf4*) (reviewed in Chambers and Tomlinson, 2009). In addition, ES cells are characterised by expression of genes known to be active in the ICM, such as *REX1* and the *Stage-specific embryonic antigen-1* (*SSEA-1*), while they lack expression of lineage specification markers. In female embryos, X chromosome reactivation takes place transiently in the pluripotent epiblast prior to implantation. ES cells exhibit the same epigenetic feature, with cells derived from female embryos containing two active X chromosomes (reviewed in Nichols and Smith, 2009; Martello and Smith, 2014).

Originally, ES cells were cultured on a layer of mitotically inactivated fibroblasts in serum containing medium (Evans and Kaufman, 1981; Martin, 1981). Later on, it was discovered that mouse embryo fibroblasts promote the self-renewal of ES cells through production of the cytokine Leukemia Inhibitory Factor (Lif) and it was shown that addition of recombinant Lif in the culture medium could support the long-term maintenance of ES cells, in the absence of feeder cells (Smith *et al.*, 1988; Williams *et al.*, 1988). Lif exerts its effect on ES cell pluripotency through activation of the Jak/Stat3 signalling pathway (reviewed in Onishi and Zandstra, 2015). In the absence of serum though, development of neural derivatives was observed in feeder free ES cell cultures, suggesting that Lif alone is not sufficient to suppress neural differentiation. Addition of Bmp4 in the serum free medium blocked the formation of neural

derivatives and promoted the maintenance of undifferentiated ES cells (Ying *et al.*, 2003). Fgf4 has been shown to stimulate ES cell differentiation through activation of the Mapk pathway; in addition, it has been reported that stimulation of Wnt/β-catenin signalling promotes ES cell propagation. Based on these observations it was found that blockage of Fgf/Mapk pathway (by the inhibitor PD184352), in combination with suppression of Gsk3 (by CHIR99021) is sufficient for robust propagation of undifferentiated ES cells in serum free conditions (2i medium) (Ying *et al.*, 2008). A summary of the signalling pathways involved in the regulation of ES cell self-renewal is presented in figure 1.6.

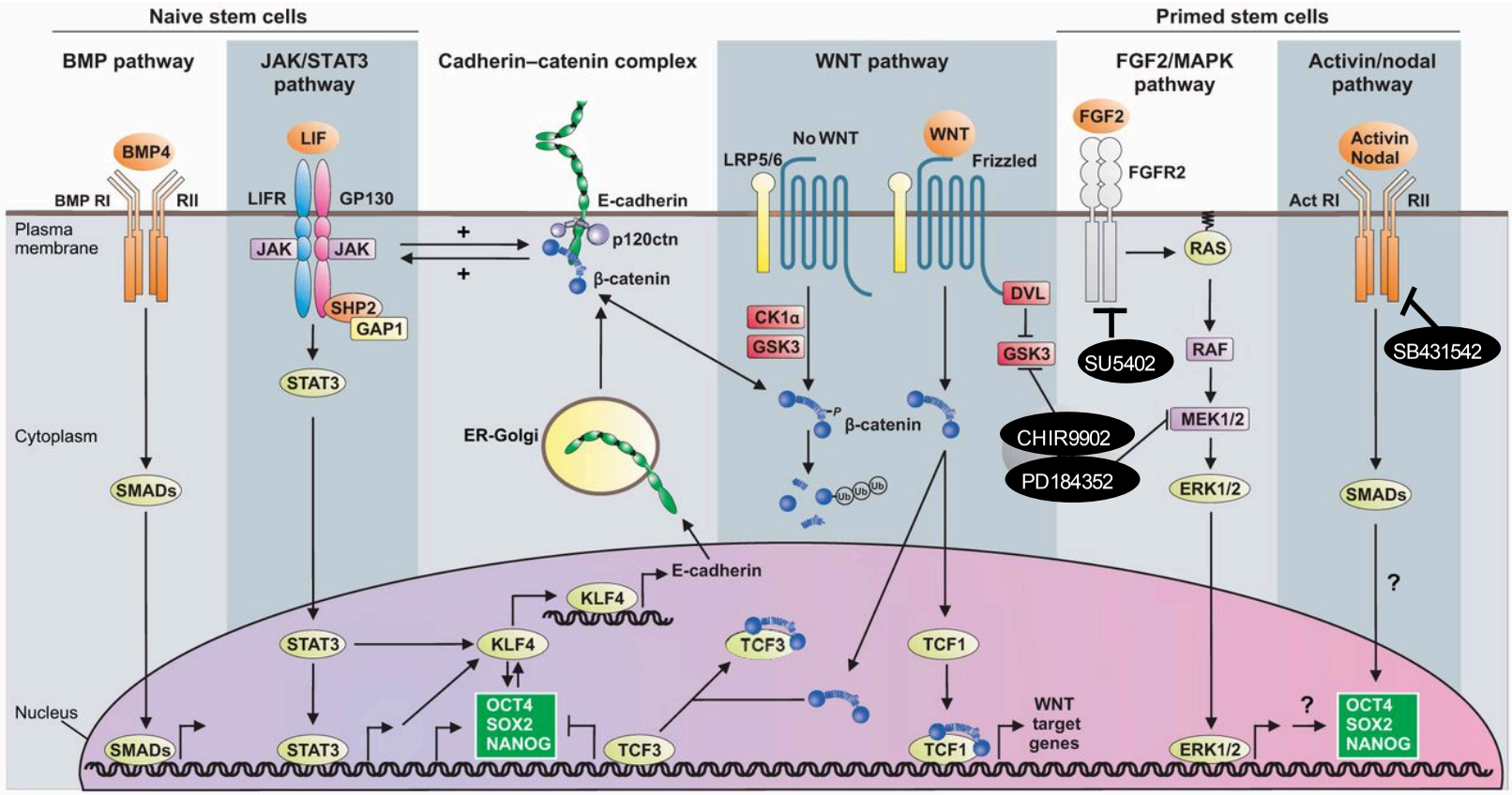


Figure 1.6: Signalling pathways involved in regulation of naive and primed pluripotency. Image modified from Pieters and van Roy (2014).

1.4.2. Capturing the mouse primed pluripotency *in vitro*: mouse EpiSCs

As discussed above, following implantation, the epiblast is characterised by co-expression of pluripotency factors and early lineage specification markers. In addition, cells from the post-implantation epiblast retain their capacity to give rise to teratocarcinomas, but they lose the ability to incorporate into the blastocyst. Therefore, it is considered that the post-implantation epiblast represent a distinct developmental state, the primed pluripotent state (reviewed in Nichols and Smith, 2009). Cell lines derived from the post-implantation epiblast, EpiSCs, recapitulate the properties of their tissue of origin and hence, they provide us with an *in vitro* model of primed pluripotency (Brons *et al.*, 2007; Tesar *et al.*, 2007). Interestingly, mouse EpiSCs share great similarities with human ES cells (growing in the presence of Fgf2 and Activin) and they are considered the rodent equivalent of the human ES cell pluripotent state (reviewed in Rossant, 2008; Nichols and Smith, 2009; De Los Angeles *et al.*, 2012).

EpiSCs can be propagated in serum free medium, containing Fgf2 and Activin A. They are capable of multilineage differentiation *in vitro* and can form teratocarcinomas when transplanted into adult mice. In contrast to ES cells, they do not incorporate into the blastocyst (Brons *et al.*, 2007; Tesar *et al.*, 2007), but they integrate and contribute to the development of all three embryonic lineages and primordial germ cells upon grafting in the post-implantation epiblast (Huang *et al.*, 2012; Kojima *et al.*, 2014). Similar to the *in vivo* epiblast cells that undergo X-chromosome inactivation shortly after implantation, EpiSCs derived from female embryos contain a silent X chromosome (Bao *et al.*, 2009; Guo *et al.*, 2009). In this cell type, pluripotency is governed by the three core factors Oct4, Sox2 and Nanog, but not Klf2 and Klf4 that are active in ES cells. Specifically, expression of *Oct4* is driven by the proximal enhancer, that is active in the post-implantation epiblast, but not in the ICM. In addition, genes found to be active in the ICM, such as the *Platelet endothelial cell adhesion molecule* (*Pecam1*), become downregulated, whereas early lineage

specification markers, such as *Tbra*, *Sox17*, *Otx2* and *Cer1* are expressed (Brons *et al.*, 2007; Tesar *et al.*, 2007). A number of studies have shown that expression of pluripotency and differentiation markers in EpiSC cultures is heterogeneous and appears to be associated with the existence of subpopulations with different properties (see chapter 3) (Hayashi and Surani, 2009; Han *et al.*, 2010; Tsakiridis *et al.*, 2014).

EpiSC lines can be generated from embryos of various developmental stages, ranging from the pre-implantation E3.5 ICM (Najm *et al.*, 2011) to the E8.0 epiblast that is about to exit pluripotency (Osorno *et al.*, 2012). Trascriptome analysis of EpiSC lines derived from mouse embryos at developmental stages between E5.5 to E8.25 has shown that EpiSCs do not retain the characteristics of their source tissue, but they converge to a cellular state that resembles the late-gastrulation epiblast (Kojima *et al.*, 2014). EpiSCs can also be generated *in vitro*, through differentiation of ES cells (Guo *et al.*, 2009). It has been reported, that the reverse transition requires overexpression of transgenes such as *Klf4* or the *Myelocytomatosis oncogene* (*c-Myc*); however, a number of studies has shown that reprogramming of EpiSCs to the ground state pluripotency can also be achieved by simply culturing EpiSCs in ES cell medium (Bao *et al.*, 2009; Hanna *et al.*, 2009; Greber *et al.*, 2010).

EpiSC self-renewal depends on Fgf2 and Activin A. Activin A belongs to the same family as Nodal, the Transforming growth factor-beta (Tgf-beta) superfamily. Activin is recognised by heterodimers of transmembrane receptors with intracellular activity of serine/threonine kinase (figure 1.6). Upon binding, Smad family members become phosphorylated and consequently transferred into the nucleus where they regulate transcription of different genes (reviewed in Shi and Massagué, 2003). Culture of EpiSCs in medium lacking Fgf2 and Activin gave rise to cells the expression profile of which resembles the anterior neural plate of E7.5 gastrulating embryos (Iwafuchi-Doi *et al.*, 2012). It has been shown that in EpiSCs, Smad2/3 signalling directly controls expression of *Nanog* (Vallier *et al.*, 2009a; Greber *et al.*, 2010). Inhibition of the Activin/Nodal signalling has been reported to induce neural differentiation (Tesar *et al.*, 2007), while under the same culture conditions, overexpression of *Nanog* was

sufficient to block specification towards the neural lineage (Vallier *et al.*, 2009a). These finding are consistent with *in vivo* studies demonstrating that in the post-implantation epiblast, *Nanog* is regulated by Nodal/Smad2 signalling (Sun *et al.*, 2014) and that *Nodal* null epiblast cells fail to form mesoderm and endoderm, but they develop neural characteristics prematurely (Camus *et al.*, 2006). In summary, Activin/Nodal signalling maintain pluripotency and inhibit neural differentiation in EpiSC cultures.

Fgf/Mapk/Erk signalling (figure 1.5 and 1.6) has been shown to have an inhibitory role in neural differentiation of EpiSCs, but it is not sufficient to block it (Greber *et al.*, 2010). As mentioned before, EpiSCs can revert to an ES cell-like state when cultured in conditions that promote the ground state pluripotency, including media supplemented with Fgf signalling inhibitors (Bao *et al.*, 2009; Greber *et al.*, 2010; Hanna *et al.*, 2009). It has been suggested that Fgf signalling is involved in the ES cell ↔ EpiSC transition; specifically, it has been shown that Fgf signalling suppresses expression of the ES cell pluripotency factor *Klf2* and in this way, it stabilises EpiSCs at the primed pluripotents state. These data agree with the study of Kunath *et al.* (2007) that demonstrates that ES cells require Fgf/ERK signalling in order to exit the naive self-renewal state and acquire the capacity to undergo lineage specification.

When EpiSCs are treated with a combination of Fgf2, Activin and Bmp4, mesoderm/endoderm-like cells emerge, whereas upon inhibition of Activin/Nodal signalling, Bmp4 induces differentiation into extra-embryonic tissues (Vallier *et al.*, 2009b). Depending on the activation levels, stimulation of Wnt/β-catenin pathway in EpiSC cultures results in the formation of subpopulations with different characteristics; low Wnt activity induces an early primitive-streak like character in cells that retain their pluripotency, whereas higher levels of Wnt signalling cause exit from the pluripotent state and give rise to mesendoderm ($Tbra^{+}Foxa2^{+}$) and neuromesodermal ($Tbra^{+}Sox2^{+}$) progenitors (Tsakiridis *et al.*, 2014) (see section 1.4.3).

1.4.3. Generation of mouse and human *in vitro* NMPs

In vivo studies have shown that Wnt and Fgf signalling are involved in the maintenance of NMPs and the formation of axial tissues in the posterior part of the embryo. Activation of these signalling cascades in ES cells and EpiSCs led to the generation of *in vitro* NMPs (figure 1.7) (reviewed in Henrique *et al.*, 2015).

Specifically, Tsakiridis *et al.* (2014) first demonstrated that stimulation of the Wnt signalling pathway (through addition of the Gsk3β inhibitor, CHIRON99021) in EpiSC cultures abolishes pluripotency and gives rise to two mutually exclusive subpopulations, consisting of *Tbra/Sox2* and *Tbra/Foxa2* coexpressing cells. As discussed before, double positivity for Tbra and Sox2 overlaps extensively with NMP identity (Wymeersch *et al.*, 2016), whereas coexpression of *Tbra* and *Foxa2* is considered to mark common mesoderm/endoderm precursors (section 1.2). Under these conditions, the upregulation of posterior *Hox* genes and genes expressed in the posterior neural late was observed, while transcription of pluripotency factors and anterior neural markers was shown to be reduced. Importantly, although self-renewing EpiSCs fail to integrate into the NSB of E8.5 embryos (Huang *et al.*, 2012), EpiSCs with elevated Wnt activity exhibited the capacity to colonise the host embryos and in some of them, they contributed to the development of neural tube and somites. Altogether, these data suggested that Wnt signalling activation promotes the formation of NMP-like cells. However, under these culture conditions, the $Tbra^{+}/Sox2^{+}$ cells were outnumbered by the $Tbra^{+}/Foxa2^{+}$ mesendoderm precursors, that most likely have been induced by Activin (Kubo *et al.*, 2004; Gadue *et al.*, 2006).

Incubation of EpiSCs for 48 to 72 hours in medium lacking Activin A, but supplemented with CHIRON99021 (CHI) and higher concentration of Fgf2 (hereafter referred to as FGF/CHI) increased the efficiency of NMP induction, with Tbra/Sox2 double positive cells accounting for approximately 80% of the culture (Gouti *et al.*, 2014). Under FGF/CHI, the expression profile of induced EpiSCs was shown to

resemble the gene activity observed in the caudal progenitor regions, while grafting experiments in mouse embryos confirmed that *in vitro* NMPs are functionally equivalent to their *in vivo* counterparts. Crucially, the dual differentiation potential of *in vitro* generated NMPs has been tested at the single cell level and it has been shown that individual Tbra-GFP^{+} cells derived from FGF/CHI treated cultures have the capacity to give rise to clones containing both neural- and mesoderm-like differentiation products (Tsakiridis and Wilson, 2015).

Similar to the EpiSC protocols, culture of ES cells in N2B27 medium supplemented with Fgf2 and CHI was shown to induce the transition from the pluripotent to the NM potent state (Gouti *et al.*, 2014; Turner *et al.*, 2014). The capacity of the ES cell-derived NMPs to contribute to the development of somites and the neural tube was confirmed *in vivo*, through grafting in chick embryos (Gouti *et al.*, 2014). *In vitro*, subsequent exposure of the bipotent progenitors to CHI gave rise to paraxial mesoderm differentiation products, such as cells expressing the *Myoblast determination protein 1* (*MyoD*); on the contrary, treatment with RA and a Shh agonist (SAG), generated neural precursors that exhibited posterior identity. Specifically, cells acquired neuronal morphology and expressed the neuronal marker *class III β-tubulin* (also known as *Tuj1*), as well as markers of motor neurons, such as the *Oligodendrocyte transcription factor 2* (*Olig2*) and the *Insulin gene enhancer protein Isl-1/2* (*Islet1/2*). Crucially, progressive upregulation of more 5' *Hox* genes, including *HoxC8* and *Hoxc9*, was observed. However, when CHI was not included in the regime for the *in vitro* generation of NMPs, RA and SAG gave rise to neural precursor with anterior characteristic. Consistent with studies suggesting that Wnt signalling is involved in posteriorising newly formed tissues (Nordstrom *et al.*, 2006; van de Ven *et al.*, 2011), altogether, these data suggest that the CHI-mediated generation of NMPs allows the further differentiation of these progenitors into spinal cord cells with posterior identity.

Activation of Wnt and Fgf signalling has been shown to induce NMP identity in human ES cells too (Gouti *et al.*, 2014; Denham *et al.*, 2015; Lippmann *et al.*, 2015). In the study of Gouti *et al.* (2014), it is presented that the mouse EpiSCs-to-NMPs

differentiation protocol is sufficient for the generation of hNMPs. Under these culture condition (FGF/CHI), 80% of the culture turned into $Tbra^{+}/Sox2^{+}$, that most likely resemble the *Tbra/Sox2* coexpressing cells found in the caudal epiblast region of human embryos (Olivera-Martinez *et al.*, 2012). However, similar to all cultures described so far, the generation of NMP was only transient, and after 72h, the Tbra/Sox2 double positive cells turned into cells that express either neural (e.g. Sox2) or mesoderm (e.g. Tbra, Tbx6) markers. Lippmann *et al.* (2015) allowed hES cells to rest for one day in E6 medium, following treatment with FGF8b for one more day and subsequent addition of CHI; this protocol generated a more homogeneous NMP-like population (75-100% $Tbra^{+}/Sox2^{+}$) that could be maintained for up to seven days. Similar to mouse cell cultures, stimulation of Fgf and Wnt signalling triggered collinear activation of *Hox* genes. Crucially, it was shown that exposure of NMP cultures to RA at different time points arrests *Hox* gene activation and induces the generation of neuroectodermal cell types with different anterior-posterior identities (Lippmann *et al.*, 2015).

As described before, the anterior and posterior nervous system have distinct origins; the anterior epiblast gives rise to the anterior neural plate, whereas the dual fated NMPs have been shown to contribute caudally to the development of the neural tube (see section 1.2). Hence, the generation of *in vitro* NMPs allows us to capture in culture the immediate precursors of posterior spinal cord cells. Importantly, evidence provided by the abovementioned studies suggests that the NMP state can be used as the first step of refined differentiation protocols that could efficiently generate neural cell types with specific anterior-posterior regional identity. Such an improvement on the differentiation protocols of human pluripotent cells could have a great impact on the success of cell replacement therapies. In addition, NMP cultures provide us with a valuable *in vitro* model that could help us to dissect the molecular mechanisms underlying the self-renewal and differentiation of these axial progenitors.

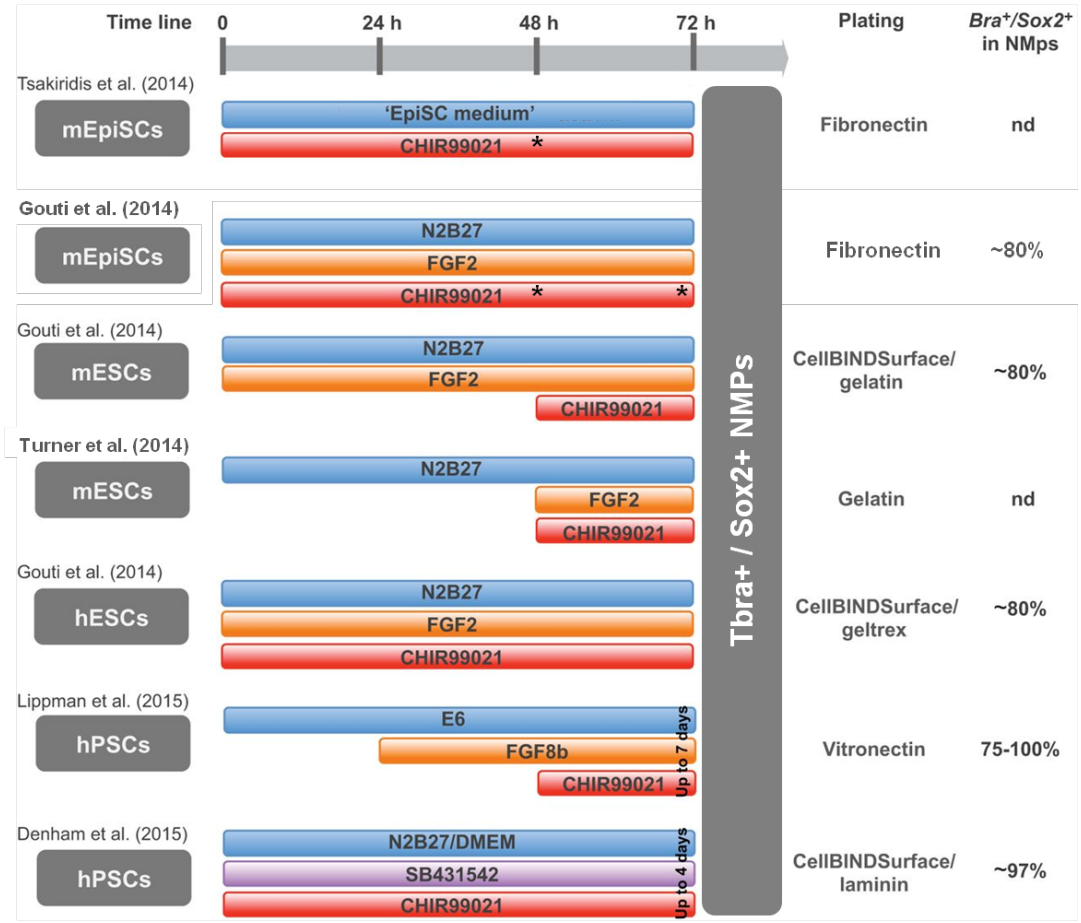


Figure 1.7: Generation of *in vitro* NMPs. Summary of the protocols used up to date for the *in vitro* differentiation of mouse and human pluripotent cells into NMPs. The percentage of Tbra/Sox2 double positive cells that emerge in each culture condition is presented. EpiSC medium refers to N2B27 supplemented with Activin A and Fgf2. CHIR99021 is a Wnt signalling agonist, while SB431542 is an inhibitor of the Activin/Nodal signalling. In some cases, 48h and 72h NMPs were generated; the incubation time is shown by the asterisks (*). m, mouse; h, human; ESCs, embryonic stem cells; EpiSCs, epiblast-derived stem cells; PSC, pluripotent stem cells; nd, not determined. Modified from Henrique *et al.* (2015).

1.5 Scope of the thesis

In vivo, once somitogenesis initiates, pluripotency disappears from the epiblast and bipotent NM progenitors are detected in the caudal part of the embryo. The post-implantation epiblast is characterised by expression of pluripotency factors, such as *Sox2*, but also of early lineage specification markers, including *Tbra* and *Sox1*. Importantly, NMP regions are marked by coexpression of *Tbra/Sox2*, while later on in development, *Sox1* expression is detected too. EpiSCs constitute the *in vitro* equivalent of the gastrula stage epiblast, while activation of Fgf and Wnt signalling in EpiSCs allows us to capture the NMP state *in vitro*. Consistent with the gene activity observed in the mouse embryo, *Tbra*, *Sox2* and *Sox1* are active in self-renewing EpiSCs, but also in EpiSC-derived NMPs.

In this study, I investigate how the expression pattern of these three transcription factors changes as the pluripotent EpiSCs turn into NMPs and I dissect the properties of the subpopulations that is marked by expression of *Tbra* and *Sox1*. Specifically, in Chapter 3, I focus on EpiSCs under self-renewing conditions and I investigate the characteristics of Tbra positive cells. In Chapter 4, I analyse the correlation of the expression patterns of *Tbra*, *Sox2* and *Sox1* in EpiSCs and NMPs and I dissect the characteristics of the Sox1 positive cells. The gene activity recorder in the caudal progenitor regions suggests that a *Tbra/Sox2* or a *Tbra/Sox1* double reporter line would a be a very useful tool that would allow us to track and isolate NMPs *in vivo* and *in vitro*. Hence, in Chapter 5, I present the generation of a *Tbra* mouse ES cell reporter line, in which targeting of *Sox2* locus has been attempted and in the future, targeting of the *Sox1* gene could be performed too.

Chapter 2: Materials and Methods

2.1. Cell culture

2.1.1. Cell culture materials

Phosphate-buffered saline (PBS) (Sigma, cat. D8537)

Components of ES cell media hereinafter termed LIF/FCS/GMEMβ are added in the order as listed below:

- 500mL - Glasgow minimum essential medium (GMEM) (Sigma, cat. G5154)
- 51mL - fetal calf serum (FCS) (Gibco, cat. 10270-106)
- 11mL - 50× glutamine/pyruvate stock solution (see recipe below)
- 5.7mL - 100× MEM non-essential amino acids (Invitrogen, cat. 11140-035)
- 570µL - 100mM (1000×) β-mercaptoethanol stock solution (see recipe below)
- 570µL - 100'000U/mL (1000×) LIF supplement (made in-house)

Glutamine/pyruvate stock solution comprised the following per 11mL:

- 5.5mL - 100mM sodium pyruvate (Invitrogen, cat. 11360-039)
- 5.5mL - 200mM L-glutamine (Invitrogen, cat. 25030-024)

100mM β-mercaptoethanol stock solution comprised the following:

- 200µL - 14.3M β-mercaptoethanol (Sigma, cat. M6250)
- 28.2mL - ultra-high purity (UHP) H_2O (Millipore, Milli-Q)

0.1% w/v gelatin solution comprised the following:

- 55mL - autoclaved 1% w/v gelatin stock solution (Sigma, cat. G1890) in UHP-H_2O.
- 500mL - PBS

Before use, the solution is incubated at 37°C for 30min to dissolve.

0.025% v/v Trypsin stock solution comprised the following:

- 500mL - solution of 0.186g EDTA (Sigma, cat. E5134) in PBS, 0.22μm filtered
- 5mL - chick serum (Sigma, cat. C5405)
- 5mL - concentrated 2.5% trypsin stock (Invitrogen, cat 15090-046)

Components of chemically defined medium (CDM) hereinafter termed N2B27 are added in the order as listed below:

- 100mL - Dulbecco's modified eagle medium (DMEM):F12 (1:1 v/v, no L-glu) (Gibco, cat. 12634-010)
- 100mL - Neurobasal (Gibco, cat. 21103-049)
- 2mL - 100× MEM non-essential amino acids (Invitrogen, cat. 11140-035)
- 2mL - 100× L-glutamine solution (Invitrogen, cat. 25030-024)
- 1mL - 100× N2 supplement (Gibco, cat. 17502-048) or home-made where indicated (see recipe below)
- 4mL - 50× B27 supplement (Gibco, cat. 17504-044)
- 200μL - 100mM (1000×) β-mercaptoethanol stock solution (see recipe above)

Where indicated, EpiSC CDM termed Activin/FGF/N2B27 are supplemented with:

- 20ng/mL (final) - Human Activin A (PeproTech, cat. 120-14E)
- 10ng/mL (final) - Human FGF2 (R&D Systems, cat. 233-FB-025/CF)

7.5µg/mL fibronectin solution comprised the following:

- 75µL - 1mg/mL fibronectin from bovine plasma (Sigma, cat. F1141)
- 10mL - PBS

Table 2.1: Tissue culture flasks, dishes and plates used.

Name / Type	Area (cm^2)	Cat.
T25 / flask	25.0	Corning 430168
T75 / flask	73.7	Corning 430720
T150 / flask	149.1	Corning 430823
10cm / dish	59.2	Sigma SIAL0599
4-wells / plate	1.89	Nunc 176740
6-wells / plate	9.62	Sigma SIAL0516
12-wells / plate	3.80	Sigma SIAL0513
24-wells / plate	1.89	Sigma SIAL0526
48-wells / plate	0.95	Sigma SIAL0548
96-wells / plate	0.26	Sigma SIAL0596

Table 2.2: Drugs used in mouse ES cultures for positive or negative selection.

Compound	Final concentration	Cat.
Puromycin	1µg/mL	Sigma, cat. P8833
Hygromycin B	100µg/mL	Roche, cat. 10843555001
Blasticidin S HCl	5µg/mL	Invitrogen, Cat. R210-01
Geneticin (G418)	200µg/mL	Gibco, cat. 11811-031
Ganciclovir	3µM	Invivogen, cat. sud-gcv

2.1.2. ES cell passaging

ES cells were cultured in LIF/FCS/GMEMβ medium, in flasks or plates pre-coated with 0.1% w/v gelatin and were incubated in a 37°C / 7% CO_2 humidified incubator. Medium was changed every second day and the cells were passaged when they reached 70-80% confluence. Cells were passaged by rinsing with pre-warmed PBS and treated with pre-warmed 0.025% v/v trypsin at 37°C for 1 to 2 minutes until cells fully detach by firmly tapping the flask. At least 5 volumes of LIF/FCS/GMEMβ was added to the flask, washing down the cells and blocking the reaction. Cells were pelleted in a 30mL universal tube (Sterilin, cat. 128A/FS) by centrifugation for 3 minutes at 290×g (Eppendorf, cat. 5702) before resuspended in LIF/FCS/GMEMβ media. Cells were split between 1:8 and 1:15 at each passage depending on the cell line.

2.1.3. EpiSC passaging

EpiSCs were maintained in Activin/FGF/N2B27 medium on plates coated with 7.5μg/mL bovine fibronectin and were incubated in a 37°C / 7% CO_2 humidified incubator. EpiSCs were passaged when they reached 70-80% confluence. Cells were passaged by washing with pre-warmed PBS and treated with 300μL of pre-warmed 1× accutase solution (Sigma, cat. A6964) at 37°C for 5 minutes. 3ml of N2B27 was added to wash down the cells and the cell suspension was centrifuged for 3 minutes at 290×g. Cell pellets were gently resuspended in 1000μL. Cell counting was performed when a specific plating cell density was required, using a hematocytometer (Neubauer), otherwise cells were split between 1:8 and 1:10 at each passage.

2.1.4. Cell freezing

ES cells and EpiSCs were frozen in CryoTube (Nunc, cat. 377224) vials. Cell pellets were collected by treatment with trypsin (ES cells) or accutase (EpiSCs) as described

in section 2.1.2 or 2.1.3 respectively and resuspended in 1mL of 10% v/v DMSO in LIF/FCS/GMEMβ (ES cells) or in knock-out serum replacement (KOSR) (Gibco, cat. 10828-028) (EpiSCs). Vials were immediately placed on dry-ice and transferred to a –80°C freezer. The following day vials were transferred to a liquid N2 tank for long term storage.

2.1.5. EpiSC differentiation assay from ES cells

ES cells were plated on fibronectin-coated plates at approximately 8320 cells/cm^2 (or the equivalent of 8×10^4 cells in a well of a 6-wells plate) in LIF/FCS/GMEMβ. Medium was removed 24 hours later and cells were washed with pre-warmed PBS before Activin/FGF/N2B27 was added. Differentiating ES cells were cultured in Activin/FGF/N2B27 continuously thereafter and were passaged to new fibronectin-coated plates as described in section 2.1.3, except that the split ratio was between 1:4 and 1:5 for the first 5 - 6 passages. When cells reached passage 8, 0.6μM Jack inhibitor (CALBIOCHEM, cat. 420099) was added in the Activin/FGF/N2B27 until the next passage, in order to eliminate any undifferentiated ES cells. Homogenous population of EpiSCs can be derived within 10 passages. EpiSC populations were tested by assessing the expression of core pluripotency factors and early differentiation markers, performing quantitative real-time PCR (QRT-PCR) (section 2.6.2) and immunocytochemistry (section 2.4.1) or flow cytometry (section 2.1.7) for the analysis of cell-surface markers.

2.1.6. Generation of *in vitro* neuromesodermal progenitors (NMPs) from mouse EpiSCs.

The generation of *in vitro* neuromesodermal progenitors (NMPs) was performed following the protocol described in (Gouti *et al.*, 2014). Mouse EpiSCs were plated on fibronectin-coated plates at 2100 cells/cm^2 in N2B27 medium, supplemented with 20ng/mL human FGF2 and 3μM Wnt agonist CHIR99021 (CHI) (Axon, cat. 1386),

hereinafter termed FGF/CHI. EpiSCs were cultured in these conditions for 48 or 72 hours, giving rise to NMPs that hereinafter will be termed 48h NMPs and 72h NMPs respectively. During the generation of 72h NMPs, the medium was changed once and fresh one was added 48 hours after the cells were plated.

2.1.7. Flow cytometry

Cultured ES cells or EpiSCs were collected as described in 2.1.2 or 2.1.3 respectively and the cell pellet was resuspended in 10% v/v FCS in PBS (FACS buffer). Cells were passed through a cell strainer-containing 5mL tube (Falcon, cat. 352235) and kept on ice. To distinguish live and dead cells, a final concentration of 0.1μg/mL DAPI (Molecular Probes, cat. D1306) was added to each sample. Cell analysis was performed using a LSRFortessa (Becton Dickinson) or a BD FACS Calibur (Becton Dickinson) cell analysers. Cell sorting was performed by Fiona Rossi and Dr. Claire Cryer, using a FACSAria II (Becton Dickinson) cell sorter. The lasers used to excite each fluorophore are shown in table 2.3.

Table 2.3: Excitation and emission wavelengths of fluorophores.

Fluorophore	**Laser**	**Emission**
GFP	488nm	B525/50
tdTomato	561nm	YG582/15
dsRed2	561nm	YG582/15
DAPI	405nm	V450/50
APC	633nm	R661/16

Quantification of the cell-surface marker Platelet and Endothelial Cell Adhesion Molecule 1 (Pecam-1)/CD31 was performed by staining cells in FACS buffer. 10^6 cells were washed twice with ice-cold PBS before resuspended in 500μL of FACS buffer containing 0.2 μg/ml allophycocyanin (APC) rat anti-mouse Pecam-1 antibody. Cells were incubated at 4°C, in dark, for 20 minutes and then washed twice with FACS buffer. The stained cells were prepared for cell analysis as described above.

2.1.8. Electroporation of plasmid DNA into mouse ES cells

Cells were collected by trypsinisation as described in section 2.1.2. Aliquots of 10^7 cells were prepared for each electroporation sample and centrifuged at 800rpm for 5 minutes. The cell pellets were washed twice by resuspending with 10mL of PBS and spinning down at 800rpm for 5 minutes. For each transfection, cells were resuspended in 600μL of PBS and transferred to a 400μm gap electroporation cuvette (BioRad, cat. 1652088). 50 to 100μg of linearised DNA (section 2.5.3) dissolved in 200μL of PBS was added to the cells and the mix was left at RT^{o}C for 3 minutes. PBS only without DNA was added to cells for negative control. The cuvette was placed in a Gene Pulser (BioRad, cat. 1652076) apparatus and the cells are electroporated at 0.8kV with 3μF resistance. Using a plugged Pasteur pipette, cells were transferred gently from the cuvette to a pre-warmed universal tube containing 9.2mL of LIF/FCS/GMEMβ. Cells are replated on pre-warmed gelatin-coated dishes at a density of $1.7 \times 10^5/cm^2$ (equivalent to 10^6 per 10cm dish). To select for the ES cells having the transfected DNA construct integrated into their genome, relevant drugs were added to the medium 24 hours later.

Cells were cultured for approximately 7 days until no viable cells remained in the negative control and visible colonies appeared in dishes containing transfected cells. Colonies were washed with pre-warmed PBS and picked using a wide-pore pipette tip containing 10 to 20μL of 0.025% trypsin solution inside. Colonies were transferred to a well of a 96 wells-plate containing 150 to 200μL of LIF/FCS/GMEMβ with the

appropriate drug selection and mixed vigorously to break up cell clumps. Medium was changed 24h later. Confluent wells were trypsinised and expanded into 24, followed by 6 wells- plate in a 1:1 split. Cells were removed from drug selection and frozen when grown in 6 wells-plate as described in section 2.1.4.

2.1.9. Nucleofection of plasmid DNA into mouse ES cells

Nucleofection of ES cells was performed using Amaxa Mouse ES cells nucleofection kit (LONZA, cat. VPH-1001). Cells were collected by trypsinisation as described in section 2.1.2. Aliquots of 3-5 × 10^6 cells were prepared for each nucleofection sample followed by centrifugation. The cells were washed by resuspending the pellet with 10mL of PBS and spinning down at 800rpm for 5 minutes. After two washes, the cell pellet was resuspended in 90µL supplemented Mouse ES Cell Nucleofector solution. 10µg of DNA was mixed with supplemented Mouse ES Cell Nucleofector solution to a total volume of 10µL. The cell suspension was combined with the DNA sample and the mixture was transferred into an amaxa certified cuvette. Supplemented Mouse ES Cell Nucleofector solution only without DNA was added to cells for negative control. Nucleofection was performed using the 4D nucleofector apparatus (LONZA), applying the mouse ES cell program A-023. When nucleofection was completed, 500µL of pre-warmed culture medium was added to the cuvette immediately and the cells were transferred to universal tubes containing 1400µL prewarmed LIF/FCS/GMEMβ, using a Pasteur pipette that is provided by LONZA in the nucleofection kit. 1000µL of the cell suspension was transferred to gelatin coated 10cm dishes containing prewarmed LIF/FCS/GMEMβ.

When the transfection reaction included plasmid DNA that involved the constitutive expression of a fluorescent protein, cell cultures were harvested 36 - 48 hours later (section 2.1.2) and transfected cells were isolated through FACS sorting (section 2.1.7). Sorted mouse ES cells were replated on gelatin coated dishes, in

LIF/FCS/GMEMβ, at 1100 cells/cm^2 (equivalent to 65000 cells per 10cm dish) and 24 hours later, selection with a relevant drug was initiated.

2.1.10. Lipofection of plasmid DNA into mouse ES cells and EpiSCs

ES cell and EpiSC were collected as described in section 2.1.2 and 2.1.3 respectively. 30 minutes before transfection, 10^6 cells were replated on a well of a 6 wells-plate coated with gelatin in 2mL of LIF/FCS/GMEMβ (ES cells) or coated with fibronectin in 2mL of Activin/FGF/N2B27 (EpiSCs) and incubated at 37°C. For each transfection, ≤3μg of circular pDNA prepared by MidiPrep or MaxiPrep (section 2.5.8) was mixed with 5μL P300 Reagent (Invitrogen, cat. L3000008) and 125μL of LIF/GMEMβ without FCS (ES cells) or DMEM/F12 (EpiSCs), while 5μL of Lipofectamine 3000 Reagent (Invitrogen, cat. L3000008) was added to 125μL of LIF/GMEMβ without FCS (ES cells) or DMEM/F12 (EpiSCs). After incubation at room temperature (RT°C) for 5 minutes, the DNA/P300 reagent mixture was combined with the lipofectamine mixture and incubated at RT°C for further 10 minutes. The DNA/lipofectamine mixture was subsequently added drop-wise to the cells. The transfection efficiency was estimated by FACS analysis 24 - 48 hours later (section 2.1.7). Selection of the transfected cells was performed by adding relevant drugs to the medium (table 2.2) or by cell sorting (section 2.1.7) 24 to 48 hours post transfection.

2.1.11. Excision of Flp recombinase recognition target (FRT-) flanked constructs from the mouse ES cell genome

In order to excise FRT-flanked constructs from the genome of mouse ES cells, ES cells were lipofected as described in 2.1.10, with a plasmid containing the gene of Flp recombinase downstream of the pCAG promoter, which is then joined, via an internal ribosome entry site (IRES), to the gene encoding for puromycin N-acetyltransferase

(Pac) (pCAG-Flp::IRES::Pac). Selection of colonies that had the FRT-flanked construct excised from their genome was performed following either protocol:

Protocol 1: 24 hours post transfection, the cells were replated on gelatin coated 10cm dishes, in LIF/FCS/GMEMβ, at a density 360 cells/cm^2. Non transfected ES cells were plated under the same conditions as control. The FRT-flanked constructs used in this study included an expression cassette for thymidine kinase from herpes simplex virus (HSV-TK), allowing for negative selection. To select the cells that had the FRT-flanked construct excised from their genome, ganciclovir was added to the ES cell culture medium 24 hours following replating. Cells were incubated in ganciclovir for 7-11 days, colonies were picked and expanded. To confirm the excision of the construct, gDNA was extracted from each clone (section 2.5.10) and analysed performing PCR analysis (section 2.5.1).

Protocol 2: 24 hours post transfection, cells were harvested from the wells of a 12-well plate, resuspended and replated at 3 different densities; 38%, 8% and 0.8% of each cell suspension was replated in 10cm dishes in LIF/FCS/GMEMβ medium. Non transfected cells, plated under exactly the same condition, were used as control. To select for the cells that have been successfully transfected, the ES cell medium was supplemented with 1μg/ml puromycin. Puromycin selection was applied for 48 hours only, and thereafter, cells were cultured without the presence of any antibiotics. Colonies were picked and expanded. The excision of the FRT-flanked construct was verified by PCR (section 2.5.1), analyzing gDNA extracted from each clone (section 2.5.10).

2.2. Embryology

2.2.1. Animal husbandry

Wildtype MF1, 129/Ola and transgenic mice strains were bred and housed within the animal unit of Medical Research Council (MRC) Centre for Regenerative Medicine (CRM) according to the provisions of the Animals (Scientific Procedures) Act (1986). All mice were maintained in a stabilized environment on a 10 hours light, 14 hours dark cycle. Embryos at specific developmental stages were obtained by setting up timed matings overnight and inspecting the presence of vaginal plugs the following morning. Embryonic day (E) 0.5 was designated to the day at noon when plugs were observed, assuming copulation occurred at midnight.

2.2.2. Collection of post-implantation embryos

Pregnant females were culled by cervical dislocation in the animal unit and were dissected within the CRM facility. The uterus was extracted from the abdominal cavity and deciduae were removed from the uterus in Ø0.22μm sterile filtered (Millipore, cat. SLGP033RB) M2 medium (Sigma, cat. M7167) at RT°C. Embryos were dissected out of the decidua in M2 medium under a zoom stereo microscope (Olympus, cat. SZH-121).

2.2.3. Morula aggregation and examination of chimeric embryos

Morula aggregation was performed by the MRC CRM transgenic service. Super ovulated female mice (CBA/BL6) were mated with males (CBA/BL6) and 8-cell staged embryos were collected and subsequently released from zona pellucida. Cells were prepared 24 hours prior to morula aggregation; 10^6 cells were plated on a well of

a 6-well plate and the cell suspension was serially diluted 1:2 through the remaining 5 wells of the 6-well plate. On the day of the aggregation setup, cells were briefly treated with trypsin and individual colonies of 6-8 cells were picked up and directly incubated with the embryos. Dishes containing the aggregates were incubated in KSOM-AA medium, at 37°C for 24 hours until embryos reached the early-blastocyst stage. Uterine transfers of aggregates into anesthetized pseudopregnant females was performed in the animal unit of the MRC CRM. The day of uterine transfer is designated E2.5.

Embryos were collected as described in 2.2.2 and imaged using a fluorescent stereomicroscope (Nikon, AZ100) attached to a colour camera (QImaging, EXi-Blue). Chimeric embryos were subsequently fixed in 4% v/v PFA/PBS for 3 hours or overnight in a 4°C fridge. Assessment of the contribution of the transgenic cells to the embryonic tissues was performed by cryostat sectioning of the chimeric embryos (section 2.3.1), followed by immunohistochemistry (section 2.4.3).

2.2.4. Blastocyst injection and transgenic mouse line generation

Blastocyst injection was performed by the MRC CRM transgenic service. Female mice (BL6) were mated with males (BL6) and E3.5 blastocysts were collected. Cells were prepared 48 hours prior to blastocyst injection; 10^6 cells were plated on a well of a 6-well plate and the cell suspension was serially diluted 1:2 through the remaining 5 wells of the 6-well plate. On the day of blastocyst injection, cells were briefly trypsinised and prepared to single cell suspension. Each blastocyst was injected with 10-12 cells and incubated for 1 hour at 37^{o}C, prior to transfer to pseudopregnant female mice (CBA/BL6).

The day of uterine transfer was designated E2.5. Pups were born 17-18 days later and the coat color was observed 8-12 days after birth. Chimeras were identified by coat color chimerism (Agouti phenotype). Once sexually mature, approximately 6 weeks

old, male chimeras were bred with female mice (CBA/BL6), so as to initiate the generation of the relevant transgenic mouse line.

2.2.5. Grafting of *in vitro* NMPs into E8.5 embryos and *ex vivo* embryo culture

Grafting of cells and *ex vivo* embryo culture was performed by Dr. Yali Huang as described in (Huang *et al.*, 2012), using genetically modified EpiSC in which a fluorescent protein was constitutively express. EpiSCs were differentiated into 48 and 72 hours NMPs (section 2.1.6) and E8.5 mouse embryos were collected (section 2.2.2). Grafting was performed by hand in a dissecting stereomicroscope using a hand-pulled micropipette. Cells were scraped from culture dishes using a 20–200ml pipette tip. The resulting cell clumps were placed close to the embryos, and a small clump was drawn into the micropipette by gentle suction. The embryo was held loosely in place with forceps while the micropipette was inserted into the NSB to create an opening. Cells were then gently expelled as the micropipette was drawn out of the embryo, leaving a short string of cells lodged in the epiblast.

Embryos were cultured according to Copp *et al.* (1990), as described in Huang *et al.* (2012). E8.5 embryos were cultured in 50% rat serum culture medium (one embryo per 1ml medium) in a roller culture apparatus (B.T.C Engineering) supplied with 5% CO_2 in air (BOC) at 37˚C. Alternatively, they could be cultured in a round-bottomed universal tube with a screw cap (Nunc) where the thread was coated with vacuum grease (Baysilone-paste, Bayer) to create an airtight seal. Tubes were gassed with 5% CO_2 in air for 1min. Up to two embryos were cultured per vial. The universals were then placed inside a roller incubator (BTC Engineering) at a slight angle from the horizontal at 37˚C. After 24 hours in culture, embryo development was checked. Embryos which had undergone turning were cultured for another 24 hours in 75% rat serum culture medium (1ml medium per embryo) in 5% CO_2, 40% O_2 balance N_2 (BOC) at 37˚C. After culture, the embryos were dissected in M2 medium

(SigmaLAldrich) from their yolk sac, amnion and developing placenta and imaged using an inverted fluorescence stereomicroscope (Olympus, IX-51) attached to a monochrome camera (QImaging, Retiga 2000R). The embryos of interest were fixed in PFA, embedded in sucrose/gelatin solutions and instantly frozen, as described in 2.3.1. Incorporation of the grafted cells was assessed by sectioning the gelatin-embedded embryos (section 2.3.1), followed by immunohistochemistry (section 2.4.3).

2.3. Histology

Methods described in this section were adapted from (Bancroft and Gamble, 2002).

2.3.1. Cryostat sectioning

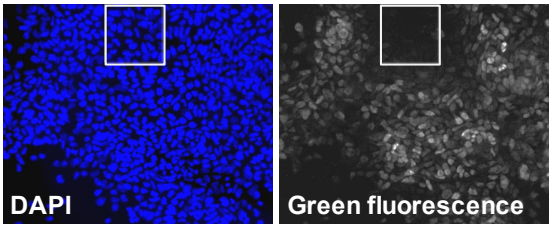
Sectioning of gelatin-embedded blocks was performed by Ronald Wilkie. Embryos were fixed in 4% v/v PFA/PBS for 3 hours or overnight in a 4°C fridge. Fixed embryos were washed 3 times in PBS and transferred to bijou vials (Sterilin, cat. 129A) containing 15% w/v sucrose (Sigma, S0389) solution for 4 hours to overnight, depending on embryo size. Embryos were carefully transferred using plastic Pasteur pipettes into liquefied 7% w/v gelatin (Sigma, cat. 48722), 15% w/v sucrose solution at 37°C and were kept at 37°C until embryos sink to the bottom of the vial. Blocks were setup by applying labelling tape (Anachem, cat. SL9355) and sealed with Parafilm M around a metallic cylinder base (Ø12mm) to create a chamber. Sunken embryos were carefully transferred into pre-warmed, gelatin/sucrose solution-filled blocks and were orientated underneath a dissecting microscope with an overlying light-source using dentist pick tools. Gelatin/sucrose solution was allowed to solidify on ice before the blocks were rapidly frozen by holding the metallic base at the surface of liquid N_2. Blocks were stored at –80°C. Prior to sectioning, samples were equilibrated overnight in a cryostat (Leica, CM1900) chamber. Sections had a thickness of 7µm, were placed onto glass slides and stored at –80°C.

2.4. Immunofluorescence staining (IFS)

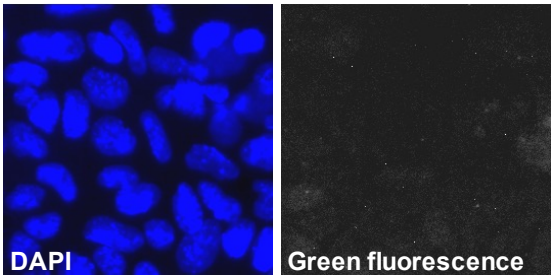
2.4.1. Immunocytochemistry on cultured cells

Cultured cells were fixed on 12 wells-plate with fresh 4% v/v paraformaldehyde (PFA) (Sigma, cat. P6148) in PBS at RT°C for 10 minutes. PFA was removed and fixed cells were washed twice with PBS. 0.1% v/v Triton-X100/PBS (PBStr) (Amersham Biosciences, cat. 17-1315-01) were used to permeabilize fixed cells for 10 minutes and excess aldehyde fixatives were blocked with 0.5M glycine pH 7.5/PBStr for 15 minutes. Cells were washed twice with PBStr and subsequently blocked with blocking buffer, which consisted of 3% v/v donkey serum (Sigma, D9663) and 1% v/v bovine serum albumin (BSA) (Sigma, A8412) in PBStr, for 1 to 2 hours at RT°C or at 4°C overnight. Cells were incubated with primary antibodies (Table 2.4) overnight, at 4°C. Cells were then quickly washed with PBStr three times, followed by four 25-minute washes, on a slow orbital shaker. Relevant fluorophore-conjugated secondary antibodies (Table 2.5) diluted in blocking buffer were added to the cells for 2 to 3 hours, in the dark, at RT°C. Cells were washed with PBStr as described above and incubated with 1µg/mL DAPI/PBS at RT°C for 5 minutes. Stained cells were kept in PBS, in the dark, at 4°C and imaging was performed using an inverted fluorescence stereomicroscope (Olympus, IX-51) attached to a monochrome camera (QImaging, Retiga 2000R). Image quantification was performed using a semi-automated image analysis pipeline developed by Guillaume Blin, as described in Osorno *et al.* (2012). Figure 2.1 illustrates an example of how the levels of background fluorescence signal were estimated (step 1-3). In the work presented in the next chapters, the gating settings were estimated analysing negative areas of 1-3 different fields; examples of fluorescence negative areas are presented throughout the result chapters. The final image quantification (as shown in figure 2.1, step 4) was performed analysing 8-10 different fields.

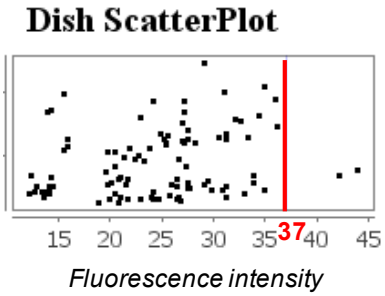
Step 1: Following immunocytochemistry and imaging, find a field containing cells that appear fluorescence negative (here marked by a white square).



Step 2: Crop the area containing the fluorescence negative cells.

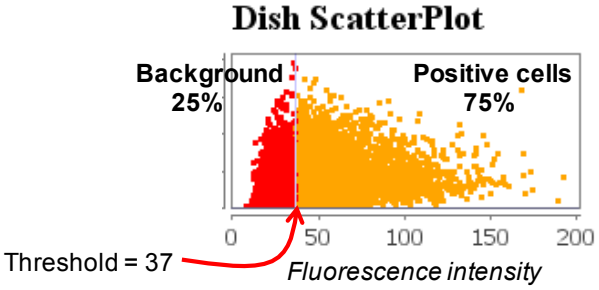


Step 3: Perform single cell immunofluorescence quantification of the cropped field.



Defining background levels of fluorescence signal

Step 4: Perform single cell immunofluorescence quantification, including all the images you wish to analyse. Use the fluorescence levels detected in the negative area (step 3) as the threshold to distinguish the background noise from the real fluorescence signal.



Quantification of fluorescence signal in images of interest

Figure 2.1: Gating settings in fluorescence image quantification.

Figure 2.1: Gating settings in fluorescence image quantification. Steps 1-3 describe the fluorescence signal quantification in a group of cells that appear negative. Step 3 shows the fluorescence intensity in each individual cell. In this example, the cut off was set at 37, as the vast majority of cells were characterised by values <37; the two cells found on the extreme right of the graph appeared quiet distant from the rest of the cells, hence it was considered that probably they are positive (with very low levels of fluorescence). This cut off was then used to quantify the number of positive and negative cells in the group of images under analysis (step 4).

2.4.2. Whole-mount immunohistochemistry

The procedure of whole-mount staining is very similar to the protocol described above for cultured cells except for the following modifications. Fixation was performed with 4% PFA/PBS at 4°C for 2 to 3 hours depending on embryo size. Embryos were permeabilized using 0.5% v/v Triton/PBS and fixatives were blocked with 1M glycine pH 7.5/PBStr. Blocking was performed overnight on a pivotal rocker in a 4°C room. Fresh primary antibodies were incubated for 24 to 48 hours in the same conditions as blocking. Secondary antibodies were incubated for 3 to 5 hours at RT^{o}C and DAPI for 20 minutes. Embryos were stored in 4°C in PBST until further analyzed. Stained embryos were imaged using an inverted fluorescence stereomicroscope (Olympus, IX-51) attached to a monochrome camera (QImaging, Retiga 2000R).

2.4.3. Immunohistochemistry on cryostat sections

Slides containing cryostat sections were thawed at RT^{o}C for 10 minutes. Individual sections were outlined with an ImmEdge hydrophobic barrier pen (Vector Labs, cat. H-4000) to keep reagents localized on tissue specimen. All steps were conducted by adding small droplets onto outlined sections in a humidified box in the dark at RT°C,

unless stated otherwise. Sections were stained using reagents similarly as described in 2.4.1, except fixation was not performed. Excess fixative were blocked with 1M glycine for 20 minutes before non-specific antigens were blocked with blocking buffer for 2 hours. Sections were incubated with primary antibodies (Table 2.4) at RT°C for 24 to 48 hours. Slides were washed with PBS 3 times, for 10 minutes each wash, and then incubated with relevant secondary antibodies (Table 2.5) for 3 hours. Excess of antibodies were removed through PBS washes, the last one of which contained DAPI. Excess of solution was removed and sections were mounted in Vectashield non-hardening medium (Vector Labs, cat. H-1200) with a cover-slip. Mounted sections were stored at 4°C and imaged using an upright fluorescence stereomicroscope (Olympus, BX-61) attached to a colour camera (Hamamatsu, ORCA-ER).

When use of this protocol did not result to a clear and strong fluorescence signal, antigen retrieval was performed prior to staining, by boiling the sections in 10mM sodium citrate pH 6.0, using a standard microwave, at maximum power for 4 × 2.5 minutes. Every 2.5 minutes the sodium citrate was left to cool down for 10 seconds and any volume lost by evaporation was replaced with DI-H_2O.

Table 2.4: Primary antibodies used in immunofluorescence staining.

Epitope	Species	Specificity	Dilution (original concentration)	Cat.
Sox2	Rabbit	Monoclonal	1:200 (0.131mg/ml)	Abcam, ab92494
Nanog	Rat	Monoclonal	1:200 (0.5mg/ml)	eBioscience, 14-5761
Oct4	Goat	Polyclonal	1:400 (0.2mg/ml)	Santa Cruz, sc8628
GFP	Mouse	Monoclonal	1:1000 (NA)	Abcam, ab38689
GFP	Chicken	Polyclonal	1:1000 (10mg/ml)	Abcam, ab13970
Cdh2	Mouse	Monoclonal	1:40 (5 μg/ml)	BD Biosciences, cat. 8C11
Tbra	Goat	Polyclonal	1:200 (0.2 mg/mL)	R&D, cat. AF2085
Foxa2	Rabbit	Polyclonal	1:200 (NA)	Abcam, ab40874
mCherry/ tdTomato	Goat	Polyclonal	1:100 (3mg/ml)	SECGEN, AB0081-200

Table 2.5: Fluorophore-conjugated secondary antibodies used.

Epitope of donkey antibody	Fluorophore conjugation (cat.)		
	Alexa-488	**Alexa-568**	**Alexa-647 / 640R**
Mouse IgG	Invitrogen, A21202	Invitrogen, A10037	Invitrogen, A31571
Rabbit IgG	Invitrogen, A21206	Invitrogen, A10042	Invitrogen, A31573
Goat IgG	Invitrogen, A11055	Invitrogen, A11057	Invitrogen, A21447
Rat IgG	Invitrogen, A21208	Sigma, SAB4600077	Sigma, SAB4600156

2.5. Molecular biology

2.5.1. *In vitro* amplification of DNA molecules

Amplification of gDNA regions and inserts in vector construction was performed by *in vitro* polymerase chain reaction (PCR) using 1 unit per 50µL reaction of Taq (New England Biolabs, cat. M0273), Phusion (Finnzymes, cat. F-530L) or Q5 (New England Biolabs, cat. M0491L) high-fidelity DNA polymerases. Each PCR reaction was setup in ice, according to the manufacturer's instructions, in a total volume of 50 - 100 µL and included the following components:

- 200µM (final) - dNTP (Invitrogen, cat. 10297-018)
- 0.5µM (final) - synthetic DNA oligonucleotide primers (Integrated DNA Technologies)
- 10 - 30ng - pDNA (section 2.5.8); or
- 100ng - gDNA (section 2.5.10)

Amplification reactions were performed using a DNA Engine thermal cycler (BioRad, cat. PTC-200), adjusting the thermocycling conditions according to the manufacturer's instructions of each polymerase. DNA amplicons were visualized by electrophoresis using 1 to 2% w/v ultra-pure agarose (Invitrogen, cat. 16500-500) gels in 0.5× Tris/borate/EDTA (TBE) containing a final concentration of 0.3μg/mL ethidium bromide (Sigma, E1510). 10× TBE (Sigma, cat. 93290) comprised 1.3M Tris, 450mM boric acid and 25mM EDTA in DI-H_2O. Estimation of the size of the PCR amplicons was performed by using an 1 kb Plus DNA Ladder (Invitrogen, cat. 10787018), an 1 kb DNA ladder (New England Biolabs, cat. N3232S) or a 100 bp DNA ladder (New England Biolabs, cat. N3231S), depending on the size of the PCR product of interest.

2.5.2. DNA purification

PCR products were purified using MinElute PCR purification kit (Qiagen, cat. 28004) according to the manufacturer's instructions. When DNA was separated in agarose gels by electrophoresis (2.5.1), a gel slice containing the desired DNA band was excised using a scalpel and a UV transilluminator (Herolab, cat. UVT-14 M) and was subject to DNA purification using the MinElute Gel Extraction kit (Qiagen, cat. 28604) according to the manufacturer's instructions.

Where indicated, DNA was also purified by adding 0.1 volume of 3M sodium acetate/H_2O (NaOAc) pH 5.2 solution to 1 volume of DNA sample before adding 2.5 volumes of cold (−20°C) 100% ethanol to precipitate double-stranded DNA. The DNA/ethanol mixture was placed in dry ice for 30 minutes before DNA was pelleted by centrifugation at 14000×g for 30 minutes. The pellet was washed and centrifuged with fresh 75% ethanol/H_2O solution. Purified DNA was obtained by air drying the pellet before its opacity reached transparency and an appropriate volume of UHP-H_2O or PBS was added to dissolve the pellet.

2.5.3. Restriction endonuclease digestion

Site-specific digestion of DNA for vector construction or diagnostic purpose was setup in a 50µL reaction volume comprising 1 to 2µg of DNA and an excess of restriction endonucleases (but ≤ 5% v/v) (New England Biolabs, Roche or Thermo Scientific). Reactions were performed using PCR thermal cyclers in conditions recommended by the manufacturer for at least 1 hour. Digested DNA bands were separated by agarose gel electrophoresis as described in 2.5.1 and, where applicable, purified from the gel as described in 2.5.2.

Linearization of pDNA for stable transfection (section 2.1.8 and 2.1.9) was performed using 50 to 100µg of pDNA. 10µL (2.5% v/v) of restriction enzyme was added for the first hour of incubation before supplementing with 10µL of restriction enzyme (typically an excess of 50 to 100 units) bringing the total volume of the reaction to 400µL. The reaction was performed overnight in conditions according to manufacturer's instructions. Linear DNA was purified by ethanol precipitation as described in section 2.5.2 and dissolved in 100 or 200µL of PBS inside tissue culture facilities.

2.5.4. DNA fragment ligation

The concentrations of restriction digestion fragments purified from an agarose gel following electrophoresis (section 2.5.2) were quantified using a micro-volume spectrophotometer (Nanodrop, ND-1000). Ligation reaction was setup in a total reaction volume of 10µL comprising 25 to 50ng of linearized vector DNA, 3-fold molar excess of insert DNA fragments and 0.5µL of T4 DNA ligase (New England Biolabs, cat. M0202L). Reactions were incubated at RT°C for 1 hour or at 16°C overnight using a thermal cycler. 4µL of purified PCR products (sections 2.5.1 and 2.5.2) were ligated to Zero Blunt TOPO vectors (Invitrogen, cat. 450245) according to manufacturer's instructions. The reaction was performed at RT°C for 10 to 15 minutes.

2.5.5. Preparation of chemically competent *E.coli* cells

Chemically competent cells were prepared using SOB medium and transformation buffer (TB). SOB medium consisted of LB supplemented with 2.5mM KCl. Following autoclaving, $MgCl_2$ was added to that at a final concentration of 20mM. TB medium was prepared by adding 3mg/ml Piperazine-N,N′-bis (2-ethanesulfonic acid) (PIPES), 2.2mg/ml Cacl2-2H_2O, 18.6mg/ml KCl in ddH_2O and then adjusting the pH to 6.7-6.8 using 5N KOH. 10.9mg/ml $MnCl_2$-4H_2O was added to that and the medium was sterilized using bottle-top filter.

An inoculate of a single colony from a plate with DH5a was picked and seeded into SOB medium. Cultures were incubated at 18°C until its optical density at a wavelength of 600nm (OD_{600}) reached 0.5 to 0.8 absorption units (AU). Cultures were then placed in ice for 10 minutes, followed by centrifugation at 3000rpm at 4°C for 15 minutes. The bacteria pellet was resuspended in chilled TB, using 33.5mL of TB per 100mL of initial bacterial culture. Following a 15-minute centrifugation at the same conditions as before, bacteria were resuspended in chilled TB, using 8ml TB per 100mL of initial bacterial culture. DMSO was added to the mix at a final concentration 7% and bacteria were incubated in ice for 10 minutes. The rest of the procedure was performed in a 4°C cold room. Aliquots of 50mL of the cell suspension were prepared and placed in liquid N_2 immediately. Competent cells were stored at -80°C.

2.5.6. Cloning of ligated plasmid DNA in *E.coli*

Ligated plasmid vectors (section 2.5.4) were transformed into chemically competent *E.coli* (section 2.5.5) by incubation on ice for 30 minutes followed by a 30 seconds heat-shock in a water-bath at 42°C. Bacteria were allowed to recover in S.O.C. medium (Invitrogen, cat. 15544-034) using a thermal mixer (Eppendorf, cat. 12709018) set at 37°C / 1000rpm for 1 hour before transferring 30 to 80µL of bacterial suspension onto LB agar plates containing relevant antibiotic drugs selection (table

2.6). The suspension was spread thoroughly on the plate by the use of sterile glass beads (~Ø2mm). Plates were inverted and cultured in a 37°C microbiological incubator (Thermo Scientific, Heratherm IGS100) for 16 to 18 hours until macroscopic colonies were visible.

Table 2.6: Antibiotic concentrations for bacterial selection.

Antibiotic	Working concentration
Ampicillin	50 μg/ml
Kanamycin	30-50 μg/ml
Chloramphenicol	20 μg/ml
Tetracycline	4-10 μg/ml

2.5.7. Colony PCR

Bacterial colonies were initially screened by PCR as described in 2.5.1, except reactions were prepared in a total volume of 20μL, using Taq DNA polymerase (Qiagen, cat. 201203). The thermocycling conditions included an initial denaturation step of 10 minutes. Colonies were picked using rounded pipette tips, prepared by transiently passing through flame and allowed to cool. The colonies were first replica-plated on a fresh LB agar plate containing the same antibiotic drug selection and then added into the PCR reaction. Where possible, compatible primers binding to the vector sequence and the insert sequence were used to validate successful ligation with proper orientation. Otherwise primer pairs binding to the vector sequence flanking the ligation site were used and positive clones were identified by amplicon size.

2.5.8. Isolation of plasmid DNA from bacterial cultures

Positive clones identified by colony PCR (section 2.5.7) were picked and inoculated in 5mL of LB medium supplemented with relevant antibiotic drug selection using 14mL snap-cap tubes (Falcon, cat. 352059). Bacteria were cultured in an incubated shaker (Benchmark Scientific, Incu-Shaker Mini H1000-MR-T1550) set at 37°C / 200rpm for 16 hours. 25% glycerol/LB stocks were made using autoclaved glycerol and stored at –80°C. Bacterial cells propagating the pDNA of interest were collected by centrifugation of the cultured suspension at 5000×g for 10 minutes. Plasmid DNA were purified from the pellet using Miniprep or Maxiprep purification kits (Qiagen, cat. 27014 or 12162) according to the manufacturer's instructions. Quantity and purity (A260/A280 and A260/A230) of the purified pDNA were determined using a micro-volume spectrophotometer (Nanodrop, ND-1000). Diagnostic restriction digestion was also performed as described in section 2.5.3.

2.5.9. Transformation of *E.coli* by electroporation

All reagents and equipment mentioned in this section with the exception of the thermal mixer were pre-cooled in a 4°C cold room prior use. Overnight-cultured bacterial suspension was diluted 1:10 to 1:50 into 1mL of fresh LB medium containing relevant antibiotic drug selection in a microfuge tube and cultured using a thermal mixer at 37°C/1000rpm until its optical density at a wavelength of 600nm (OD_{600}) reached 0.35 to 0.4 absorption units (AU). Bacterial cells were collected by centrifugation at maximum speed for 1 minute. Cells were washed twice, by resuspending the pellet in ice cold UHP-H_2O and performing centrifugation again. The resultant pellet was resuspended in 20µL of UHP-H2O and 10ng of pDNA or 500ng of purified PCR amplicons were added to the bacterial suspension. The bacteria/DNA mix was transferred to a 100µm gap electroporation cuvette (BioRad, cat. 1652089). The cuvette was placed in a MicroPulser (BioRad, cat. 165-21000) apparatus and cells were electroporated using a preset program for *E.coli* at 1.8kV. 1mL of fresh ice cold LB

medium was added and bacterial cells were transferred to a microfuge tube for 1 to 2 hours incubation using a thermal mixer set at 30°C or 37°C / 1000rpm. A series of volumes from 50µL of a 1:100 dilution to 100µL of undiluted bacterial suspension were plated on individual LB agar plates with relevant antibiotics drug selection. Plates were incubated at 30°C for 48 hours or 37°C for 16 to 18 hours.

2.5.10. Genomic DNA isolation from cultured cells

ES cells were cultured in T25 flasks in LIF/FCS/GMEMβ until cells reached 80% confluence before they were collected by trypsin treatment as described in 2.1.2. DNeasy kit (Qiagen, cat. 69504) was used to extract gDNA from cell pellets according to the manufacturer's instructions. Quantity and purity (A_{260}/A_{280} and $A_{260}/_{A230}$) of the purified gDNA were determined using a micro-volume spectrophotometer (Nanodrop, ND-1000).

2.5.11. Southern blot analysis

Restriction digestion reactions were prepared containing 5µg of purified gDNA (section 2.5.10) in a total volume of 40µL and incubated overnight according to the manufacturer's instructions. Digested DNA was separated by electrophoresis (section 2.5.1) that was performed overnight at 30V, using 0.8% w/v ultra-pure agarose gels in 1× Tris/acetate/EDTA (TAE). Gels were incubated for 30 minutes in 1µg/mL ethidium bromide/H_2O solution (Sigma, E1510) using an orbital shaker. After nicking DNA fragments by exposure to 254nm UV for 1 minute, the gel was incubated twice for 15 minutes in a 0.5M NaOH, 1M NaCl solution and washed in a 0.5M Tris, 3M NaCl pH 7.4 solution. DNA was wet transferred overnight to a Hybond XL membrane (General Healthcare, cat. RPN303-S) soaked in 2× SSC solution by capillarity driven diffusion from the underlying 20× SSC solution, which comprised 3M NaCl, 0.3M tri-sodium citrate. Transfer was performed for 24 to 36 hours and the membrane was washed in 2× SSC before dried using an oven set at 80°C. The dried membrane was wetted with

2×SSC before it was blocked with PerfectHyb solution (Sigma, cat. H7033) containing 100μg/mL herring sperm DNA (Sigma, cat. D7656) in a rolling glass tube (Techne, cat. FHB-12) using a roller oven (Techne, cat. HB-1D) set at 68°C. The probe was prepared by PCR amplification of a specific non-repetitive locus near the region of interest using high-fidelity DNA polymerases as described in 2.5.1. The probe was purified using Qiaquick PCR purification kit (Qiagen, cat. 28104) and the quantity was determined using a micro-volume spectrophotometer. 25ng of probe DNA in Tris-EDTA (TE) buffer (10mM Tris-HCl pH 8.0, 1mM EDTA) were labelled with P32 α-dCTP using a DecaLabel DNA Labeling Kit (Thermo Scientific, cat. KO622) according to manufacturer's instructions. Excess unlabelled α-dCTP were removed using micro Bio-Spin P-30 gel columns (Bio-Rad, cat. 732-6223). The membrane was hybridized with the labelled probe overnight at 68°C, then rinsed and washed twice with 0.5× SSC, 0.1% w/v SDS for 90 and 30 minutes at 68°C. The membrane was exposed to Hyperfilm ECL (Amersham, cat. 28906837) in a dark room equipped with a safelight and the film was chilled at −80°C for 1 to 7 days depending on the signal intensity.

2.5.12. High Resolution Melting (HRM) assay

Each reaction of High Resolution Melting (HRM) assay was setup by first mixing the following components in a total volume of 15μL:

- 1x LightCycler 480 High Resolution Melting Master (Roche cat. 04909631001),
- 3.5mM $MgCl_2$ (Roche cat. 04909631001),
- High Performance Liquid Chromatography (HPLC)-purified primers at 0.2 μM each primer (Integrated DNA Technologies),
- PCR-grade H_2O.

100ng of gDNA, extracted from mouse ES (section 2.5.10) was diluted in PCR-grade H_2O in a final volume of 5 μL and added to each reaction. A non-template control reaction was also prepared by replacing the gDNA with H_2O. Each reaction was

prepared in triplicate. The assay was performed in a LightCycler 480 II (Roche, cat. 05015243001) instrument, using the program Gene Scanning 384 II and modifying the annealing temperature and the elongation time according to the manufacturer's instructions. The results were analyzed using the Light cycler 480 Melting Curve Genotyping Software.

2.5.13. Designing and functional testing of CRISPR/Cas9n sgRNAs

The RNA guided Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR-associated protein-9 (Cas9) nuclease system was used to introduce double strand breaks into the mouse ES cells genome and facilitate in this way gene targeting (Cong *et al.*, 2013). To achieve high specificity, a double nicking strategy was followed that combines the D10 mutant nickase version of Cas9 (Cas9n) with a pair of offset single guide RNAs (sgRNAs) complementary to opposite strands of the target site (Ran *et al.*, 2013).

sgRNAs were designed using a free online software, available at *http://crispr.mit.edu*, that scans the genomic sequence of interest for possible CRISPR guides [20 nucleotides followed by a Protospacer adjacent motif (PAM) sequence: NGG] and also predicts offtarget matches throughout the whole genome. The outcome of the analysis consists of a ranked list of all possible guides in the query sequence ordered according to a quantitative specificity analysis on the effects of base-pairing mismatch identity, position and distribution. For each target sequence, the two pairs of sgRNAs that presented the highest specificity were selected to be tested functionally.

sgRNAs were cloned into the plasmid pSpCas9n(BB)::2A::GFP (PX461) (gift from Feng Zhang, Addgene plasmid # 48140) (figure 2.2) following a protocol provided by the ZhangLab CRISPR Genome Engineering Resources (see appendix, figure S2.1).

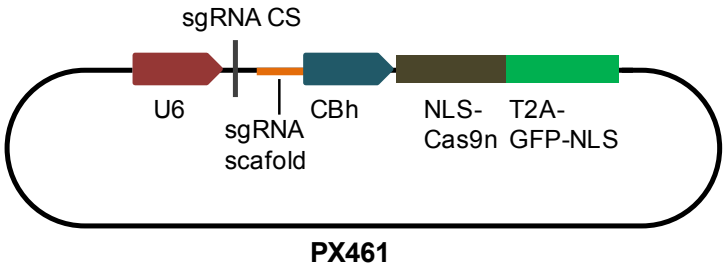


Figure 2.2: Cas9n-GFP plasmid used for genome targeting in mouse ES cells. Cas9n (D10A nickase mutant) from *S. pyogenes* with 2A-GFP, and cloning backbone for sgRNA.

In order to clone the target sequences into the pX461 backbone, each sgRNA was designed in the form:

```
5′ – CACCGNNNNNNNNNNNNNNNNNNN – 3′
3′ –      CNNNNNNNNNNNNNNNNNNNNCAAA – 5′
```

Each pair of oligos (synthesized by IDT) was phosphorylated and annealed. The vector PX461 was digested with the restriction enzyme BbsI (NEB, cat. R0539S), followed by dephosphorylation, using Alkaline Phosphatase, Calf Intestinal (CIP) (NEB, cat. M0290S) and subsequently gel purification (section 2.5.2). A ligation reaction was set up containing the digested vector and each sgRNA, followed by treatment with a Plasmid-Safe™ ATP-Dependent DNase (Epicentre, Cat. E3101K) to prevent unwanted recombination products. The ligation product was transfected into *E.coli* as described in section 2.5.6. Insertion of sgRNAs into the vector was tested by colony PCR (section 2.5.7) and plasmid DNA was extracted from the positive colonies using a Maxiprep kit (section 2.5.8). Orientation of the inserts and potential insertion of concatamers was further assessed by sequencing analysis (performed by DNA sequencing & services, University of Dundee).

To test the sgRNAs functionally, mouse ES cells were lipofected with each pair of sgRNAs and allowed to expand for 24 hours. GFP positive cells were then FACS sorted (section 2.1.7), replated in LIF/FCS/GMEMβ medium and allowed to grow for 24-36 hours. Cells lipofected with a mix in which DNA was substituted by H_2O were also plated as control. gDNA was extracted from each culture (section 2.5.10) and the sequence of the target site of each pair of sgRNAs were analyzed by HRM assay (section 2.5.12).

2.5.14. Karyotype determination of cultured cells

1 to 2 × 10^6 ES cells were plated in gelatin-coated T25 flask and cultured for 24 hours as describe in section 2.1.2 until cells reached 50 to 70% confluence. 300μL of KaryoMAX Colcemid solution (Gibco, cat. 15212-012) was added to LIF/FCS/GMEMβ medium and cells were cultured for a further 3 hours before harvested by trypsin treatment. Cell pellets were washed once in PBS and resuspended in 5mL of hypotonic solution (17mM sodium citrate, 33.6mM KCl in sterile UHP-H_2O) for 15 minutes at RT°C. Cells were collected by centrifugation at 290×g for 3 minutes and 2.5mL of hypotonic solution was substituted with 2.5mL of fixative solution (25% v/v acetic acid in methanol) which was added slowly, under gentle mix. The hypotonic/fixative mix was replaced with 5mL of fixative after cells were collected by centrifugation. The cells were then washed twice in 5mL fixative solution before concentrated in 0.5mL fixative solution. Cells were dropped and spread onto a glass slide before dried at RT°C and subsequently stained with 1μg/mL DAPI (Molecular Probes, cat. D1306) solution for 10 minutes. Slides were then washed with H_2O, dried and fixed with a cover-slip for imaging with an inverted fluorescence stereomicroscope (Olympus, IX-51) attached to a monochrome camera (QImaging, Retiga 2000R).

2.6. Gene expression quantification

2.6.1. Total RNA isolation and cDNA preparation

Plated cells or cell pellets were lysed by using RLT buffer (Qiagen, cat. 74104) supplemented with 1% β-mercaptoethanol (RLT/β) following the manufacturer's instructions. The cell lysate was homogenized by vigorous pipetting and total RNA was purified using RNeasy mini kit (Qiagen, cat. 74104) in NF-H_2O according to the manufacturer's instructions. Quantity and purity (A_{260}/A_{280} and A_{260}/A_{230}) of the purified total RNA were determined using a micro-volume spectrophotometer (Nanodrop, ND-1000).

Each reverse transcription (RT) reaction was setup by first mixing the following components in a total volume of 13µL:

- Random hexamers (Invitrogen cat. 8080127) -1µL of 50ng/µL stock
- dNTP (Invitrogen, cat. 10297-018) - 1µL of 10mM stock
- RNA - 0.3 to 1µg RNA in 11µL H_2O

RNA/primer mix was then denatured at 65°C for 5 minutes using a thermal cycler before chilled on ice for 2 minutes. Each RT reaction was subsequently supplemented with 7µL of a mix containing:

- First-Strand buffer (SuperScript III first-strand synthesis kit, Invitrogen, cat. 18080-093) - 4µL of 5x stock
- Dithiothreitol (DTT) (SuperScript III first-strand synthesis kit) - 1µL of 0.1M stock
- RNaseOUT inhibitor (Invitrogen, cat. 10777-019) - 1µL of 40U/µL stock
- SuperScript III reverse transcriptase (in SuperScript III first-strand synthesis kit) - 1µL of 200U/µL

Reverse transcriptase minus (RT-) controls were prepared for each RNA sample, following exactly the same protocol, except for adding H_2O instead of reverse transcriptase. The reactions were incubated at RT for 10 minutes. Using a thermal cycler, first-strand synthesis was performed, at 50°C for 1 hour, followed by inactivation at 72°C for 15 minutes. The resultant cDNA mix was diluted 1:10 with NF-H_2O.

2.6.2. mRNA level quantification by real-time PCR

Quantitative real-time PCR (QRT-PCR) was performed using a LightCycler 480 II (Roche, cat. 05015243001) instrument and the LightCycler 480 Software, Version 1.5.0.39. Reactions were prepared in 384 wells-plates (Roche, cat. 04729749001). Each reaction was setup containing 3µL of cDNA (section 2.6.1) and 7µL of mixture composed of LightCycler 480 SYBR Green I Master (Roche cat. 04887352001) or Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent, Cat. 600882) and primers at a final concentration of 0.5µM. Each reaction was performed in triplicate. A reaction in which cDNA was replaced by H_2O was prepared for each pair of primers as a negative control. Analysis of RT minus cDNA samples (section 2.6.1) was performed alongside the samples of interest. Standard curves of each pair of primers were prepared separately and used for the estimation of the relative mRNA concentrations of each sample. Values derived for each gene were normalized to that of TATA-box binding protein (TBP).

Chapter 3:
Identification of distinct subpopulations in EpiSC cultures

3.1. Introduction

EpiSC cultures can be considered an *in vitro* equivalent of the mouse post-implantation epiblast. Although they can be derived from mouse embryos of a wide range of developmental stages, EpiSC populations do not retain the transcriptional profile of the developmental stage of the source tissue, but they converge to a cellular stage similar to the epiblast of the late-gastrula-stage embryo (Kojima *et al.*, 2014). Under self-renewing conditions, EpiSCs express both the core pluripotency factors *Oct4*, *Sox2* and *Nanog*, and also early lineage specification genes, including *Tbra* (Brons *et al.*, 2007; Tesar *et al.*, 2007; Han *et al.*, 2010). The expression of genes associated with pluripotency or differentiation is heterogeneous and seems to mark distinct subpopulations that are in a state of dynamic equilibrium (Hayashi and Surani, 2009; Han *et al.*, 2010). Interestingly, it has been shown that EpiSC cultures, under conditions that promote pluripotency, include subsets of cells that have progressed into different stages of lineage specification (Hayashi and Surani, 2009).

Tbra is expressed in several cell types in the mouse embryo. It marks epiblast that is about to ingress through the primitive streak, as well as ingressing and nascent mesoderm. It is also expressed in the node and notochord, as well as in NMPs. The latter cell type also coexpresses *Sox2* (Tsakiridis *et al.*, 2014; Wymeersch *et al.*, 2016). Since EpiSCs at the population level have been reported to express both *Tbra* and *Sox2*, the aims of the work presented in this study were to:

- investigate whether Tbra^{+} EpiSCs represent a distinct cell state from their Tbra^{-} counterparts,

- determine whether Tbra$^+$ EpiSCs correspond to any known embryonic cell types, in particular NMPs.

3.2. EpiSC cultures contain a major subpopulation of cells with early primitive streak-like character

To investigate whether EpiSC cultures contain NMPs, the correlation between expression of *Sox2* and *Tbra* was analysed at the single cell level. Since loss of pluripotency in the epiblast coincides with extinction of *Nanog* expression (Osorno *et al.*, 2012), Nanog was included into the analysis as a marker that would allow the distinction between cells that are in the pluripotent state and cells that may have progressed into lineage commitment. Triple immunocytochemistry for Tbra, Sox2 and Nanog proteins in EpiSCs confirmed the heterogeneous character of EpiSC cultures and demonstrated that around 64% of the culture consists of Sox2/Nanog double positive cells, confirming that cells coexpressing pluripotency markers constitute the majority of the EpiSC population. Around 45% of the culture consists of Tbra/Sox2 double positive cells (figure 3.1). Crucially, the vast majority of this fraction (84%) was found to be Nanog positive, implying that the Tbra/Sox2 double positive cells have not exited pluripotency. Although the majority of Tbra$^+$ cells expressed either or both pluripotency markers (figure 3.1), negative correlation between expression levels of Tbra and Sox2/Nanog was observed, with increasing levels of Tbra corresponding to lower levels of each of the pluripotency factors. In contrast to that, cells with higher Sox2 levels were shown to express higher levels of Nanog. In conclusion, the expression of both *Sox2* and *Nanog* in the Tbra$^+$ cells suggests that EpiSCs do not express *Tbra* as a result of spontaneous differentiation, but the vast majority of Tbra$^+$ cells seems to be at the primed pluripotent state. The small fraction of Tbra$^+$/Nanog$^-$ /Sox2$^-$ cells may represent cells that have progressed into lineage commitment, while the detection of Tbra$^+$/Sox2$^+$/Nanog$^-$ cells may indicate the presence of a small subset of NMP-like cells.

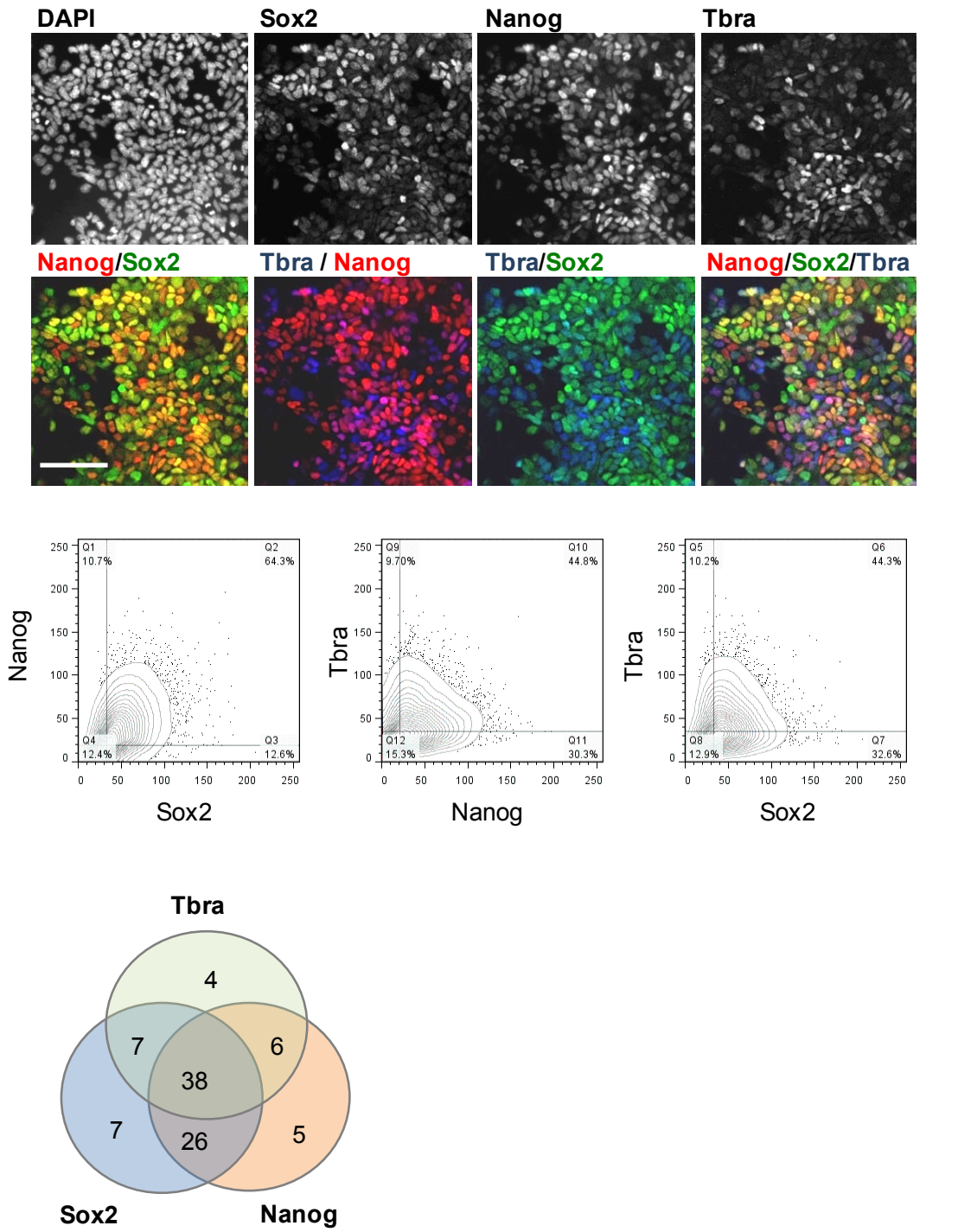


Figure 3.1: Correlation of Tbra, Sox2 and Nanog expression in EpiSC cultures. Tbra, Sox2 and Nanog immunocytochemistry in undifferentiated, wild type EpiSCs. Graphs: Immunofluorescence quantification following single cell image analysis. Venn diagram: Percentages of cells expressing any combination of the three markers. Scale bar: 100µm.

The next step was to dissect further the properties of the Tbra^{+} EpiSCs. For that purpose, a previously published Tbra-GFP mouse ES cell reporter line (Fehling *et al.*, 2003) was used and an EpiSC line was derived from it *in vitro*. FACS analysis of undifferentiated EpiSC cultures demonstrated that 30-60% of the cells were GFP^{+} (figure 3.2, A). Double immunocytochemistry for GFP and endogenous Tbra confirmed very good correlation between the expression levels of the two proteins, although GFP was not detected in all Tbra^{+} cells, probably due to different degradation rates of the two proteins (figure 3.2, B).

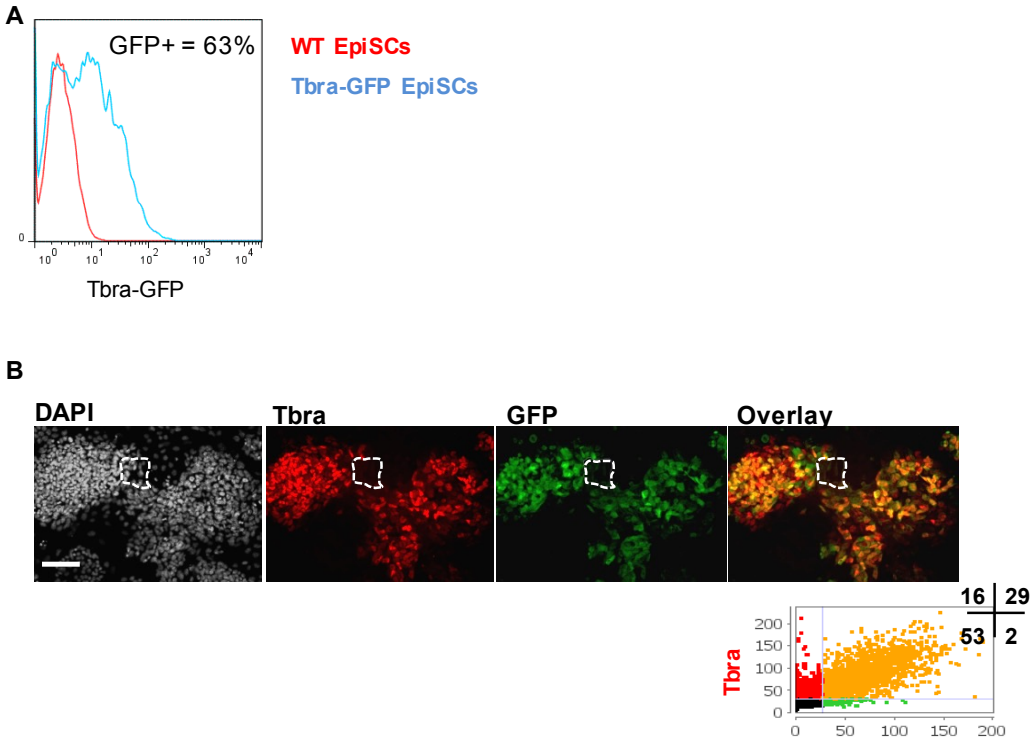


Figure 3.2: Expression of Tbra-GFP in EpiSC cultures. Tbra-GFP expression in undifferentiated EpiSC cultures assessed by flow cytometry. **B)** Tbra and GFP immunocytochemistry in EpiSC cultures. Graph: Immunofluorescence quantification following single cell image analysis. The numbers show the percentages in each quadrant. The areas marked on the immunofluorescence images were used to determine the background fluorescence signal in image quantification. WT, wild type. Scale bar: 100µm.

To analyse the expression profile of the Tbra-GFP$^+$ cells, EpiSC cultures were FACS sorted into Tbra-GFP positive and negative fractions and QRT-PCR was performed for the analysis of pluripotency and lineage specification markers. Transcripts of all of the genes were detected in both EpiSC fractions, but interestingly, distinct expression patterns were observed (figure 3.3). More specifically, expression of the core pluripotency factors *Oct4*, *Sox2* and *Nanog,* as well as expression of the neural related genes *Otx2* (Simeone *et al.*, 1992; Acampora *et al.*, 1995), the *Achaete-scute* homolog *Mash1* (Guillemot and Joyner, 1993) and *Sox3* (Wood and Episkopou, 1999) were not significantly enriched in either population, but a tendency for these markers to favour the Tbra-GFP$^-$ cells was visible. The Tbra-GFP$^+$ subpopulation was predominantly marked by higher levels of the early primitive streak and organiser markers *Foxa2* (Sasaki and Hogan, 1993), *Noto* (Abdelkhalek *et al.*, 2004), *Fgf8* (Crossley and Martin, 1995), *Mix paired-like homeobox* (*Mixl1)* (Robb *et al.*, 2000), *Mesoderm posterior bHLH transcription factor 1* (*Mesp1*) (Saga *et al.*, 1996) and *LIM homeobox 1* (*Lhx1*) (Shawlot and Behringer, 1995). Expression of the late primitive streak/ NMP markers *Wnt3a* (Takada *et al.*, 1994), *Nkx1.2* (Schubert *et al.*, 1995), *Cdx2* (Beck *et al.*, 1995; van den Akker *et al.*, 2002) and *Evx1* (Bastian and Gruss, 1990; Cambray and Wilson, 2007) were also enriched in the Tbra-GFP$^+$ cells, but to a lesser extent, compared to those activated earlier on.

The expression analysis of some of the genes was extended via immunocytochemistry (figure 3.4). Double immunocytochemistry for Tbra-GFP and each of the pluripotency factors Sox2 and Nanog reproduced the expression pattern that was observed in the wild type EpiSC cultures; specifically, it was observed that the vast majority of Tbra-GFP$^+$ EpiSCs expresses these pluripotency factors, with increasing Tbra-GFP levels corresponding to lower levels of Sox2 or Nanog. Analysis of the expression pattern of the endoderm/axial mesoderm marker *Foxa2* indicated a degree of co-localisation with Tbra-GFP, in contrast to the expression of the neural and ingressed mesoderm marker *N-cadherin* (*Cdh2*) (Radice *et al.*, 1997) that was mainly observed in the Tbra-GFP negative cells. Taken together these data demonstrate that under self-renewing

conditions, expression of *Tbra* marks an EpiSC fraction that is enriched in early primitive streak-like and possibly axial mesoderm cells.

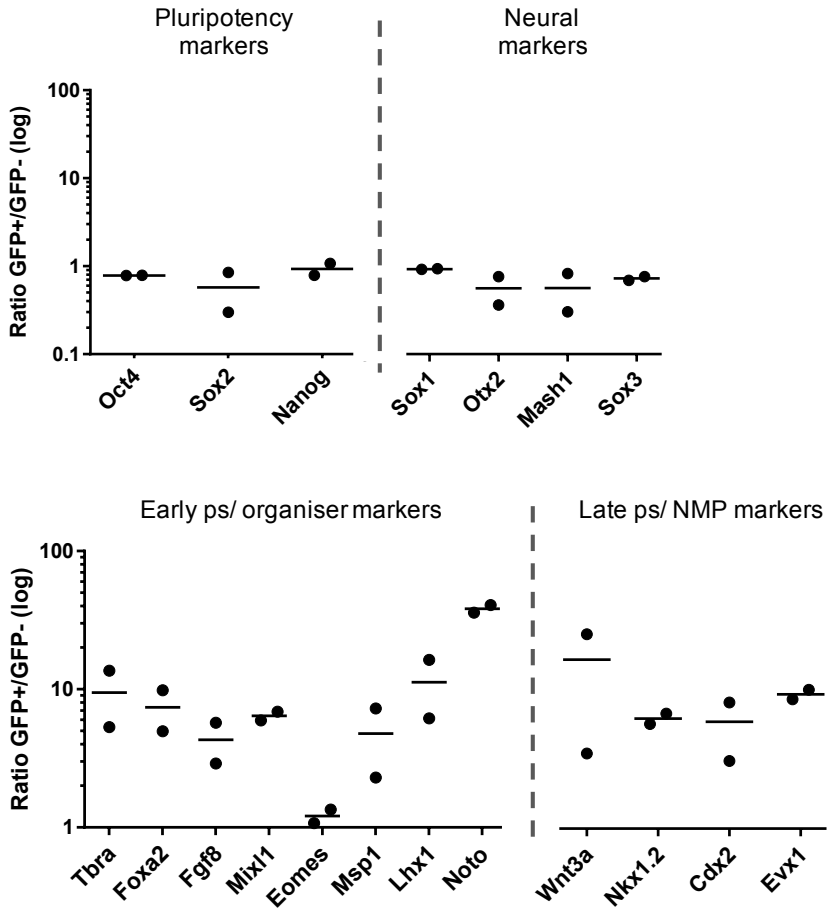


Figure 3.3: Expression analysis of Tbra-GFP positive and negative EpiSCs. QRT-PCR analysis of indicated markers in sorted Tbra-GFP positive and negative EpiSC fractions. Results are represented as $\log_{10}$ ratio of expression versus GFP negative cells. N=2. Ps, primitive streak.

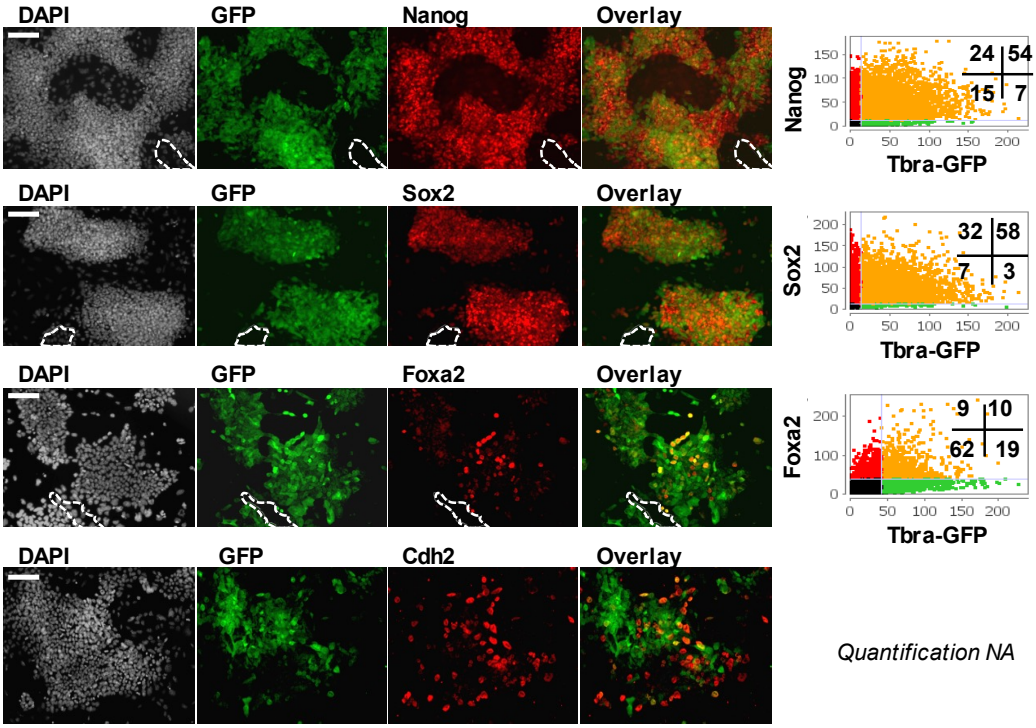


Figure 3.4: Expression analysis in Tbra-GFP EpiSCs. Immunocytochemistry in EpiSC cultures. Graph: Immunofluorescence quantification following single cell image analysis. The numbers show the percentages of cells in each quadrant. The areas marked on the immunofluorescence images were used in image quantification to determine the background fluorescence signal. Image analysis of Chd2 immunofluorescence was not possible due to the localisation of the Cdh2 protein on the cell membrane. Scale bar: 100µm.

To assess whether Tbra-GFP$^+$ and Tbra-GFP$^-$ cells can interconvert, cells of sorted fractions were replated in EpiSC cultures conditions in three different cell densities. Colonies with the characteristic EpiSC morphology arose from both GFP$^+$ and GFP$^-$ cells, while daily FACS analysis demonstrated that both sorted subpopulations re-equilibrated the original GFP levels within three days (figure 3.5), suggesting that the the primitive streak-like Tbra-GFP$^+$ EpiSC subpopulation exists in dynamic equilibrium with GFP$^-$ cells. Interestingly, GFP$^-$ cultures regained 40% GFP$^+$ cells within 24 hours, whereas the GFP$^+$ replated fractions did not exhibit such a fast interconversion rate, probably due to a slower degradation rate of the GFP protein. The initial plating density did not seem to have an effect on that. Although the GFP$^-$

cultures exhibited a rapid increase of GFP^+ cells within 24 hours, the GFP levels increased at a much lower rate thereafter. 72h post plating, the levels of GFP^+ cells reached 60% and 40% in the cultures that arose from GFP^- cells plated at density 10,000 and 100,000 cells/well, respectively (figure 3.5). Over the course of many experiments I observed that higher density cultures contained lower percentages of GFP^+ cells (results not shown), which is consistent with these findings. Altogether, these data suggest increased density has a negative effect on the levels of Tbra-GFP.

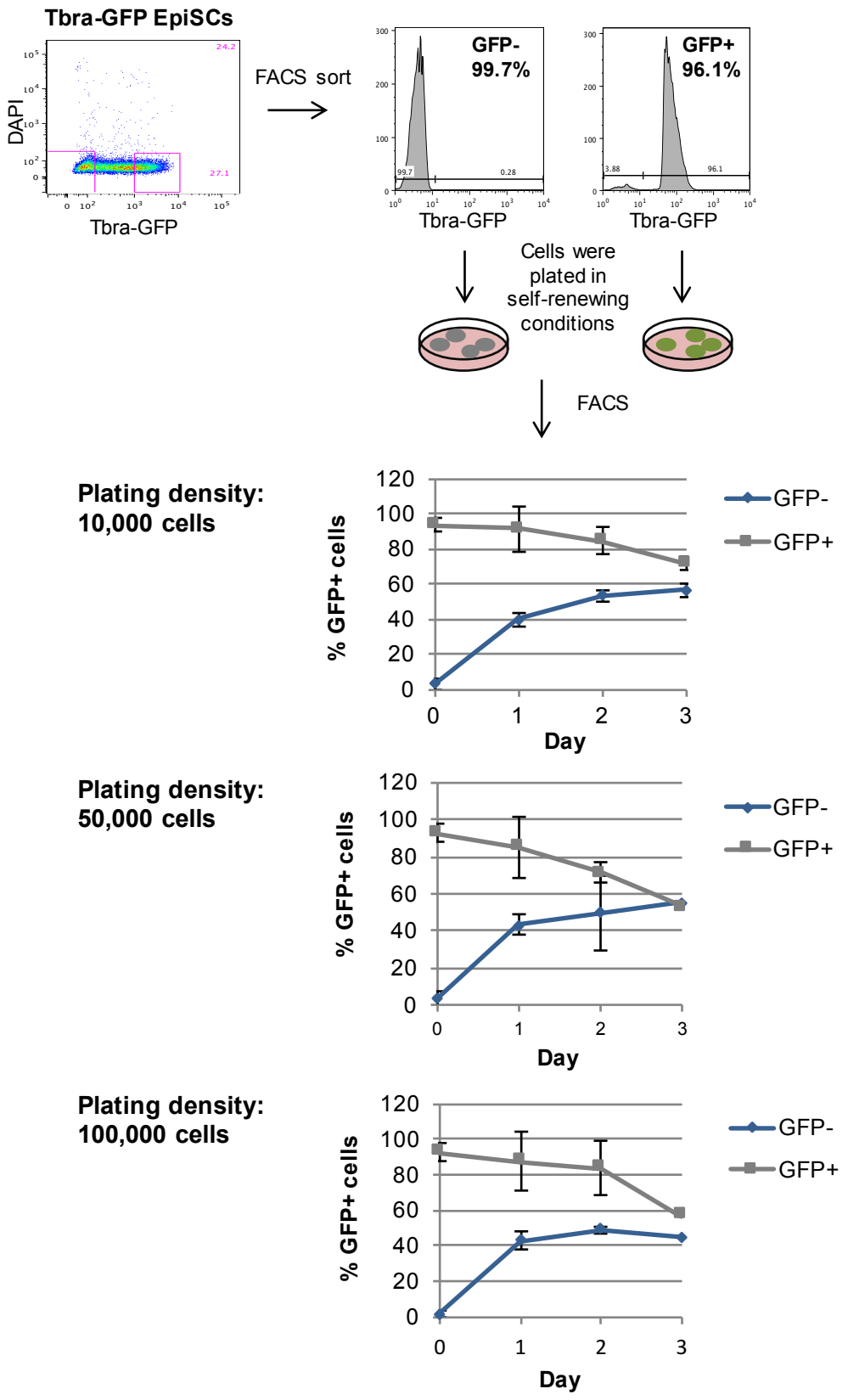


Figure 3.5: Interconversion of sorted Tbra-GFP positive and negative EpiSCs.

Figure 3.5: Interconversion of sorted Tbra-GFP positive and negative EpiSCs. Tbra-GFP EpiSC cultures were FACS sorted into GFP^+ and GFP^- subpopulations. The graphs on the right show the purity of the sorted fractions. Cells were replated in self-renewing conditions at the indicated cell densities. The graphs present the (%) percentage of GFP^+ cells in the cultures that emerged from the originally sorted GFP positive and negative fractions. N=3. Error bars represent standard deviation.

I next sought to investigate whether the differential expression of neural-related and primitive streak-related genes in the GFP^+ and GFP^- fractions reflects variations in the differentiation potential of the two subpopulations. In order to test the capacity of GFP^+ and GFP^- cells to progress into neural specification, sorted EpiSCs were plated on matrigel pre-coated dishes (~10^4 cells/cm^2), in N2B27 medium without Fgf2 and Activin, for 2 days (modified from Iwafuchi-Doi *et al.*, 2012) and the induction of neural markers *Sox1*, *Sox3* (Wood and Episkopou, 1999), *Mash1* (Casarosa *et al.*, 1999), *Paired box 6* (*Pax6*) (Walther and Gruss, 1991) and *Neuronal differentiation 1* (*NeuroD*) (Lee *et al.*, 1995) was assessed by QRT-PCR. Both GFP^+ and GFP^- cells showed upregulation of the neural markers, but interestingly, higher levels of the neural related transcripts were detected in the cultures that emerged from GFP^+ cells (figure 3.6).

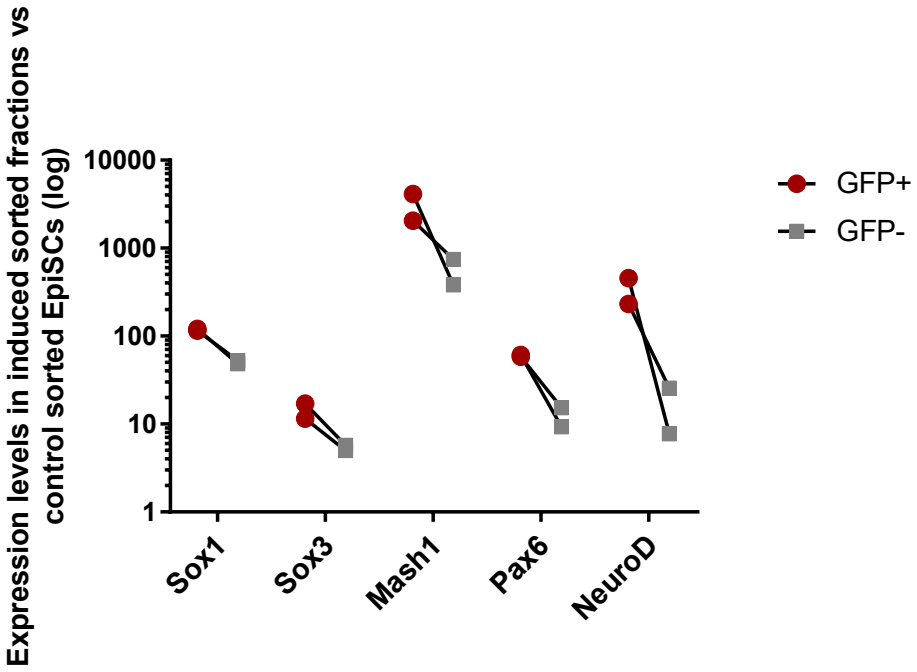


Figure 3.6: Neural differentiation ability of Tbra-GFP EpiSC fractions. RTQ-PCR analysis of neural markers in sorted Tbra-GFP positive and negative fractions, that were cultured in neural inducing conditions. Results are represented as log_{10} ratio of expression versus EpiSC control fractions. N=2 (two replicates with cells from one FACS sorting).

Given the expression profile of the Tbra-GFP$^+$ EpiSC fraction, the observation that under neural inducing conditions, the Tbra-GFP$^+$ cells exhibited higher levels of neural related transcripts was rather surprising. Hence, the next step was to investigate these data further, by analysing the expression pattern of factors that are involved in neural induction during gastrulation. *In vivo*, the BMP antagonists *Chordin* and *Noggin* are expressed in the node and they are necessary for development of the anterior nervous system (McMahon *et al.*, 1998; Bachiller *et al.*, 2000). The Wnt inhibitor Dkk1 (Mukhopadhyay *et al.*, 2001; del Barco Barrantes *et al.*, 2003) and the Nodal/Wnt/BMP antagonist Cer1 (Shawlot *et al.*, 1998; Piccolo *et al.*, 1999) are found in the anterior mesendoderm, that is known to maintain the overlying neuroectoderm (reviewed in Arnold and Robertson, 2009). Homozygous *Dkk1* or *Cer1* mutant mice develop truncations of the anterior central nervous system. Based on the fact that *Tbra* is expressed in both the node and the nascent mesoderm, I investigated whether there is any correlation between expression of Tbra and expression of those inhibitors in EpiSC cultures. QRT-PCR analysis of sorted Tbra-GFP positive and negative fractions demonstrated that the Tbra-GFP$^+$ cells contained higher levels of *Dkk1*, *Cer1*, *Noggin* and *Chordin* (figure 3.7, A), providing a possible explanation for the ability of Tbra-GFP$^+$ cells to reach greater induction levels under neural differentiation conditions.

Altogether, these data suggest that Tbra-GFP$^+$ EpiSCs are not spontaneously differentiated cells, but a dynamic, early primitive-streak like subpopulation that expresses the pluripotency factors and retains the ability to undergo neural differentiation.

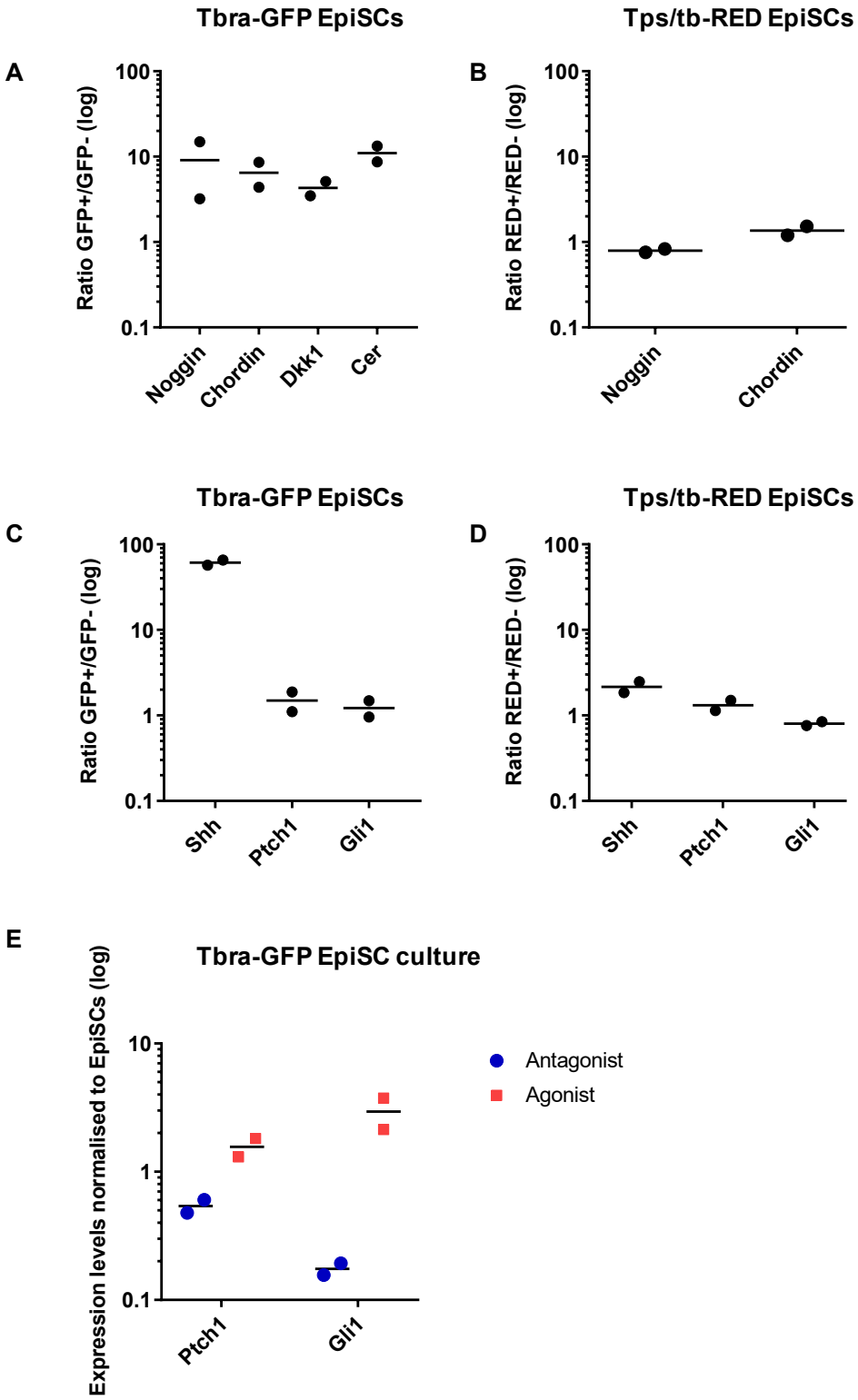


Figure 3.7: Expression study of signalling molecules in EpiSC cultures.

Figure 3.7: Expression study of signalling molecules in EpiSC cultures. A, C) QRT-PCR analysis of indicated markers in sorted Tbra-GFP positive and negative EpiSCs or **B, D**) sorted Tps/tb-RED positive and negative EpiSCs, under self-renewing conditions. Results are represented as $\log_{10}$ ratio of expression versus GFP^- or $dsRED^-$ cells. **E)** QRT-PCR analysis in whole EpiSC cultures (plating density: $4x10^3$ cells/cm^2) treated with 0.4μM purmophamine (Shh agonist) or 6μM Shh antagonist for three days. Results are represented as $\log_{10}$ ratio of expression versus untreated control EpiSCs. N=2.

3.3. EpiSC cultures contain a subset of Tbra-GFP$^+$ cells with node/notochord like characteristics

As mentioned before, *in vivo*, expression of *Noto*, *Noggin* and *Chordin* is restricted to the node and the notochord, where *Tbra* and *Foxa2* are also co-expressed. Based on the observation that EpiSC cultures contain Tbra/Foxa2 double positive cells and the Tbra-GFP$^+$ fraction is enriched in *Noto*, *Noggin* and *Chordin* transcripts, we hypothesized that Tbra-GFP$^+$ EpiSCs contain a subpopulation with node/notochord characteristics. To test this idea further, a second EpiSC reporter line, the Tps/tb-RED, was used, that contains a *dsRed2* (RED) transgene under the transcriptional control of a randomly integrated *Tbra* promoter fragment that is active specifically in the primitive streak and the tail bud, but not the notochord (Clements *et al.*, 1996; Tsakiridis *et al.*, 2014; Economou *et al.*, 2015). Since the Tps/tb-RED reporter is designed so as not to mark node/notochord cells, it was predicted that if a node/notochord like subpopulation exists in the EpiSC cultures, then it should not be marked by RED. QRT-PCR analysis of sorted RED$^+$ and RED$^-$ EpiSC fractions showed no differences in the expression levels of *Noggin* and *Chordin* in the two subpopulations (figure 3.7, B), an outcome that is consistent with our hypothesis. I next sought to analyse the expression of the morphogen *Shh* that is also expressed in the node/notochord (Echelard *et al.*, 1993), as well as its receptor *Patched* (*Ptch1*) and the downstream transcriptional activator *GLI-Kruppel family member GLI1* (*Gli1*) (Cohen *et al.*, 2015). Transcripts of the three genes were detected in both positive and negative fractions of the two reporter lines, but importantly, Tbra-GFP$^+$ EpiSCs were shown to be enriched in *Shh*; Tps/tb-RED$^+$ cells also presented higher expression levels of *Shh* than the Tps/tb-RED$^-$ cells, although the difference between the two RED fractions (2-fold increase in RED$^+$) was much smaller than the difference between the two GFP subpopulations (61-fold increase in GFP$^+$) (figure 3.7, C and D). Altogether, these data support the notion that Tbra-GFP$^+$ EpiSCs may contain a subset of cells with node/notochord characteristics.

The detection of the *Shh*, *Gli1* and *Ptch1* transcripts in EpiSCs prompted me to investigate whether Shh signalling occurs in EpiSC cultures. For that purpose, the Shh agonist purmorphamine (Calbiochem, cat: 540223) and a Shh antagonist (Genentech, see Souilhol *et al.*, 2016) was added in EpiSC cultures for three days and subsequently, the expression levels of *Ptch1* and *Gli1* were assessed by QRT-PCR (Cohen *et al.*, 2015) (figure 3.7, E). In the presence of the Shh antagonist, expression of *Gli1* and *Ptch1* was upregulated, whereas treatment with the Shh agonist induced higher expression levels of the two genes. Although titration of the concentration of the two compounds is probably necessary to achieve greater induction or inhibition of Shh signalling, these preliminary data suggest that the Shh signalling pathway is active in EpiSC cultures. Taking into account the substantially higher expression of Shh in the Tbra-GFP$^+$ fraction, it is possible that the Tbra-GFP$^+$ EpiSCs, and more specifically, the node/notochord-like cells are the source of the Shh signalling.

3.4. Discussion

EpiSCs exist in the so-called primed pluripotent state (reviewed in Nichols and Smith, 2009), and similar to their tissue of origin, the post-implantation epiblast, they express pluripotency factors and early lineage specification genes. Like mouse and human ES cells (Canham *et al.*, 2010; Blauwkamp *et al.*, 2012), EpiSC cultures appear to be very heterogeneous and contain subsets of cells with different characteristics. In this chapter, I analyse the properties of the Tbra positive EpiSCs. However, it should be noted that although Tbra has been used as a prominent example of an early differentiation marker detected in EpiSC cultures, it has been reported that not all EpiSC lines express *Tbra* (Bernemann *et al.*, 2011). Importantly, comparison of different cell lines revealed that the *Tbra* RNA levels correlate negatively with the capacity of cells to undergo neural differentiation and their ability to revert to an ES cell-like state (Bernemann *et al.*, 2011). In the work presented here, I examine the heterogeneous expression of *Tbra* within individual EpiSC cultures and I identify subpopulations that seem to represent the *in vitro* counterparts of distinct embryo tissues (summarized in figure 3.8).

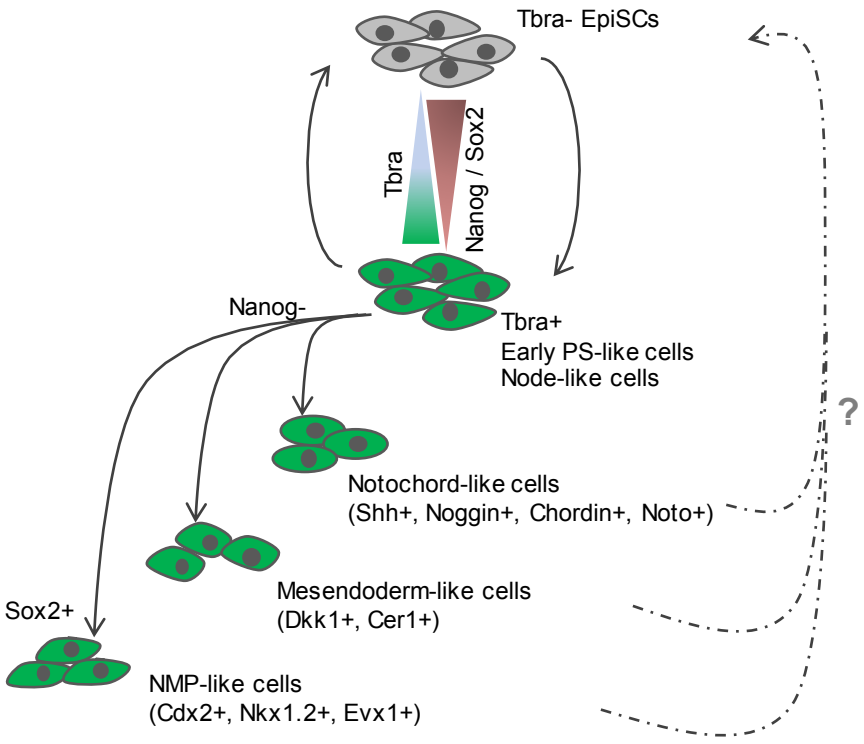


Figure 3.8: Model illustrating the fluctuation of EpiSCs between different cell states. In self-renewing EpiSC cultures, the protein levels of Tbra present negative correlation with the levels of Nanog and Sox2 protein. The Tbra^{+} subpopulation, that contain early primitive streak-like and node-like cells, can revert back to a Tbra^{-} state. Based on the absence of Nanog in a subset of cells, it has been speculated that these cells resemble the post-ingression epiblast cells that have progressed into lineage specification. Identification of the different subpopulations is based on the enrichment of Tbra-GFP^{+} cells in transcripts of genes that mark distinct embryonic tissues *in vivo*; these markers are shown in brackets. Whether the different subsets of post-ingression like cells can revert back to a Nanog ^{high} / Tbra ^{low} state is not known.

The majority of Tbra$^+$ EpiSCs resemble the pre-ingression epiblast cells and exhibit early primitive streak-like characteristics

In vivo, during gastrulation, expression of *Sox2* is predominantly in the anterior epiblast (Wood and Episkopou, 1999; Uchikawa *et al.*, 2011; Cajal *et al.*, 2012; Iwafuchi-Doi *et al.*, 2012), whereas *Tbra* is found in the primitive streak and the node (Kispert and Herrmann, 1994). Although *Tbra* and *Sox2* seem to have a rather mutually exclusive expression pattern, recent studies have shown that the node-streak border of E7.5 embryos contains a small proportion of Tbra/Sox2 double positive cells (Tsakiridis *et al.*, 2014; Wymeersch *et al.*, 2016). *Nanog* is also regionally expressed, occupying the proximal-posterior part of the epiblast (Hart *et al.*, 2004; Hatano *et al.*, 2005), while cells lose expression of *Nanog* as they move through the primitive streak (Osorno *et al.*, 2012). *Oct4* is expressed throughout the embryo until the late bud stage and in more posterior regions thereafter (Scholer *et al.*, 1990; Yeom *et al.*, 1996). Crucially, loss of pluripotency occurs at the beginning of somitogenesis and is coincident with extinction of Nanog, but before Oct4 becomes undetectable (Osorno *et al.*, 2012). After somitogenesis begins, coexpression of *Tbra* and *Sox2* is detected at the NSB of E8.5 embryos where bipotent axial progenitors, NMPs, are found (Wymeersch *et al.*, 2016). In this work, triple immunocytochemistry of Tbra, Sox2 and Nanog in EpiSCs shows that the vast majority of Tbra$^+$ EpiSCs are also expressing Nanog, suggesting that these cells are pre-ingression epiblast-like cells that have not exited pluripotency. These data are in agreement with grafting experiments that we have previously performed in our lab and that have shown that EpiSCs can readily generate chimeras when grafted to postimplantation embryos at stages earlier than E7.5, but that they do no engraft well in E8.5 embryos (Huang *et al.*, 2012).

The use of a Tbra-GFP reporter cell line allowed me to investigate the properties of the Tbra-expressing EpiSCs further. Expression analysis at the population but also at the single cell level, demonstrated that the pluripotency factors Oct4, Sox2 and Nanog were detected in the vast majority of Tbra-GFP$^+$ EpiSCs. The Tbra-GFP$^+$ fraction is

shown to be enriched in early primitive streak and organiser markers. Late primitive streak/ NMP markers are also expressed at higher levels in the Tbra^{+} fraction, but not as predominantly as the genes expressed in the primitive streak earlier on. On the contrary, expression of *Tbra* seems to be negatively correlated with expression of neural related genes. These data agree with the general notion of EpiSCs representing the primed pluripotent state (reviewed in Nichols and Smith, 2009) and indicate that the heterogeneous expression of *Tbra* in EpiSC cultures does not mark spontaneously differentiated cells, but primitive streak like cells that have not exited pluripotency (as shown by expression of the core pluripotency factors), but seem to be poised for mesoderm/endoderm differentiation.

Accumulating evidence suggests that the heterogeneous expression of mesoderm/endoderm markers, including *Tbra*, is due to endogenous activity of the Wnt/β-catenin signalling pathway in EpiSCs (Sumi *et al.*, 2013; Tsakiridis *et al.*, 2014; Kurek *et al.*, 2015). Specifically, inducible deletion of both β-catenin alleles in EpiSCs led to ablation of *Tbra* expression (Tsakiridis *et al.*, 2014). In addition, it has been shown that inhibition of the Wnt pathway in EpiSCs under self-renewing conditions represses the expression of primitive streak markers, whereas the opposite effect has been reported upon Wnt signalling stimulation (Sumi *et al.*, 2013; Tsakiridis *et al.*, 2014; Kurek *et al.*, 2015). These findings are consistent with the *in vivo* role of Wnt signalling in the initiation of the primitive streak (Liu *et al.*, 1999) and the identification of *Tbra* as its direct target (Yamaguchi *et al.*, 1999).

In the work presented here, the Tbra-GFP positive and negative fractions exhibit the ability to interconvert. As shown in the purity tests of the sorted fractions (figure 3.5), the sorted GFP^{+} subpopulations were contaminated with a small percentage of GFP^{-} cells. Could the apparent interconversion of the two fractions be due to selective growth of the GFP negative cells over the GFP positive ones? Although the expansion rates of the two fractions have not been estimated in the work presented here, a former member of our lab has performed the same experiment using a different *Tbra* reporter EpiSC line (see below) and has shown that the two sorted populations expand at a

comparable rate (Tsakiridis *et al.*, 2014). Hence, it seems unlikely that the interconversion pattern the GFP^+ cultures exhibited is due to expansion of the contaminating GFP^- cells, but rather it represents the ability of GFP^+ cells to switch off expression of *Tbra* and/or give rise to GFP^- cells, suggesting that the GFP^+ cells are at an early reversible stage of their progression towards mesoderm/endoderm specification. Interestingly, 40% of GFP GFP^- cells became GFP^+ within 24 hours. Given that Nodal signalling plays a central role in the induction of most mesodermal and endodermal cell types *in vivo* (Schier, 2003), it is essential for initiation of *Tbra* expression in the primitive streak (Conlon *et al.*, 1994) and it has been shown to induce *Tbra* expression in ES cells under differentiation (Gadue *et al.*, 2006), it is possible that the immediate upregulation of *Tbra* in the GFP^- cells is induced by the presence of Activin in the culture medium. Interestingly, after 24h, the GFP levels did not increase as rapidly thereafter. Moreover, over the course of many experiments I noticed that higher density cultures contained lower percentages of GFP^+ cells. Hence, it is possible that paracrine/ autocrine factors produced by the Tbra-GFP^+ cells are responsible for the restriction of GFP under certain levels. Altogether, the balance that seems to exist between the GFP^+ and GFP^- fractions can be explained by the presence of two opposing forces: the culture medium probably induces expression of Tbra-GFP, whereas paracrine/ autocrine factors secreted by the GFP^+ cells inhibit it.

Although the expression profile of Tbra-GFP^+ cells would suggest differentiation bias towards mesoderm/endoderm rather than neuroectoderm, when sorted GFP^+ and GFP^- fractions were plated in conditions that are known to promote the development of neural plate cells (N2B27 medium without cytokines) (modified from Iwafuchi-Doi *et al.*, 2012), Tbra-GFP^+ cells exhibited greater induction of neural related genes compared to the GFP^- fraction. These findings can be explained by the enrichment of Tbra-GFP^+ cells in *Noggin*, *Chordin*, *Dkk1* and *Cer1* mRNA, that act as inhibitors of Wnt, BMP and Nodal signalling, playing a key role in neural induction during gastrulation (reviewed in Arnold and Robertson, 2009). This notion is further supported by the study of Matsuda and Kondoh (2014) demonstrating that when EpiSCs are cultured under neural inducing conditions (N2B27 medium without cytokines), overexpression of *Dkk1* leads to upregulation of the anterior forebrain

precursor markers *Homeobox gene expressed in ES cells* (*Hesx1*) and *Sine oculis-related homeobox 3* (*Six3*).

While this dissertation was under preparation, a new study was published which involves the use of the same Tbra-GFP reporter line and investigates the heterogeneity that EpiSC cultures exhibit under self-renewal conditions (Song *et al.*, 2016). This study, used the same culture conditions as we use in our lab, except for pre-coating dishes with FBS. The expression analysis performed by Song *et al.* (2016) agrees with the data presented n this chapter. In addition, it is shown that GFP^+ and GFP^- fractions can interconvert; similar to my observations, the authors notice that GFP^+ cells emerge in the GFP^- cultures within 24 hours, but the rate of this interconversion is not presented. When sorted GFP positive and negative fractions were plated in medium lacking Fgf2 and Activin, but supplemented with 10% FBS, the Tbra-GFP^+ cells presented bias towards mesoderm/endoderm differentiation, while the Tbra-GFP^- cells exhibited bias towards the neuroectoderm lineage. Hence, it was confirmed that Tbra-GFP^+ EpiSCs are not spontaneous differentiated cells, but they retain the ability to undergo neural differentiation. However, it appears that the culture conditions affect the way the GFP positive and negative fractions progress into neural specification.

Analysis of the Tbra-GFP^+ EpiSCs has also been performed by Kurek *et al.* (2015). Using the ability of sorted cells to form EpiSC colonies and embryoid bodies (EBs) as a measure of pluripotency, the authors state that a major fraction of the Tbra-GFP positive cells have lost pluripotency. However, a closer look at the data reveals that the capacity of cells to form colonies and EBs, as well as the expression levels of *Sox2* and *Oct4*, are not completely absent in GFP^+ cells; they exist, but they exhibit negative correlation with the levels of Tbra-GFP. This observation is consistent with the conclusions of the work presented here and support the hypothesis that Tbra-GFP high cells may represent mesoderm committed cells. As mentioned earlier, inhibition of endogenous Wnt activity have been shown to decrease the expression of primitive streak markers. Kurek *et al.* (2015) suggest that Wnt inhibition enhances the self-renewal of EpiSCs and state that some cell lines, including the Tbra-GFP EpiSC line,

could not be maintained in an undifferentiated state, unless Wnt inhibitors were added. This is rather surprising, as addition of any supplement apart from Fgf2 and Activin A was never necessary for us in order to keep EpiSC lines in a self-renewing condition. The reason behind such a discrepancy is not yet understood.

Early lineage specification in EpiSC cultures has also been studied by a former member of our lab (Tsakiridis *et al.*, 2014). In that study, our group made use of a different Tbra reporter line, the Tps/tb-RED reporter line, that contains a dsRed2 transgene under the transcriptional control of a randomly integrated *Tbra* promoter fragment that is active specifically in the primitive streak and the tail bud, but not the notochord (Clements *et al.*, 1996; Economou *et al.*, 2015). Given that GFP and RED mark different fractions of *Tbra* expressing EpiSCs, comparison of the two studies allowed us to dissect further the EpiSC heterogeneity and identify subsets of $Tbra^{+}$ cells with distinct properties. EpiSC cultures were shown to contain 15-30% RED^{+} cells, as opposed to 30-60% GFP^{+} cells. Expression analysis in sorted RED positive and negative fractions indicated a similar expression pattern of primitive streak and neural markers as the one presented here and like the GFP subpopulations, RED positive and negative fractions were able to interconvert. Generation of embryoid body-like aggregates from sorted subpopulations demonstrated that RED^{+} cells have strong bias towards mesoderm differentiation, while when sorted fractions were plated in neural inducing conditions (N2B27 medium supplemented with the MEK/Erk and Activin/Nodal inhibitors) expression of *Sox1* and *Sox2* was higher at the RED^{-} cultures, but not significantly.

Thus, altogether, it is suggested that $Tbra^{+}$ EpiSCs represent a major fraction of cells that have not exited pluripotency, but have adopted an early primitive streak-like character. Similar to the cells of the posterior part of the gastrula staged epiblast, these EpiSCs are at a reversible stage where they exhibit bias towards the mesoderm/endoderm lineages, but at the same time they retain their ability to undergo neural differentiation when exposed to neural inducing conditions. Interestingly, Tsakiridis *et al.* (2014) demonstrated that in EpiSC cultures under self-renewing

conditions, *Tbra* expression is mutually exclusive with expression of *Sox1-GFP*, that seems to mark a distinct subpopulation with neural characteristics.

The Tbra-GFP$^+$ EpiSC fraction may contain subsets of cells with distinct properties

A small fraction of cells (7%) was shown to be Nanog$^-$, but Tbra$^+$/Sox2$^+$. Could these cells resemble the NMPs of the E8.5 NSB? Due to lack of a Tbra/Sox2 double reporter cell line at the time this study was conducted, besides the lack of unique markers for NMPs (see section 1.3.3), this study does not provide direct evidence for that. However, NMP-related transcripts were detected in EpiSC cultures. In addition, the study of Huang *et al.* (2012) has shown that when EpiSCs were grafted into the NSB of two to five somite-stage (E8.5) embryos, cell integration was compromised, but a few dispersed cells were observed. Given that EpiSCs seem to be compatible with the postimplantion epiblast at stages before pluripotency is lost, it is likely that the few cells that can incorporate at the E8.5 embryos are a minor fraction of cells that have exited pluripotency and possibly have progressed into the NMP state. Hence, it is possible that the Tbra$^+$/Sox2$^+$/Nanog$^-$ cells represent NMP-like cells. However, as it is discussed later on, a more extensive expression analysis at the single cell level is necessary in order to validate this hypothesis.

Comparison of the expression profile of Tbra-GFP and Tps/tb-RED fractions demonstrates that a subset of the Tbra$^+$ EpiSCs resembles the organiser of the late gastrula staged embryo, the node and its derivative the notochord. *In vivo*, the node and the notochord are marked by expression of *Noto*, *Shh*, *Noggin* and *Chordin*. As it would be expected, expression of dsRED, that is designed not to label node/notochord cells, did not exhibit strong correlation with expression of these genes. In contrast to that, Tbra-GFP$^+$ cells were characterised by enrichment in *Noto*, *Shh*, *Noggin* and *Chordin* transcripts, suggesting that at least some of the Tbra$^+$ cells have a node/notochord-like character. Since cells that ingress through the node to form the

notochord switch off Nanog, we would expect that the node like cells are found in the $Tbra^{+}Nanog^{+}$ fraction, while it is possible that some of them progress further into lineage specification, they switch off Nanog activity and they become notochord like cells. Interestingly, preliminary data indicate that EpiSC cultures contain active Shh signalling and suggest that the node/notochord-like Tbra-GFP^{+} cells may be the main source of this morphogen.

As mentioned above, the Tbra-GFP^{+} cells were shown to be enriched in *Dkk1* and *Cer1* transcripts. However, *in vivo*, *Dkk1* and *Cer1* is not expressed by the pluripotent epiblast, but by the anterior mesendoderm, which rises from the node. Hence, we would expect that *Dkk1* and *Cer1* is expressed by the $Tbra^{+}/Sox2^{-}/Nanog^{-}$ EpiSC fraction, that most likely include cells that have progressed further into lineage specification and resemble cells of early mesendoderm.

Fluctuation between the different cell states

Based on detection of Nanog protein in the vast majority of the *Tbra* expressing cells, it was concluded that the $Tbra^{+}/Nanog^{+}$ cells represent pre-ingression epiblast cells. Importantly, it was shown that the Tbra-GFP positive and negative EpiSCs are in a dynamic equilibrium and can interconvert, suggesting that these cells are at a reversible state. Based on the observation that a small proportion of cells was $Tbra^{+}/Nanog^{-}/Sox2^{-}$, it was speculated that a small subset of cells progresses further into lineage specification, while the detection of certain tissue specific markers suggested that these cells could represent notochord-like and mesendoderm-like subpopulations. Similarly, it has been suggested that the $Tbra^{+}/Sox2^{+}/Nanog^{-}$ cells may represent a minor NMP-like fraction. However, it is not known whether the cells that seem to resemble post-ingression epiblast cells are at a reversible stage or they are lineage committed. A previous study in ES cultures has shown that Nanog negative cells can revert to a state of Nanog positivity (Chambers *et al.*, 2007). Although we do not know whether the same equilibrium exists in EpiSC cultures, it is implied that *in*

vitro downregulation of Nanog does not necessarily mark irreversible exit from the pluripotent epiblast state, and that some of the subpopulations we described here may retain the ability to reverse to a $Tbra^-/Nanog^+$ state (figure 3.8).

Interconversion between distinct subpopulations has been reported in the past, in studies in EpiSCs (Hayashi and Surani, 2009; Han *et al.*, 2010; Tsakiridis *et al.*, 2014) and ES cells (Hayashi *et al.*, 2008). In the future, live imaging microscopy using a *Tbra* reporter cell line would complement these data and would help us to analyse how the *Tbra* gene activity possibly oscillates within single EpiSCs. According to a model that attempts to explain the heterogeneity in stem cell cultures, oscillatory expression of synergistically and antagonistically acting transcription factors cause stem cells to fluctuate between different phenotypical states and develop distinct differentiation biases (reviewed in Graf and Stadtfeld, 2008). It is believed that fluctuating levels of pluripotency factors offer the opportunity to the cells to respond to environmental signals and go into differentiation (Osorno and Chambers, 2011). For example, it has been shown that ES cells expressing low levels of Nanog have an increased propensity to differentiate (Chambers *et al.*, 2007). That agrees with the data shown in figure 3.1 and 3.4, presenting negative correlation between the protein levels of Tbra and each of the pluripotency factors, Nanog and Sox2. What generates stochastic fluctuations of transcription factors in the first place and how the environment affects the balances between the different subpopulations is not yet understood.

Next steps

In this study, the expression profile of $Tbra^+$ EpiSCs and the identification of distinct cell subpopulations have been performed using a combination of immunocytochemistry and QRT-PCR in two different reporter lines. Nevertheless, our data demonstrate that pluripotency and lineage specification markers are expressed in EpiSC cultures in a very heterogeneous way, therefore the model presented here needs to be validated by expression analysis at the single cell level. Hence, in the future,

single cell QRT-PCR will allow us to dissect further the expression patterns this study revealed and will offer a more detailed description of the different cell subsets that EpiSC cultures contain under self-renewing conditions.

Importantly, the different groups of cells that have been described here seem to express regulators of signalling pathways involved in tissue specification. More specifically, Shh plays a key role in the dorso-ventral patterning of the neural tube (reviewed in Dessaud *et al.*, 2008), while the BMP antagonists Chordin and Noggin are important in forebrain development (Bachiller *et al.*, 2000). The Wnt/Nodal inhibitors Cer1 and Dkk1 are also involved in maintaining the neuroectoderm layer of the gastrula-staged epiblast (reviewed in Arnold and Robertson, 2009). The presence of these factors within the Tbra-GFP$^+$ cell population suggests that the different subsets of cells that we have described here could pattern adjacent cells in a culture dish. For example, a previous study from our lab has shown that EpiSC cultures contain a Sox1-GFP$^+$ neural-like subpopulation that possibly contains cells committed to neural differentiation (Tsakiridis *et al.*, 2014). Hence, in the future, it will be interesting to investigate the hypothesis that Tbra$^+$ cells induce neural character to their neighbour cells, by using image analysis software that will allow us to model the relative positions of Tbra$^+$ cells and cells that express neural markers. Moreover, it will be interesting to use a Noto-GFP reporter line (Winzi *et al.*, 2011) and investigate the patterning effects that the node/notochord-like cells potentially have on their neighbour cells, under self-renewing but also in neural differentiation conditions.

A recent study presented a differentiation protocol that generates node/notochord like cells from mouse ES cells (Winzi *et al.*, 2011). In the first step of the protocol, ES cells are treated with Activin for 3 days and that results to generation of Tbra$^+$Foxa2$^+$ cells. Then, cultures are treated with Activin, Fgf2, Noggin, Dkk1 and a RA receptor inhibitor for 2 days. As a result, Noto is upregulated; however, the efficiency of the differentiation protocol is very low and Noto-GFP$^+$ cells account only for 10% of the culture. Given that the post-implantation epiblast cells are the immediate precursors of node/notochord and most importantly, based on the finding that suggest that EpiSC

cultures already contain a subset of node/notochord-like cells, it seems likely that EpiSCs will serve as a better *in vitro* system for the development of more efficient differentiation protocols and the generation of more homogeneous node/notochord-like cultures that will help us to investigate the gene networks that are active in these developmentally important structures. Hence, in the future, we could use a combination of the Noto-GFP and Tps/tb-dsRED reporter lines in order to investigate the node/notochord-like EpiSC fraction in more detail and identify culture condition that efficiently converts EpiSCs to node/notochord like cells.

Overall, our data enhance the notion that EpiSC cultures are an *in vitro* system that is physiologically relevant to the gastrulation stage epiblast, therefore they can serve as a good model for the development of differentiation protocols and the analysis of early lineage specification events. In the next chapter, I focus on *in vitro* NMPs generated from EpiSC cultures and I present how the correlation of *Tbra*, *Sox2* and *Sox1* expression changes as the self-renewing EpiSCs turn into bipotent NMPs.

Chapter 4: Assessing Sox1 as a marker for late NMPs

4.1. Introduction

NMPs are characterized by coexpression of the mesodermal marker *Tbra* and the neural/pluripotency marker *Sox2* (Wymeersch *et al.*, 2016). Expression of *Sox1*, another neural specific marker, has also been detected in NMP regions. Examination of Sox1-GFP mice (Aubert *et al.*, 2003) demonstrated that at E8.5, expression of Sox1-GFP spreads from the anterior end of the embryo to the rostral half of the node (Cambray and Wilson, 2007). Importantly, Sox1-GFP was also detected in the CLE of late E8.5 embryos (N. Cambray, unpublished data). At E10, Sox1-GFP, as well as *Sox1* mRNA was detected at relatively high levels in the CNH. Sox1-GFP was also found at lower levels in the TBM, but *Sox1* mRNA was almost undetectable in this region (Cambray and Wilson, 2007). Interestingly, cells with high Sox1-GFP levels were present at the caudal end of the notochord, while some were continuous with two ventrolateral horns of the neural tube, possibly representing cell migration from the axial progenitor region to the elongating notochord (Cambray and Wilson, 2007). In summary, *in vivo* data suggest that NMPs upregulate *Sox1* as they transit from the primitive streak (early NMPs) to the tail bud stage (late NMPs).

As described in the previous chapter, EpiSC cultures contain $Tbra^{+}Sox2^{+}$ cells. However, the core pluripotency factor Nanog is detected in most of these cells, while grafting experiments demonstrated that EpiSCs do not incorporate efficiently when transplanted into the NSB of E8.5 embryos (Huang *et al.*, 2012), suggesting that at least the vast majority of $Tbra^{+}Sox2^{+}$ EpiSCs are pluripotent rather than bipotent NMPs (discussed in more detail in chapter 3). Under self-renewing conditions, expression of *Tbra* and *Sox1-GFP* are mutually exclusive (Tsakiridis *et al.*, 2014). As presented in chapter 3, Tbra is detected in 30-60% of EpiSCs and marks a dynamic subpopulation of cells that is primed, but not committed, to mesoderm/endoderm differentiation. In Sox1-GFP EpiSCs, GFP is expressed in 20-30% of the population. Based on the failure of Sox1-GFP positive cells to undergo efficient interconversion

between positive and negative cells and to propagate efficiently as EpiSCs, it has been suggested that the Sox1-GFP^{+} EpiSC fraction is poised to undergo neural differentiation and may contain cells that have already progressed into neural commitment (Tsakiridis *et al.*, 2014).

Recently, it was shown that NMPs can be generated from both mouse EpiSCs and human ES cells upon a 48-72h treatment with Fgf2 and the Gsk-3 antagonist/Wnt agonist CHIRON99021 (thereafter referred to as FGF/CHI) (Gouti *et al.*, 2014). In these conditions, expression of *Nanog* becomes undetectable, whereas expression of genes active in caudal progenitor regions, such as *Wnt3a*, *Cdx2* and *Nkx1.2*, as well as *Hox* genes, increase. The changes in the transcriptome of FGF/CHI treated EpiSCs coincides with an increase in the fraction of Tbra^{+}Sox2^{+} cells in the culture to around 80%. *Sox1* RNA has been detected in NMP cultures (Gouti *et al.*, 2014), but the correlation between *Sox1* and *Tbra* expression has not been investigated so far.

Given that Sox1 is present in the caudal progenitor regions *in vivo*, we hypothesized that the NMP cultures could contain Tbra^{+}Sox1^{+} NM bipotent cells. Most importantly, the upregulation of *Sox1* at the tail bud stages raised the hypothesis that *Sox1* expression in NMP cultures could mark the transition from an early- to a late-like NMP state *in vitro*. Based on all the above, the aim of this project was to:

- investigate whether *Sox1* expressing cells found in NMP cultures exhibit an NM bipotent character,
- assess whether *Sox1* expression marks the early to late NMPs transition *in vitro*.

4.2. Tbra/ Sox1-GFP double positive cells emerge in NMP cultures

Expression analysis in microdissected E8.5 NSB and E10.5 CNH regions demonstrated an increase of *Sox1* mRNA levels at the tail bud stage, in contrast to Sox2, the expression levels of which remained stable (figure 4.1- Costas Economou, unpublished data). These data are consistent with the findings of Cambray and Wilson (2007) described above and support the hypothesis that *Sox1* expression marks late NMPs.

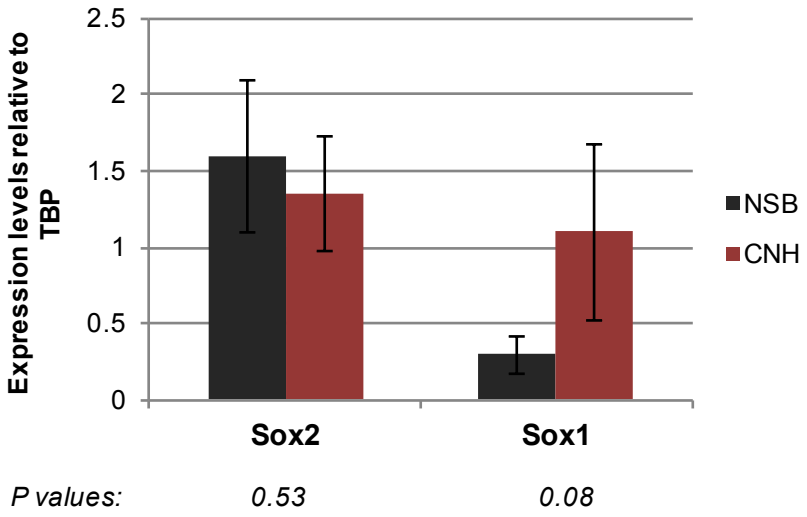


Figure 4.1: Expression of Sox2 and Sox1 in the NMP regions of the mouse embryo. QRT-PCR analysis of Sox2 and Sox1 in the NSB and the CNH of E8.5 and E10.5 embryos, respectively. Error bars represent standard deviation. N=2-6. Statistical significance of the data was assessed performing t-test (two-tailed, unpaired, assuming equal variances); the p values are presented on the bottom of each graph. Data generated by Costas Economou.

As mentioned before, *in vitro* NMPs can be generated upon treatment of EpiSCs with FGF/CHI for 48-72h. In order to analyse the *Sox1* expression levels in NMP cultures, a Sox1-GFP ES cell reporter line was used (Aubert *et al.*, 2003) and an EpiSC line was generated from it. FACS analysis of EpiSC cultures treated with FGF/CHI demonstrated that the number of Sox1-GFP^{+} cells increased from 25% in EpiSCs to 40% and 45% in 48h and 72h NMP cultures, respectively (figure 4.2). In this study, I focused on the analysis of Sox1-expressing cells in 72h NMPs.

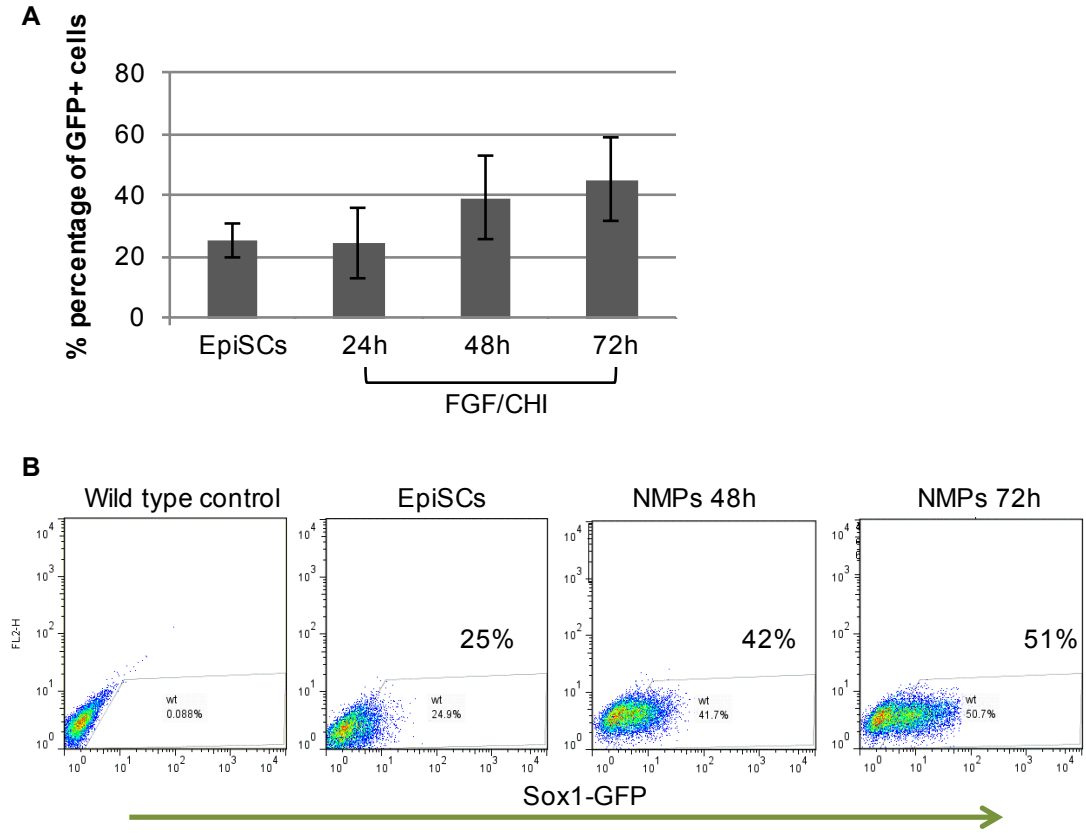
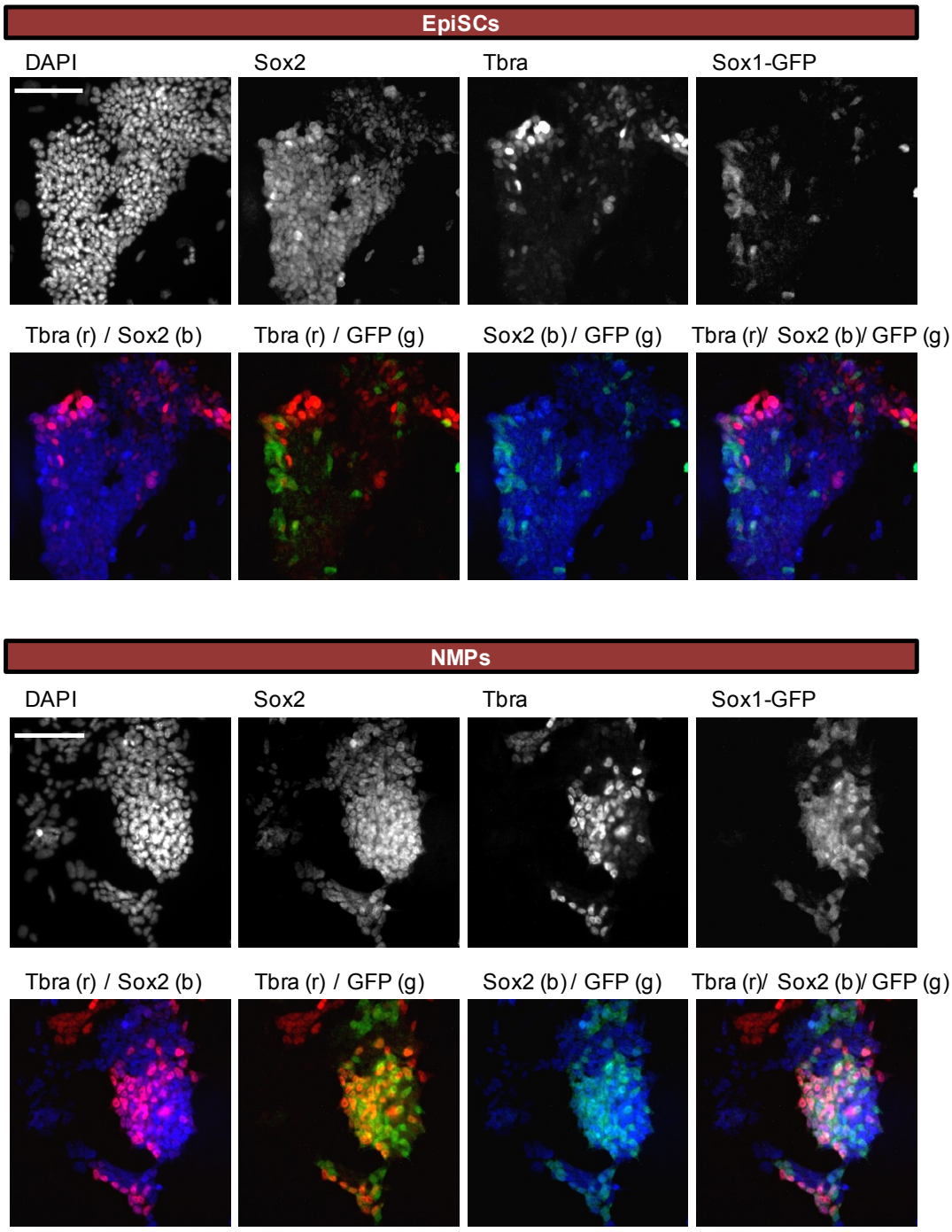


Figure 4.2: Sox1-GFP expression in *in vitro* Sox1-GFP NMP cultures. A) Average percentage (%) of Sox1-GFP cells in EpiSCs and in EpiSC cultures treated with FGF/CHI for 24-72 hours, estimated by FACS analysis. Error bars represent standard deviation. N=6. **B)** FACS analysis of EpiSC, 48h NMP and 72h NMP Sox1-GFP cultures.

Triple immunocytochemistry for Sox1-GFP, Sox2 and Tbra in EpiSCs and 72h NMP cultures, followed by single cell image quantification (figure 4.3 and S4.1), confirmed the mutually exclusive expression pattern of *Tbra* and *Sox1-GFP* that has previously been observed in EpiSCs (Tsakiridis *et al.*, 2014) and shed light on the correlation of Sox1-GFP with Tbra and Sox2, in *in vitro* NMPs. Due to lack of a reliable anti-Sox1 antibody, immunocytochemistry was performed in the Sox1-GFP reporter line, using an antibody that recognises the GFP protein. In NMP conditions, the number of $Tbra^{+}$ cells rose compared to EpiSC cultures, while the levels of $Sox2^{+}$ cells remained the same (figure 4.3, C). An increase of Tbra/Sox2 double positive cells in NMP conditions from ~20% to ~50% was confirmed. In agreement with the FACS analysis shown in figure 4.2, treatment of EpiSCs with FGF/CHI led to an increased number of Sox1-GFP expressing cells (figure 4.3, B and C). Crucially, in NMP cultures, a large fraction of Tbra/Sox1 double positive cells emerged (~37%). The majority of $Sox1^{+}$ cells seemed to express lower levels of Tbra, although $Sox1^{high}$/ $Tbra^{high}$ cells were also observed. Sox2 was detected in almost the entire Sox1-GFP^{+} fraction, with the levels of the two proteins showing an apparently positive correlation between them. Interestingly, Sox1-GFP was not detected in all $Tbra^{+}Sox2^{+}$ cells, but in approximately 75% of the Tbra/Sox2 double positive NMPs. More specifically, around 51% of the cells exhibited coexpression of Tbra and Sox2, but only 37% of the culture was shown to be Tbra/Sox2/Sox1 triple positive (figure 4.3, D and S4.1).

A


EpiSCs
DAPI
Sox2
Tbra
Sox1-GFP
Tbra (r) / Sox2 (b)
Tbra (r) / GFP (g)
Sox2 (b) / GFP (g)
Tbra (r)/ Sox2 (b)/ GFP (g)
NMPs
DAPI
Sox2
Tbra
Sox1-GFP
Tbra (r) / Sox2 (b)
Tbra (r) / GFP (g)
Sox2 (b) / GFP (g)
Tbra (r)/ Sox2 (b)/ GFP (g)

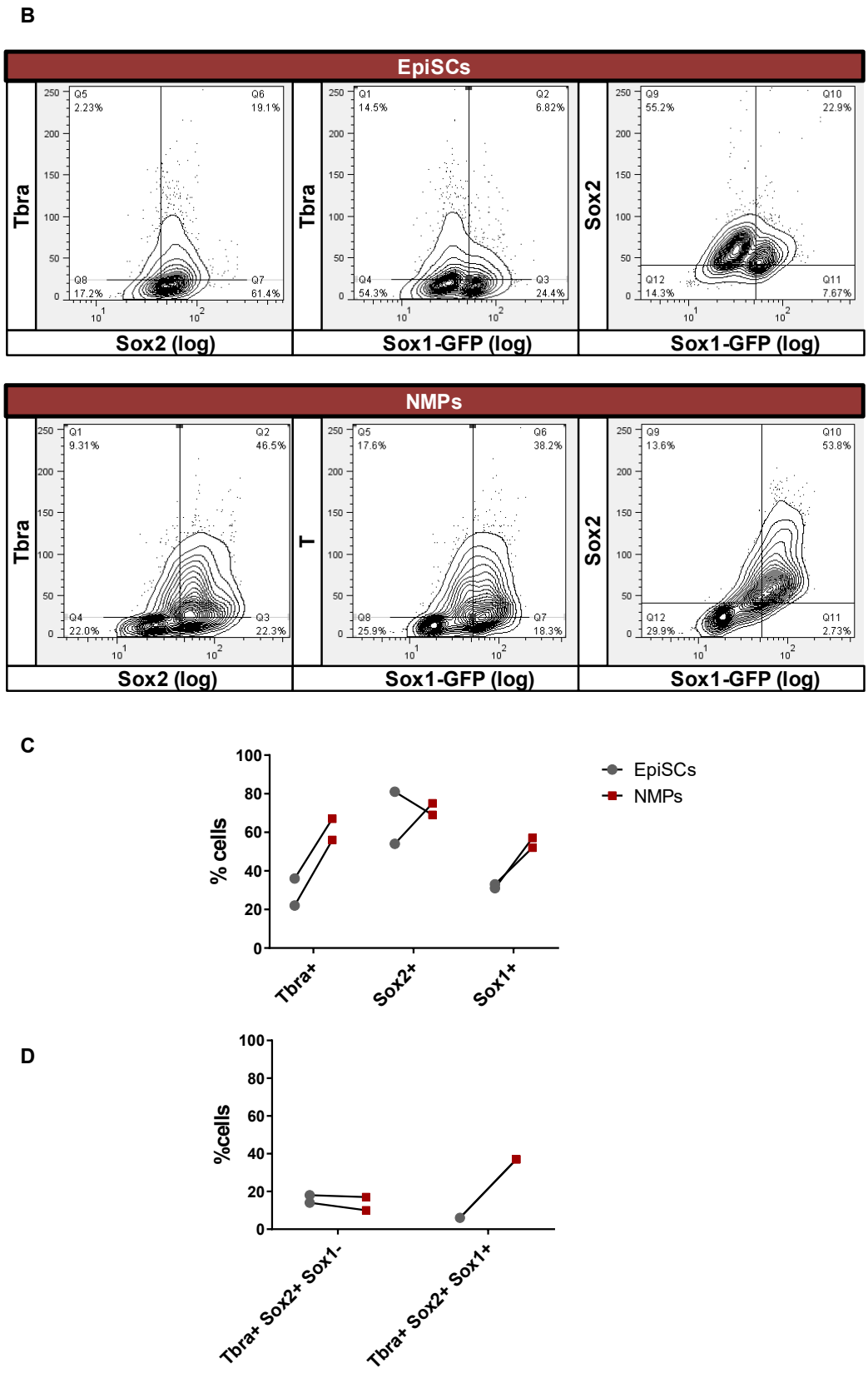


Figure 4.3: Correlation of Sox1-GFP, Sox2 and Tbra expression patterns in EpiSC and NMP cultures.

Figure 4.3: Correlation of Sox1-GFP, Sox2 and Tbra expression patterns in EpiSC and NMP cultures. A) Triple immunocytochemistry for Sox1-GFP, Sox2 and Tbra in EpiSC and 72h NMP cultures. r, b and g denote the red, blue and green channel, respectively. **B-D)** Immunofluorescence quantification following single cell image analysis. Expression pattern of Sox1-GFP, Sox2 and Tbra and the correlation among them; analysis is shown in contour plots, with x-axis in log scale (B). Percentage (%) of $Tbra^{+}$, $Sox2^{+}$ and Sox1-GFP^{+} cells in EpiSC and NMP cultures (C). Percentage (%) of $Tbra^{+}Sox2^{+}Sox1^{-}$ and $Tbra^{+}Sox2^{+}Sox1^{+}$ cells in EpiSC and NMP conditions (D). N=2. Scale bars: 100μm. A more detailed analysis of image quantification is shown in supplementary figure S4.1.

4.3. Sox1-GFP positive cells express NMP markers and higher levels of Hox genes

Sox1 is a marker of proliferating neural precursors (Pevny *et al.*, 1998). In order to address whether Sox1-GFP marks NMPs or cells that have exited the NMP state and have progressed into neural commitment, 72h NMP cultures were FACS sorted and the mRNA levels of neural and NMP markers were analysed by QRT-PCR (figure 4.4). It was observed that sorted GFP positive and negative fractions exhibited the same levels of the caudal progenitor region markers *Evx1*, *Cdx2*, *Wnt3a* and *Nkx1.2*. High levels of *Tbra* transcript were detected in both fractions, but a small enrichment of *Tbra* mRNA was observed in the Sox1-GFP negative cells. Expression of *Sox2* and the neuronal marker class *III b-tubulin* (also known as *Tuj1*) did not exhibit any differences between the two subpopulations, while the posterior neural marker *Gastrulation brain homeobox 2* (*Gbx2*), expression of which expands over the NMP regions (Rhinn *et al.*, 2004), was shown to be upregulated in the Sox1-GFP positive cells. In summary, Sox1-GFP cells exhibit the expression profile of NMPs, while the upregulation of *Gbx2* in the Sox1-GFP^{+} fraction, in combination with the detection of a number of Sox2/Sox1-GFP double positive, but Tbra negative cells (figure S4.1), implies that this subpopulation may also contain a group of neural committed cells. Hence, the expression analysis data, collectively, suggest that *Sox1-GFP* is expressed

in NMPs, but possibly also in cells that are likely to have progressed into neural commitment.

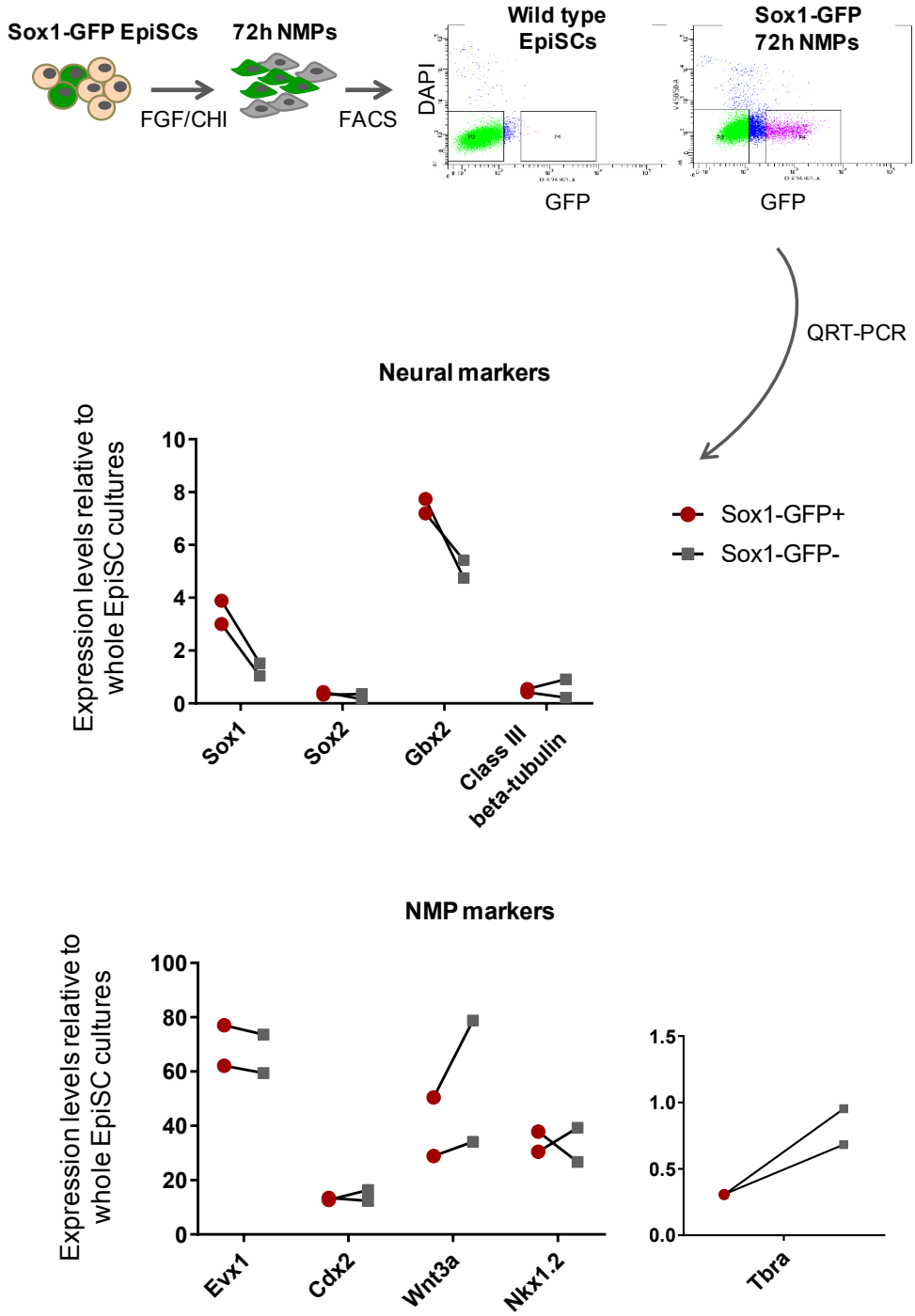


Figure 4.4: Expression analysis of Sox1-GFP fraction in 72h NMP cultures. QRT-PCR analysis of neural and NMP markers in FACS sorted Sox1-GFP positive and negative 72h NMP fractions. Values are normalised to whole EpiSC cultures. N=2.

In vivo, Sox1 expression is upregulated in NMP regions during the transition from primitive streak to tail bud (Cambray and Wilson, 2007). Hence, it was hypothesized that Sox1-GFP could mark cells that constitute the *in vitro* equivalent of tail-bud-stage axial progenitors. Interestingly, during this transition, major *Hox* gene alterations also occur, with many *Hox* genes from different paralogous groups being upregulated between E8.5 and E10.5 (unpublished data, see section 1.3.3). Based on these data, the hypothesis that Sox1-GFP cells represent late NMPs was tested by performing expression analysis of a number of *Hox* genes in FACS sorted GFP positive and negative NMP cultures. QRT-PCR demonstrated enrichment of the Sox1-GFP$^+$ fraction in the transcripts of most of the *Hox* genes that were upregulated in NMP conditions (figure 4.5). More specifically, Sox1-GFP$^+$ cells exhibited higher transcript levels of *HoxB1*, *HoxA4* and *HoxC5*. *Hox* genes from the paralogous groups 8, 9 and 10 showed strong induction under FGF/CHI and almost all of them (6 out of 7) were expressed at higher levels in Sox1-GFP$^+$ cells. *HoxA5* exhibited very low upregulation compared to EpiSC cultures and the same expression levels between the Sox1-GFP positive and negative fractions. Very low induction was observed for three *Hox* genes belonging to the 5' paralogous groups 11-13, all of which exhibited similar transcript levels in the two subpopulations. *HoxC10* was one of the most highly upregulated genes in the Sox1-GFP positive NMP fraction; detection of HoxC10 protein in Sox1-GFP/Tbra double positive NMPs was confirmed by immmunocytochemistry (figure 4.5, B). Overall, the enrichment of Sox1-GFP$^+$ cells in *Hox* genes, especially posterior *Hox* genes of the paralogous groups 9-10, resemble the alteration of *Hox* gene activation that take place in the caudal progenitor regions during the transition from early NMPs (E8.5) to late NMPs (E9.5-10.5), and hence support the hypothesis that Sox1-GFP marks NMPs that correspond to the axial progenitors found at the tail bud stages.

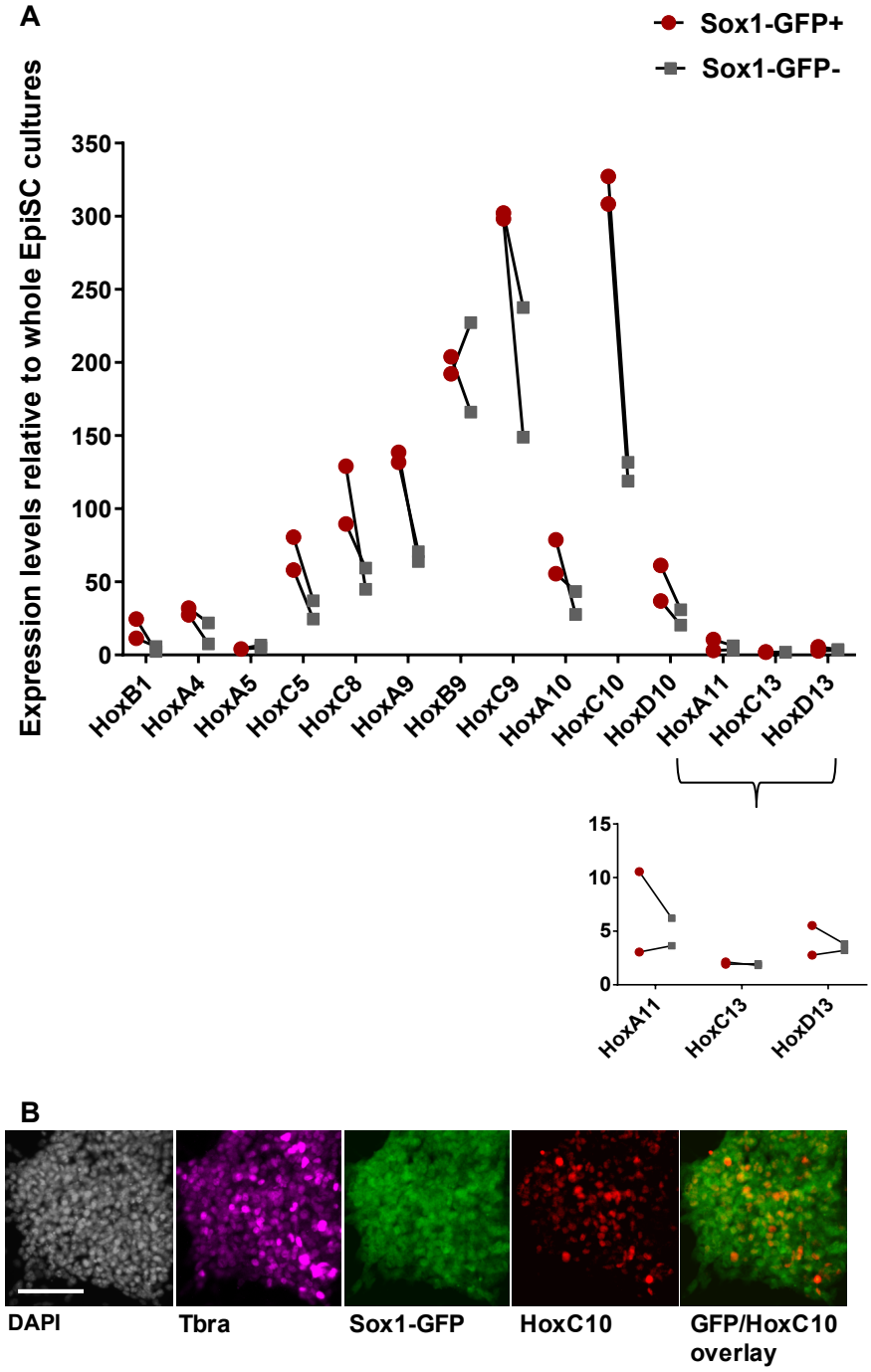


Figure 4.5: Hox gene expression in sorted Sox1-GFP fractions of 72h NMP cultures. A) QRT-PCR analysis of *Hox* genes in FACS sorted Sox1-GFP positive and negative 72h NMP fractions. FACS analysis of the cell sorting is shown in figure 4.4. N=2. **B)** Tbra, Sox1-GFP and HoxC10 immunocytochemistry in 72h NMPs. Scale bar: 100μm.

4.4. Fate of Sox1-GFP positive cells under prolonged culture in FGF/CHI

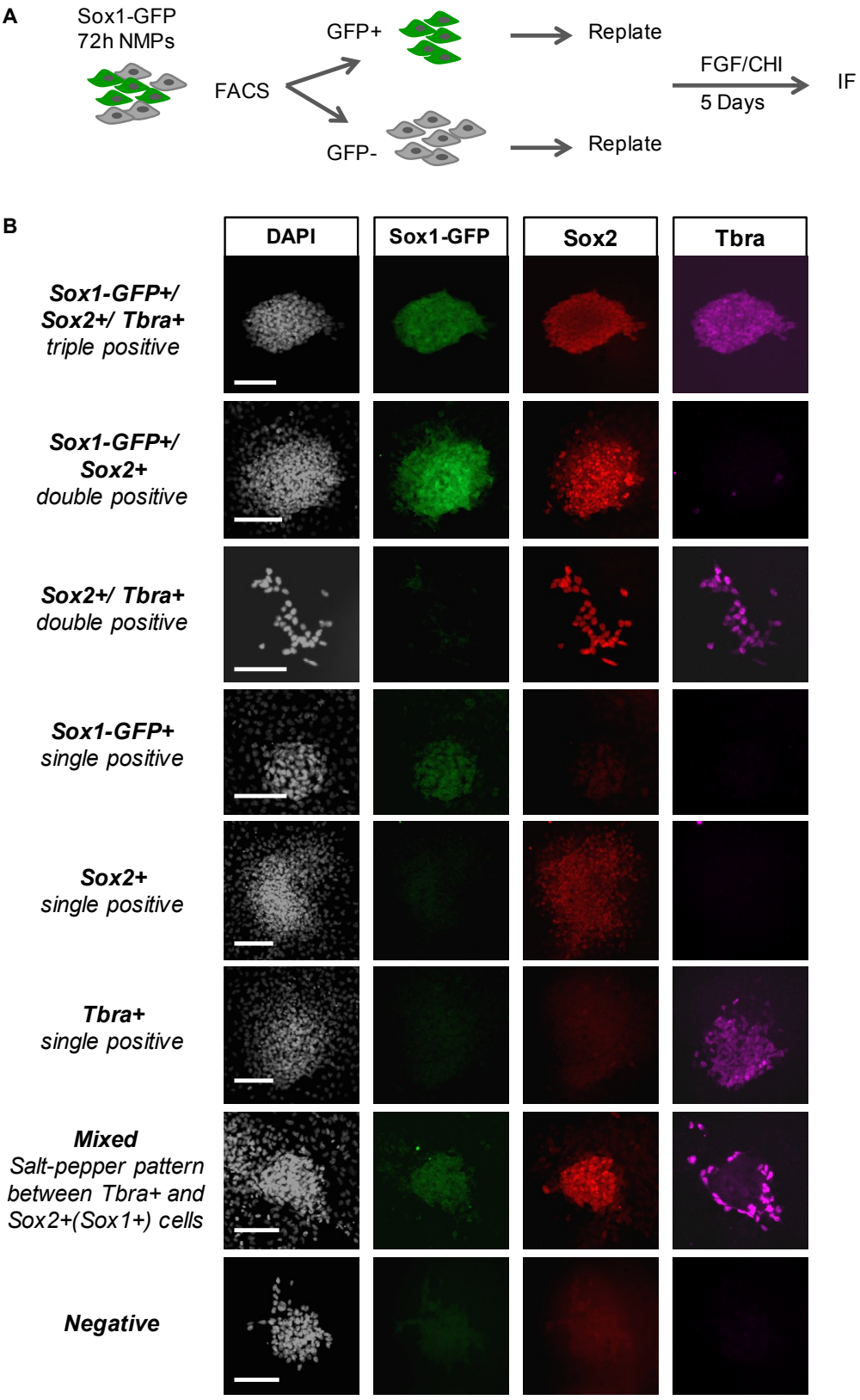
It has been previously shown that prolonged (i.e. >72 hours) culture in FGF/CHIR induces further differentiation of NMPs into mutually exclusive mesoderm (Tbra or Tbx6 single positive) and neuroectoderm (Sox2 single positive) cells (Gouti *et al.*, 2014; Tsakiridis and Wilson, 2015). Hence, these culture conditions allow NMPs to progress into both mesoderm and neural lineage specification and give rise to their natural differentiation products. Based on that, the NM potency of single Sox1-GFP$^+$ cells was assessed through prolonged culture in FGF/CHI. More specifically, 72h NMP cultures were FACS sorted and the GFP positive and negative fractions were replated at clonal density (Tsakiridis and Wilson, 2015). After 5 days, the resulting colonies were analysed by immunofluorescence for Tbra, Sox2 and Sox1-GFP and categorized based on their composition. The results are presented in figure 4.6. In summary:

- A large fraction of Tbra$^+$Sox1$^+$Sox2$^+$ and Tbra$^+$Sox2$^+$ colonies, that most likely represent self-renewing NMPs, were detected in one of the two replicates of the experiment. In this experiment, the GFP negative fraction gave rise to Tbra$^+$Sox1$^+$Sox2$^+$ (21%) and Tbra$^+$Sox2$^+$ (26%) colonies, while the GFP positive cells formed Tbra$^+$Sox1$^+$Sox2$^+$ (24%) clones only. When this experiment was repeated, only one Tbra$^+$Sox2$^+$ colony was observed. Tbra$^+$Sox1$^+$ cells were not observed in any of the two experiments (figure S4.2).
- 38% of the clones obtained from the GFP negative fraction were composed of Sox1/ Sox2 single or double positive cells (11% Sox1$^+$Sox2$^+$, 19% Sox2$^+$, 8% Sox1$^+$); strikingly, the neural-like colonies that resulted from GFP positive cells accounted for 71% of the clones (26% Sox1$^+$Sox2$^+$, 5% Sox2$^+$, 40% Sox1$^+$), compared to 38% from GFP negative cells, indicating a stronger neurogenic capacity of the Sox1-GFP positive fraction.
- Tbra single positive cells were detected in 24% of the colonies that were formed from GFP negative cells, but only in 1% of the clones obtained from the GFP positive cells. The purity tests of the FACS sorted GFP positive fractions demonstrated 1-4% contamination with GFP negative cells (figure S4.2, A). Hence, it is very

possible that the Tbra$^+$ mesoderm-like colonies detected in the GFP$^+$-derived cultures resulted from contaminating Sox1-GFP$^-$ cells.

- As mentioned above, Tbra$^+$Sox1$^+$Sox2$^+$ and Tbra$^+$Sox2$^+$ cells represent self-renewing NMPs, while colonies consisting of single positive cells indicate unilinear differentiation into either mesoderm or neuroectoderm. Crucially, neuromesodermal potency is demonstrated by clones consisting of a mix of single positive Tbra$^+$ and Sox2$^+$ or Sox1$^+$ cells. In this study, clones exhibiting a combination of single positive cells accounted for 12% of the colonies that raised from GFP negative cells, but only for 2% of the clones that emerged from the GFP positive fraction.
- Finally, both fractions gave rise to 15-20% of colonies that were negative for all three markers (15-20%), possibly reflecting cells that have progressed further into lineage differentiation and therefore have downregulated the three transcription factors.

Collectively, these data suggest that the Sox1-GFP negative fraction contains bipotent NM progenitors, but also cells committed to neural or mesoderm differentiation. However, under these culture conditions, the Sox1-GFP positive cells were shown to be strongly biased towards neuroectoderm lineage specification, whereas there is no adequate evidence to support the notion that Sox1-GFP$^+$ cells can undergo mesoderm differentiation. Although a very small number of Tbra$^+$ and mixed colonies (1% of each category) was detected in the GFP positive-derived colonies, we cannot exclude the possibility that the Tbra$^+$ cells were produced by the contaminating GFP negative cells (1-4%). In order to assess whether the low cell density could potentially have affected the ability of Sox1-GFP$^+$ cells to undergo mesoderm differentiation, the same experiment was performed plating sorted cells at a higher density. However, the same results were observed (data not shown), suggesting that under these culture conditions, the ability of Sox1-GFP$^+$ cells to give rise to Tbra$^+$ cells was not affected by cell density.


A
Sox1-GFP
72h NMPs
FACS
GFP+
GFP-
Replate
Replate
FGF/CHI
5 Days
IF
B
DAPI
Sox1-GFP
Sox2
Tbra
Sox1-GFP+/ Sox2+/ Tbra+ triple positive
Sox1-GFP+/ Sox2+ double positive
Sox2+/ Tbra+ double positive
Sox1-GFP+ single positive
Sox2+ single positive
Tbra+ single positive
Mixed Salt-pepper pattern between Tbra+ and Sox2+(Sox1+) cells
Negative

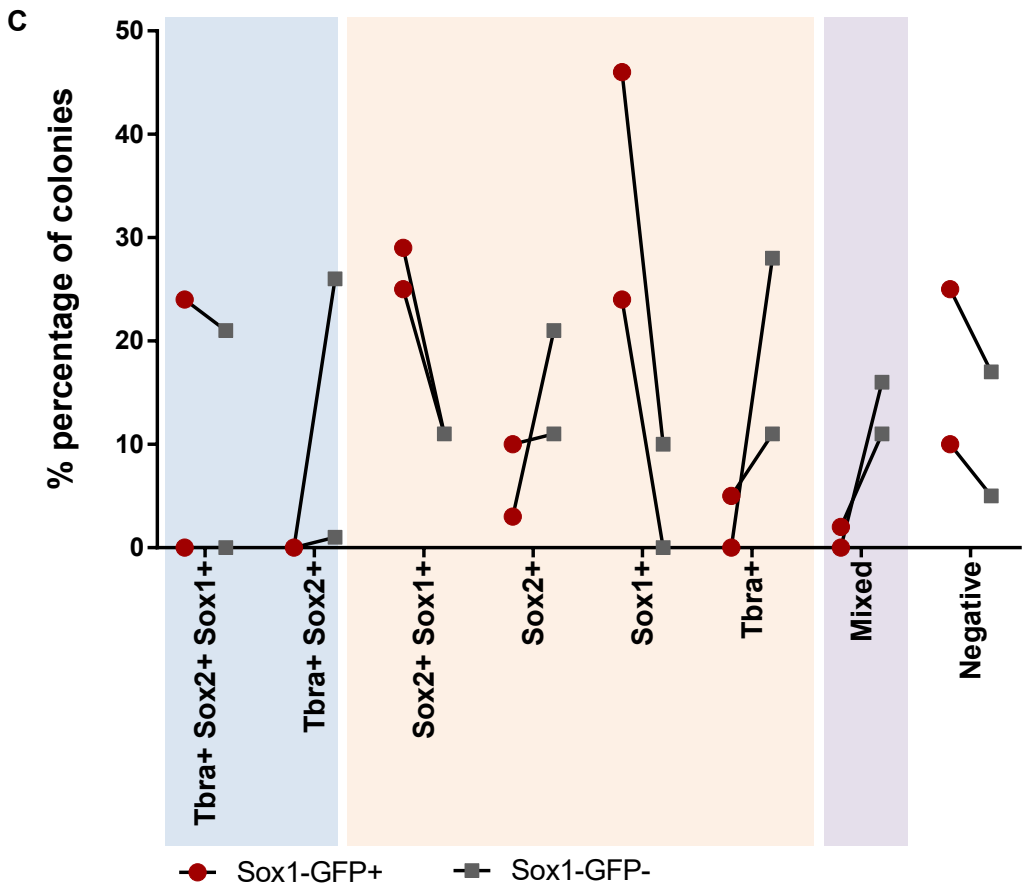


Figure 4.6: Fate of Sox1-GFP positive and negative cells under prolonged FGF/CHI treatment. Scheme depicting FACS sorting of Sox1-GFP positive and negative cells of 72h NMP cultures, followed by replating at clonal density and culture in FGF/CHI for 5 days. **B)** Immunofluorescence analysis for Tbra, Sox1-GFP and Sox2, showing representative examples of the clones obtained after culture of single sorted cells in FGF/CHI for 5 days. **C)** Percentages of the different types of colonies obtained after clonal plating of Sox1-GFP^{+} and Sox1-GFP^{-} cells. Mixed colonies consist of single Tbra^{+} cells mixed with single Sox2^{+} or single Sox1^{+} or Sox2^{+}Sox1^{+} double positive cells. Scale bars: 100µm. N=2. Analysis of each individual experiment is shown in supplementary figure S4.2 (see appendix).

4.5. Discussion

Sox1 is expressed in proliferating neural precursors and it has been shown to play a key role in maintaining these progenitors in the undifferentiated state, by counteracting the activity of proneural proteins and inhibiting neurogenesis (Bylund *et al.*, 2003). Hence, expression of *Sox1-GFP* has been used as a readout for the successful development of neural differentiation protocols and the purification of neural precursor cells (Aubert *et al.*, 2003; Ying *et al.*, 2003; Chung *et al.*, 2006; Abranches *et al.*, 2009). Sox1 is one of the early lineage specification markers expressed in self-renewing EpiSC cultures (Tsakiridis *et al.*, 2014). More specifically, Sox1-GFP was detected in 20-30% of EpiSCs, with half of this population being positive for Nanog or Oct4 (Tsakiridis *et al.*, 2014). Sorted Sox1-GFP$^+$ cells were shown to contain lower transcript levels of primitive streak markers, but higher levels of the neural marker *Pax6*, as compared to their negative counterparts (Tsakiridis *et al.*, 2014). When sorted Sox1-GFP positive and negative cells were re-plated in EpiSC culture conditions, the original ratio of positive and negative cells was not restored, but a degree of interconversion was observed (Tsakiridis *et al.*, 2014). Collectively, these data suggested that the Sox1-GFP$^+$ EpiSC subpopulation is poised to neural differentiation and possibly contains cells that have already progressed into neural commitment (Tsakiridis *et al.*, 2014). Crucially, *Sox1* expression has also been detected in *in vitro* NMP cultures generated from mES cells or EpiSCs (Gouti *et al.*, 2014; Turner *et al.*, 2014). However, it is not known whether under these conditions *Sox1* is expressed in bipotent NM progenitors or marks the emergence of neural committed cells.

As discussed in section 1.3.3, co-expression of the mesoderm marker *Tbra* and the neural/pluripotency marker *Sox2* seems to overlap extensively with NMP identity (Wymeersch *et al.*, 2016). However, there are no unique markers for the identification of NMPs, but the genes that are activated in the axial progenitor regions are also expressed in different tissues of the mouse embryo. *Sox1* is one of the genes the expression of which expands over the NMP regions (Cambray and Wilson, 2007; unpublished data). However, Sox1 is particularly interesting because of two main

differences with Sox2. First of all, Sox1 is an early marker of neural specification exclusively, in contrast to Sox2 that also functions as a core pluripotency factor (Masui *et al.*, 2007; Gagliardi *et al.*, 2013). For instance, we have shown that EpiSC cultures contain a major fraction of Tbra/Sox2 double positive cells, which are pluripotent rather than bipotent NMPs (Tsakiridis *et al.*, 2014; chapter 3). In contrast to that, under self-renewing conditions *Tbra* and *Sox1* expression exhibit an almost mutually exclusive pattern, marking cells that are poised to mesoderm and neural differentiation, respectively (Tsakiridis *et al.*, 2014). Hence, the emergence of Tbra/Sox1 double positive cells could potentially be a more specific indicator of NMP identity, at least *in vitro*. Second, expression of *Sox1* rises later on compared to *Sox2* (Wood and Episkopou, 1999; Cajal *et al.*, 2012) and importantly, it is upregulated in the NMP regions during their transition from the primitive streak to tail bud stages (Cambray and Wilson, 2007; figure 4.1). Hence, it was hypothesized that Sox1 could be used as marker of late NMPs.

The expression profile of Sox1-GFP$^+$ cells found in *in vitro* NMP cultures matches the gene activity observed in the axial progenitor regions *in vivo*

NMPs can be generated *in vitro* upon treatment of EpiSCs with FGF/CHI for 48-72h (Gouti *et al.*, 2014). In the present study, it was shown that Sox1-GFP levels rise as EpiSCs differentiate into NMPs, with the percentage of GFP$^+$ cells reaching up to 60% in 72h NMP cultures. Crucially, immunofluorescence analysis revealed that under these conditions, Tbra$^+$Sox2$^+$Sox1$^+$ triple positive cells emerge. In contrast to EpiSC cultures, *Tbra* and *Sox1-GFP* expression exhibited overlap in a major part of the NMP cultures, with Sox1-GFP protein being detected mainly in cells with low levels of Tbra. Expression analysis in FACS sorted GFP positive and negative subpopulations demonstrated that Sox1-GFP$^+$ cells express late primitive streak/NMP markers at the same levels as the Sox1-GFP negative fraction, while *Tbra* RNA was shown to be slightly enriched in the Sox1-GFP$^-$ cells. Analysis of neural markers demonstrated that *Gbx2*, that is found at the posterior neural plate including the NMP regions, was the only gene exhibiting upregulation in the Sox1-GFP$^+$ fractions. Collectively, these data

demonstrate that the Sox1-GFP$^+$ subpopulation that emerges in FGF/CHI-treated EpiSC cultures resembles the profile of NMPs. Single cell immunofluorescence image analysis demonstrated that Tbra is not detected in all Sox1-GFP cells, but Sox2/Sox1-GFP double positive and Sox1-GFP single positive cells are present in NMP cultures too. This expression pattern, in combination with the upregulation of *Gbx2* in the Sox1-GFP$^+$ cells, implies that the Sox1-GFP$^+$ fraction may also include neural committed cells.

The properties of Sox1-GFP populations found in *in vitro* NMP cultures have also been described by Turner *et al.* (2014). However, in that study, *in vitro* NMPs were generated through Wnt and Fgf signalling mediated differentiation of mES cells and not EpiSCs. More specifically, Sox1-GFP mES cells were cultured in N2B27 for 5 days, with a 24h pulse between days 2 and 3 of either CHI and FGF or CHI only. Expression analysis of Sox1-GFP negative, low and high subpopulations revealed that *Tbra* and *Nkx1.2* were expressed in the negative and low fractions, but they were absent in the GFP^{high} cells. Importantly, single cells analysis of the Sox1-GFP negative and low fractions confirmed the presence of Sox1$^+$Tbra$^+$ and Sox1$^+$Nkx1.2$^+$ double positive cells and similar to our observations, it revealed that high *Tbra* RNA levels relate to relatively lower levels of *Sox1*. The *Tbra/Sox1* expression pattern observed in *in vitro* NMP cultures resembles the results of a recent *in vivo* study demonstrating that high expression levels of *Tbra* and *Sox2* were detected in regions fated for mesodermal and neural differentiation respectively, while Tbra$^+$Sox2$^+$ *in vivo* NMPs were characterized by low-to-medium expression levels of both genes (Wymeersch *et al.*, 2016).

In the future the generation of a *Tbra/Sox1* double reporter line will shed light on the differences between the Tbra, Sox1 double or single positive subpopulations and importantly, it will allow us to isolate a more homogeneous population of NMP-like cells. As described in chapter 5, a new Tbra-GFP mES cell reporter line has already been made, while the targeting construct that was used for the generation of the Sox1-GFP cell line (Aubert *et al.*, 2003) has been modified and the GFP gene has been

replaced by the sequence of tdTomato (data not shown). In this way, a Sox1-tdTomato/Tbra-GFP double reporter line can be generated through targeting of the *Sox1* locus in the Tbra-GFP mES cells.

Sox1-GFP$^+$ cells exhibit the gene expression profile of late NMPs

In vivo, axial progenitors are exposed to Wnt and Fgf signals (Wilson *et al.*, 2009). These signalling pathways are necessary for axis elongation and play a key role in the rostrocaudal patterning of *Hox* gene expression, specifying in this way, the regional identity of the newly formed tissues (Liu *et al.*, 2001; Dasen *et al.*, 2003; Nordstrom *et al.*, 2006; Mazzoni *et al.*, 2013). Microarray analysis of dissected caudal progenitor regions demonstrated that the *Hox* code undergoes profound transcriptional changes as NMPs transit from E8.5 primitive streak to E9.5 tail bud stages (unpublished data, see 1.3.3). Interestingly, this pattern coincides with the *Sox1* upregulation in NMP regions (Cambray and Wilson, 2007; figure 4.1). As it would be expected, upregulation of *Hox* genes has been observed in the FGF/CHI-induced *in vitro* NMP cultures (Gouti *et al.*, 2014; Turner *et al.*, 2014). Hence, based on all the above, it was predicted that if the hypothesis of *Sox1* expression marking late NMPs is correct, then *Hox* genes should be differentially expressed in the Sox1-GFP positive and negative fractions. Expression analysis showed that genes from the paralogous groups 4-10 were upregulated in the Sox1-GFP$^+$ cells, recapitulating the burst of activity observed *in vivo*, in NMP regions, between E8.5 and E10.5 and in this way, supporting the notion that Sox1-GFP cells constitute the *in vitro* equivalent of the tail bud stage NMPs. Interestingly, genes from the groups 11 and 13 exhibited the lowest induction in NMP cultures and no differences in the two Sox1-GFP fractions. This can be explained by the 'temporal co-linearity' of *Hox* gene code (Deschamps and van Nes, 2005). More specifically, since 3' *Hox* genes are expressed first, whereas more 5' Hox genes are expressed later and sequentially, within the 72 hours of exposure to FGF/CHI, the most posterior *Hox* genes "spent less time being transcriptionally active, resulting in lower induction levels; it is possible that the short period of transcriptional activation does not allow any differences between the two subpopulations to reveal.

The single cell immunofluorescence image analysis and the QRT-PCR data in the sorted Sox1-GFP fractions highlights the heterogeneity of *in vitro* NMP cultures and reveals the presence of a NMP-like subpopulation that seems to undergo maturation, developing a more posterior character. What is the reason behind this non-synchronous maturation of the population is not known. As presented in chapter 3 and in a number of published studies (Hayashi and Surani, 2009; Han *et al.*, 2010; Tsakiridis *et al.*, 2014), EpiSC cultures consist of distinct subpopulations with different characteristics. Hence, it is possible that the non-synchronous maturation of the NMP-like cells is the result of the heterogeneity in the starting EpiSC cultures.

In the future, it will be very interesting to test the differences between the Sox1-GFP positive and negative fractions functionally. More specifically, sorted subpopulations could be cultured under neural or paraxial mesoderm inducing conditions and the differentiation products could be assessed for the expression of posterior *Hox* genes. If the hypothesis that Sox1-GFP$^+$ cells represent mature progenitors is correct, then the derivatives of the Sox1-GFP expressing fraction should exhibit a more caudal character, being enriched in 5' *Hox* genes. In that case, the identification of Sox1-GFP as a marker of late *in vitro* NMPs could contribute to the development of refined differentiation protocols that aim to generate skeletal muscles or spinal cord neurons of lower axial levels.

Sox1-GFP+ cells undergo neural differentiation under prolonged culture in FGF/CHI

In vivo, the existence of bipotent NM progenitors has been confirmed through genetic marking of single cells and their derivatives using the *LaacZ* system in mouse embryos (Tzouanacou *et al.*, 2009). *In vitro*, the bipotent nature of EpiSC-derived NMPs have been tested through prolonged exposure of cells to FGF/CHI and analysis of the resulting clones (Tsakiridis and Wilson, 2015). More specifically, sorted Tbra-GFP$^+$

cells of 72h NMP cultures were replated at clonal density in FGF/CHI for an extra 48h and the colonies that emerged were analysed for expression of *Tbra* and *Sox2*. Around 55% of the resulting colonies were $Sox2^+$, while $Tbra^+$ clones accounted for 9% of the culture, suggesting that under these conditions cells undergo mainly unilinear differentiation into either neurectoderm or mesoderm, exhibiting strong bias towards the former. Importantly, 12% of the colonies contained a mix of Tbra and Sox2 single positive cells, confirming the bipotent status of Tbra-GFP^+ *in vitro* NMPs.

In the present study, following replating at clonal density and prolonged culture (5 extra days) in FGF/CHI, sorted Sox1-GFP negative cells were shown to be able to give rise to mesoderm-like and neural-like colonies, but also to clones containing a mix of differentiation products, reproducing the results of the abovementioned study. However, sorted Sox1-GFP^+ cells revealed strong neurogenic capacity, but inability to generate $Tbra^+$ colonies. In fact, *Tbra* expression was detected in 2 out of 86 colonies, but given that the sorted fractions contained 1-4% contaminating Sox1-GFP^- cells, it is possible that these colonies have resulted from the GFP negative cells. Hence, it appears that although the Sox1-GFP positive and negative fractions exhibit the same expression profile, their fate under prolonged exposure to FGF/CHI differs. Interestingly, large fractions of $Tbra^+Sox1^+Sox2^+$ and $Tbra^+Sox2^+$ colonies, that most likely represent self-renewing NMPs, were detected in only one of the two replicates of the experiment, whereas only one $Tbra^+Sox2^+$ colony that was observed in the second replicate. This could possibly be due to differences in the gating applied in each FACS sorting (appendix, figure S4.2, A). More specifically, the first time the experiment was performed, the sorted fraction contained cells with intermediate and high levels of GFP; in the second replicate, the 72h NMP cultures contained two clearly distinct populations and the sorted fraction composed of cells with relatively high GFP levels. Hence, it is likely that the variable Sox1-GFP levels represent cells with distinct properties. This is consistent with the study of Turner *et al.* (2014), demonstrating that the NMP status coincides with low expression levels of *Sox1*. Overall, the results of this experiment suggests that Sox1-GFP cells are biased towards neural lineage specification, while the question whether they can undergo mesoderm differentiation remains unanswered. In order to address this issue, sorted Sox1-GFP

subpopulations could be cultured in medium containing CHI, but lacking Fgf2. As discussed in 1.3.4, canonical Wnt signalling has a key role in mesoderm formation (Yamaguchi *et al.*, 1999) and importantly, it has been shown that exposure of NMPs (mES cell-derived or human NMPs) to CHI induces paraxial mesoderm identity (Gouti *et al.*, 2014; Turner *et al.*, 2014). Hence, if Sox1-GFP positive cells are not committed to neural differentiation, but retain the ability to form mesoderm derivatives, then paraxial mesoderm-like ($Tbx6^+$) colonies should emerge from the Sox1-GFP^+ fractions. Importantly, it has been reported that under prolonged exposure to FGF/CHI, the capacity of cells to undergo paraxial mesoderm differentiation (as demonstrated by expression of Tbx6) is affected by the cell density (Tsakiridis and Wilson, 2015). Hence, in the future, it will be necessary to test a range of different cell densities, when paraxial mesoderm differentiation is induced.

The developmental potential of Sox1-GFP^+ cells could also be tested by transplantation into mouse embryos. Previous studies have shown that upon grafting in the NSB of E8.5 embryos, *in vitro* NMPs contribute to paraxial mesoderm and neural tube (Gouti *et al.*, 2014). Therefore, if Sox1-GFP^+ cells are *bona fide* NM bipotent progenitors, then the signalling environment in the NMP region will allow the grafted cells to incorporate into both neural and mesodermal tissues. However, grafting of cells collected immediately after cell sorting has proven to be technically challenging. In order to overcome this problem, FACS sorting could be performed in 48h NMP cultures and the sorted positive and negative fractions could be replated in FGF/CHI for an extra 24h, before transplantation into mouse embryos. Although this protocol ensures that cells are not exposed to FGF/CHI for more than 72h, it involves the possibility of FACS sorting and replating affecting the expression profile and maybe the properties of the cells. Hence, characterization of the 72h cultures will be necessary before we proceed to transplantations.

Does *Sox1* expression mark the same cell type in 48h and 72h NMP cultures?

As mentioned before, NMPs can be generated upon FGF/CHI treatment for 48h and 72h, while Sox1-GFP expression is detected in both time points (figure 4.2). In this study, I have investigated the properties of Sox1-GFP$^+$ cells in 72h NMP cultures. Could the properties of Sox1-GFP cells have changed as the culture expands over the 48 to 72h? The study of Gouti *et al.* (2014) indicate that expression of *Hox* genes increases as EpiSCs and hES cells are exposed to FGF/CHI, suggesting progressive maturation of *in vitro* NMP cultures. The *in vitro* clonal analysis presented by Tsakiridis and Wilson (2015) demonstrates that single cells isolated from 48h and 72h NMP cultures exhibit the same capacity to form colonies containing a mix of mesoderm-like and neural-like cells, suggesting that at 72h, cells remain in the bipotent NM state. It has been previously shown that the developmental potential of *in vitro* NMPs can be accessed *in vivo* through cell transplantation in the NSB of the E8.5 mouse embryo. So far, only 48h NMP cultures have been tested *in vivo* (Gouti *et al.*, 2014). However, it is important to confirm the NMP-like nature of 72h cultures *in vivo* too, as if the developmental potential of NMP cultures changes over 48h to 72h, then it is possible that the character of the Sox1-GFP$^+$ cells is altered as well.

In order to address this issue, we have started performing grafting of 48h and 72h NMPs in E8.5 embryos. The results of the first transplantations are presented in supplementary figure S4.3 (see appendix). The grafting experiment and the *ex-vivo* culture was performed by Yali Huang. In summary, transplantation of 48h NMPs in the NSB of E8.5 embryos resulted to incorporation of the labelled cells in 4 out of the 5 embryos. Sections from these embryos revealed integration of transplanted cells into the paraxial mesoderm and the neural tube of the posterior trunk, while contribution to endoderm was not observed. Moreover, labelled cells exhibited integration into the tail bud (3/4 embryos). Overall, the contribution of 48h grafted cells resembles the integration pattern that transplanted *in vivo* NMPs have exhibited (Cambray and Wilson, 2002, 2007; Wymeersch *et al.*, 2016). Grafted 72h NMP cultures exhibited

integration in 1 out of 5 host embryos, which exhibited aberrant morphology. In this embryo, contribution was expanded as caudally as the tail bud, with labelled cells being detected in paraxial mesoderm and the neural tube. Surprisingly, contribution to the formation of neural crest cells was detected too. The integration of 72h NMPs in only 1/5 embryos, as compared to incorporation of 48h NMPs in 4/5 embryos, raises the question whether there is any difference in the status of the progenitor cells at the two time points. At the moment, the small number of grafted embryos does not allow us to draw any conclusions, but more grafting experiments need to be performed in the future. Based on the possibility that the developmental potential of 48h and 72h NMP cultures may be different, it will be very interesting to investigate the expression profile and most importantly, the differentiation potential of the Sox1-GFP^{+} cells that emerge in the 48h NMP cultures.

In the next chapter, I present the generation of two reporter mouse ES cell lines that in the future will help us to perform more refined studies of the distinct subpopulations that are found in the *in vitro* NMP cultures.

Chapter 5:
Generation of a Tbra-GFP/Sox2-tdT double reporter cell line

5.1. Introduction

As NMPs are marked by the coexpression of *Tbra* and *Sox2* (Wymeersch *et al.*, 2016), a *Tbra/Sox2* double reporter line would be a valuable tool that would allow us to track and isolate these progenitor cells *in vivo* and *in vitro.* A Tbra-GFP reporter ES cell line has already been described in the past (Fehling *et al.*, 2003); however, in this cell line, part of the first exon of one of the *Tbra* alleles was replaced by a GFP cassette carrying a polyadenylation signal and as a result transcription of the targeted *Tbra* allele was blocked. Since *Tbra* heterozygous mice present a defect in migration of mesoderm away from the primitive streak and are characterised by truncated and often slightly kinked tails (Beddington *et al.*, 1992; Wilson *et al.*, 1995), this reporter cell line is not suitable for axis elongation studies, but one that retains the integrity of both *Tbra* alleles is required.

The generation of a *Sox2* reporter mouse ES cell line has been reported in a few studies so far (Avilion *et al.*, 2003; Ellis *et al.*, 2004; Taranova *et al.*, 2006). However, these reporter lines are heterozygous for a null allele of *Sox2*, while the targeting vector has replaced part or all of the 3' untranslated region (UTR), a region that could be important for transcriptional regulation. In our lab, we have previously targeted the *Sox2* locus, replacing the stop codon with a histone H2B-tandem dimer Tomato (H2B-tdTomato) cassette (Frederick Wong, unpublished data.). *Sox2* reporter cells faithfully recapitulated the endogenous *Sox2* expression and contributed to the development of chimeric embryos; however, chimeras exhibited neural tube defects and abnormal development of the genitalia.

Based on all the above, the aim of this part of the project was to generate a reliable *Tbra/Sox2* double reporter line that would minimally disrupt the endogenous loci to produce a physiologically normal transgenic mouse line.

5.2. Targeting of *Tbra* locus using CRISPR/Cas9 technology

The targeting construct of the *Tbra* locus was generated by Frederick Wong (figure 5.1). The targeting vector was designed to replace the stop codon with a cassette consisting of a glycine-serine-glycine spacer- Thosea asigna virus 2A peptide sequence (GSG-T2A-)linked (Szymczak *et al.*, 2004; Wang *et al.*, 2015) H2B-GFP ORF (Kanda *et al.*, 1998), which is then joined, via an internal ribosome entry site (IRES), to the gene encoding for puromycin N-acetyltransferase (Pac) [Tbra::GSG-T2A::H2B-GFP::IRES::Pac]. This sequence was followed by a flippase recombinase recognition target (FRT-)flanked cassette containing a neomycin/geneticin (G418) resistance gene driven by a merged eukaryotic phosphoglycerate kinase 1 (PGK) and bacterial EM7 promoter, as well as an expression cassette for thymidine kinase from herpes simplex virus (HSV-TK) [FRT::PGK/EM7-Neo::MC1-TK::FRT]. The targeting vector was linearised and transfected into E14tg2a ES cells by electroporation. G418-resistant clones were picked and expanded, and correct gene targeting was confirmed by southern blot. In this experiment, ninety clones were screened, but none of them presented successful targeting (data not shown), suggesting very low targeting efficiency for this locus.

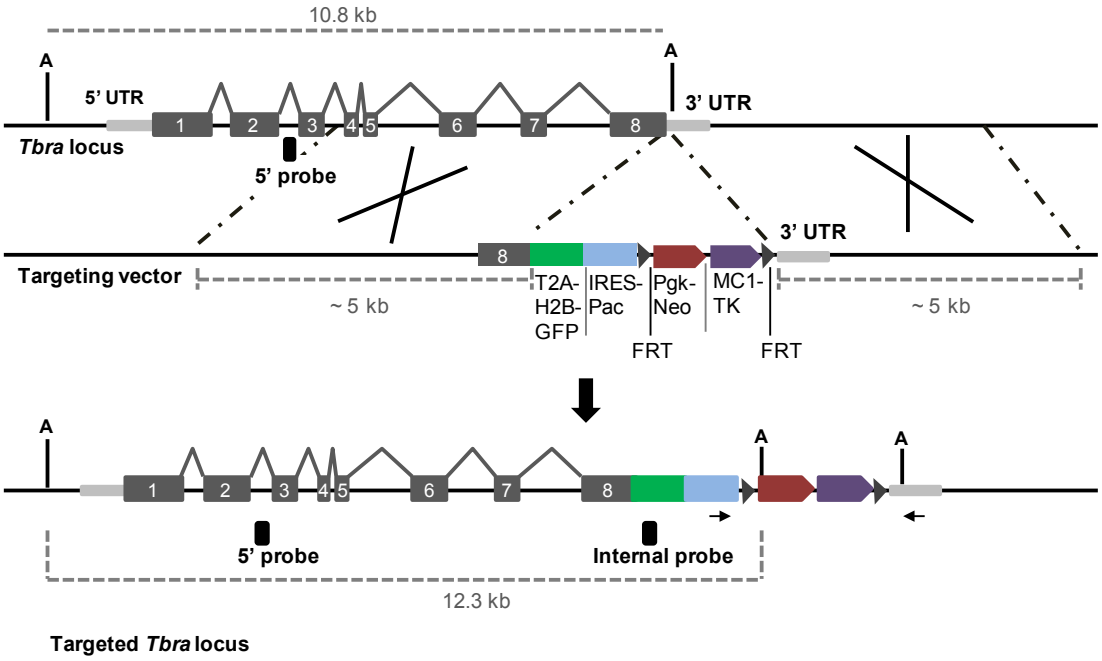


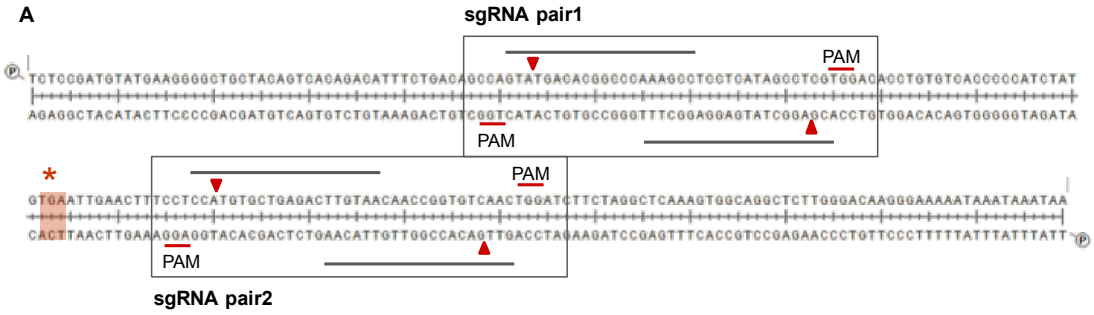
Figure 5.1: *Tbra* locus targeting strategy. Schematic illustration of the endogenous *Tbra* locus, the targeting construct and the insertion of the reporter cassette downstream of the *Tbra* coding region through homologous recombination. The numbered grey boxes represent exons. The dark dotted lines indicate the homology region between the endogenous locus and the targeting vector. "A" denotes the cutting site of the restriction enzyme AseI that was used for southern blot analysis. Digestion of wild-type and targeted loci resulted in gDNA fragments of 10.8kb and 12.3kb, respectively. The binding sites of the 5' probe and the internal probe used for the southern blot analysis are shown. The arrows next to the diagram of the targeted locus indicate the binding sites of the primers used to verify the excision of the FRT-flanked cassette.

In order to increase the targeting efficiency of *Tbra* locus, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR-associated protein-9 nuclease (Cas9) technology was employed (Cong *et al.*, 2013). Wild-type Cas9 nuclease promotes genome editing by creating double strand breaks, at a target genomic locus. The nuclease is driven to the target site by a sequence of just 20bp, the single guide RNA (sgRNA). The complementarity between the sgRNA and the target DNA tolerates mismatches, therefore the system is prone to generation of unwanted off-target mutations. To achieve higher specificity, without reducing the efficiency levels, the aspartate-to-alanine (D10A) mutant form of Cas9 was used, that is only capable of cutting one of the DNA strands, creating a nick instead of a double strand break (Cas9nickase - Cas9n) (Ran *et al.*, 2013). More specifically, this strategy includes the use of a pair of sgRNAs that are complementary to opposite strands of the target site. As a result, at the target site, nicks are generated on both DNA strands and result in a staggered double strand break. In the presence of a template, such as a targeting vector, the DNA damage is repaired through the pathway of homology directed repair. Alternatively, the ends of a double strand break are rejoined through non-homologous end joining, which can result in random indel mutations at the site of junction. Off-target binding of the sgRNAs still occurs, but the binding of both of the paired sgRNAs close enough together, on opposite strands, to generate a double-strand break is highly unlikely. Thus, off-target binding is most likely to generate individual nicks that can be generated by each of the sgRNAs. These are predominantly repaired by the high fidelity base excision repair pathway.

In this study, two candidate pairs of sgRNAs were designed (section 2.5.13) to bind around the stop codon of *Tbra*, where the GFP-cassette had to be inserted (figure 5.2, A). As shown in figure 5.2, B, both pairs exhibited high target specificity and none of them was predicted to create a double strand break at an off-target site. Each sgRNA was cloned into the pSpCas9n-2A-GFP (PX461) vector (chapter 2, figure 2.2) and the two pairs were tested functionally. As described in section 2.5.13, the ability of the two candidate pairs to induce a double strand break at the *Tbra* gene was assessed by

lipofecting E14tg2a ES cells with each pair of cloned sgRNAs and then analysing the gDNA sequence of the target sites. As explained above, in the absence of a repair template, a double strand break should result in insertions or deletions close to the target site. The generation of indels was assessed by performing High Resolution Melting (HRM) assay (Vossen *et al.*, 2009), comparing the melting curves of the two transfected samples with that from an non-transfected control. This showed that the gDNA sequence of the sample transfected with the sgRNA pair2 was clustered with the sequence of the wild-type control, whereas the gDNA sequence of the cell sample that was transfected with sgRNA pair1 stood out as a distinct sequence (figure 5.2, C). This finding indicated that Cas9n when driven by sgRNAs pair1 can create a double strand break upstream of the stop codon of the *Tba* locus.

A



B

sgRNA		Sequence	Off target sites (within genes)	Off target sites per pair	Pair score (%)
Pair 1	A	GGCTTTGGGCCGTGTCATAC	69 (9)	0	75
	B	AAAGCCTCCTCATAGCCTCG	134 (23)		
Pair 2	A	TTACAAGTCTCAGCACATGG	251 (14)	0	59
	B	TTGTAACAACCGGTGTCAAC	35 (2)		

C

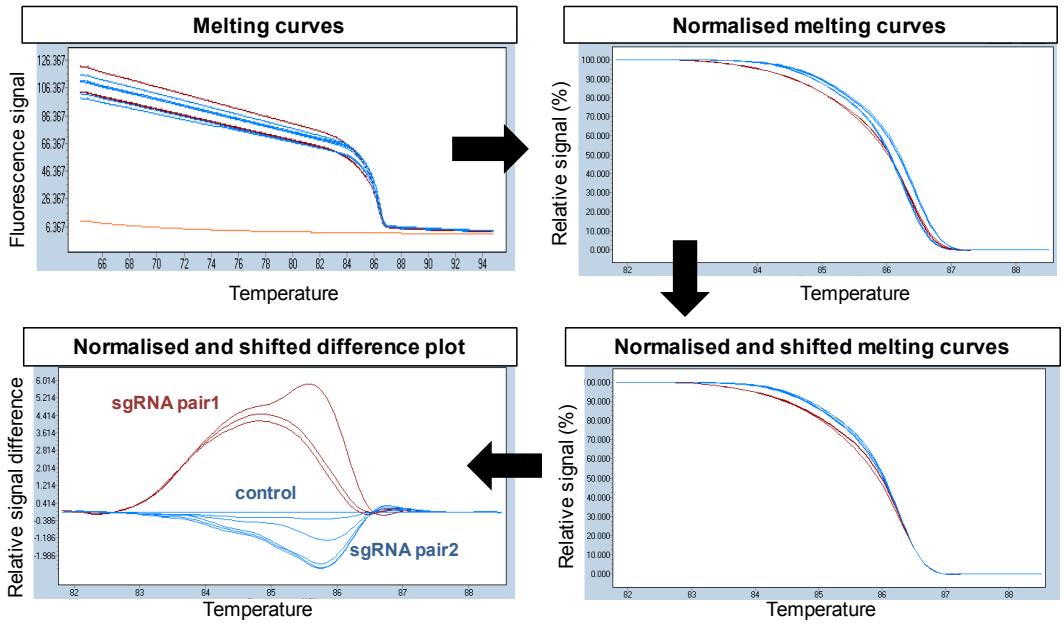


Figure 5.2: Design and functional testing of CRISPR/Cas9n sgRNAs. A) sgRNAs were designed using a free online software package as described in section 2.5.13. The top diagram presents the DNA sequence of *Tbra*, including the end of the last coding exon and the beginning of 3' UTR. The red asterisk highlights the location of the stop codon. The binding sites of two pairs of sgRNAs and the PAM sequences are marked

on the DNA strands with grey and red lines respectively. The red arrow heads indicate the sites where the nicks will occur. **B)** The table shows the sequence of each sgRNA, the number of off-target sites and the total score of each pair, as predicted by the designing software. **C)** High Resolution Melting assay was performed to compare the gDNA sequence at the CRISPR/Cas9n target sites of ES cells transfected with each pair of sgRNAs with the sequence of non-transfected ES cells (each reaction was performed in triplicate). The raw melting curves were normalised by setting pre-melt (initial fluorescence) and post-melt (final fluorescence) signals of all samples to uniform values (% percentage). Then the normalised curves were shifted along the temperature axis, to equalize the point at which the double strand DNA in each sample becomes completely denatured. The normalised and shifted melting curves were subtracted from the curve of the control sample, to get a clearer display of the differences in melting curve shape.

To generate the reporter cell line, E14tg2a ES cells were nucleofected with the targeting vector we originally used (figure 5.1) and the two Cas9n expressing plasmids, each carrying a single sgRNA from pair 1 (Ran *et al.*, 2013). Ninety-six G418-resistant colonies were picked and expanded. gDNA was extracted from each clone and correct targeting was assessed by southern blot. Of these 96, 45% of the clones had integrated the GFP cassette in one of the two alleles, 3% of the clones were homozygous for the integration and 15% presented non-targeted integration events (figure 5.3, A).

As explained above, the target site of Cas9n is just upstream of the stop codon (figure 5.2, A). Importantly, this region is included in the homology arms of the targeting construct (figure 5.1), therefore it is possible that after homologous recombination took place, the Cas9n could still target either of the two alleles, creating indels at the *Tbra* locus. To screen for undesired mutations, gDNA was extracted from 28 correctly targeted clones and the sequence of both *Tbra* alleles was examined, performing HRM assay and sequencing. 47% of the targeted clones did not carry any mutations in the *Tbra* ORF, 14% of the clones had mutations in both alleles, 25% presented mutations in the targeted allele and 14% carried mutations in the non-targeted allele. Seven correctly targeted clones carrying the wild type *Tbra* sequence in both alleles were selected for further analysis. To verify that there was only one copy of the targeting vector integrated into the genome, southern blot was performed using a probe complementary to the GFP cassette (figure 5.1 and 5.3, B). Karyotype analysis of the correctly targeted was subsequently performed which showed clones 34, 77 and 78 had a modal chromosome count of 40 (figure 5.3, C). Clones 34 and 78 presented the highest frequency of cells containing 40 chromosomes, therefore they were selected for the cell line generation.

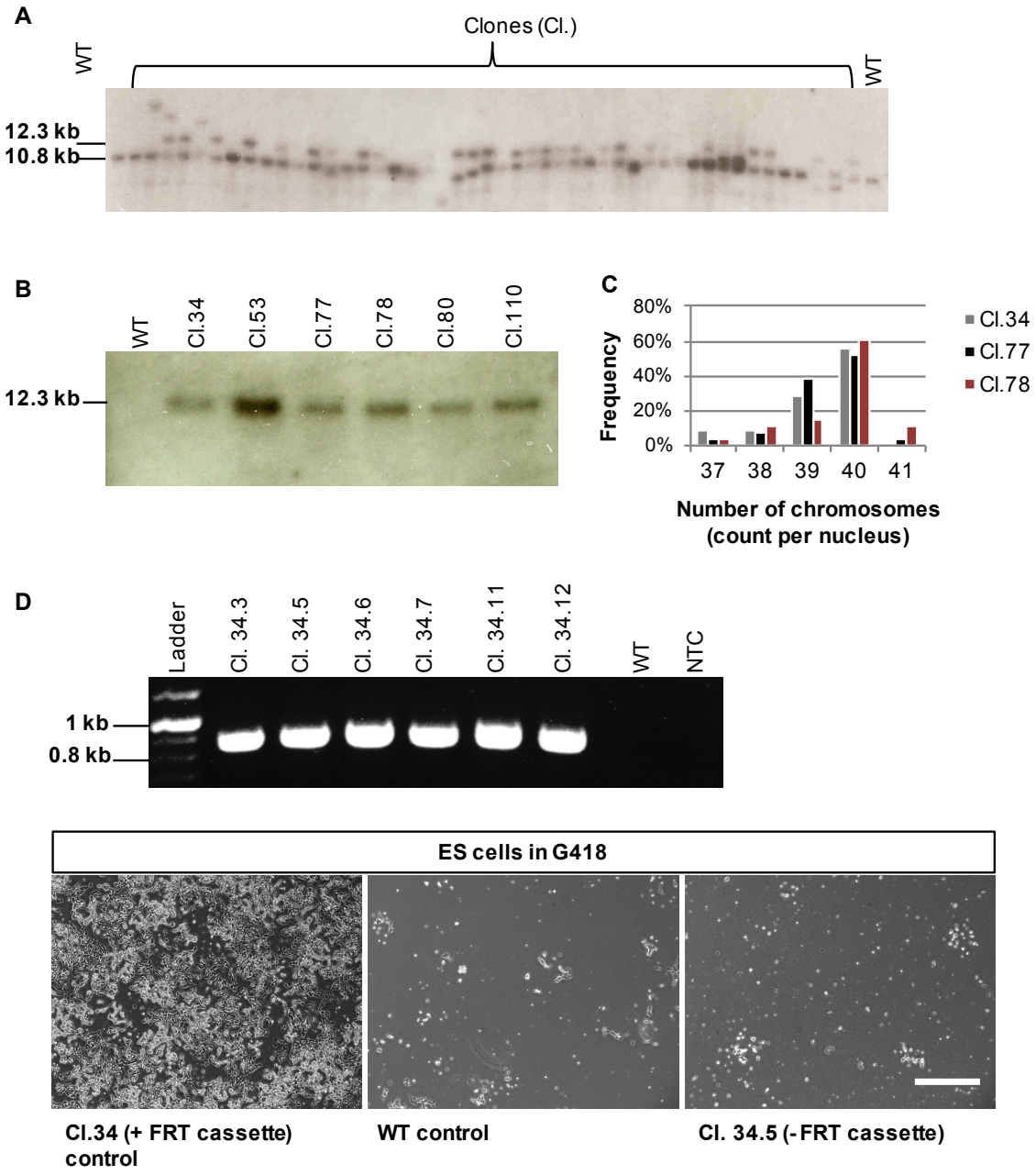


Figure 5.3: Generation of a Tbra-GFP-Pac mouse ES cell reporter line. Southern blot using the 5' probe, for the identification of clones with correct gene targeting. DNA fragments of wild-type (WT) clones correspond to 10.8kb, whereas targeted clones present bands of 12.3kb. **B)** Southern blot using the internal probe, for the identification of clones with a single integration event**. C)** Karyotypic analysis of correctly targeted clones. **D)** PCR for the confirmation of successful FRT cassette

excision. The binding sites of the primers used for this analysis are shown in figure 5.1. When FRT cassette has been excised, PCR gives rise to a product of 840 bp. NTC denotes non-template control. FRT cassette removal was further confirmed by incubating clones in G418 (lower panel). Scale bar: 600μm.

It has been shown that expression of HSV-TK can cause sterility in male transgenic mice (Alshawi *et al.*, 1988; Braun *et al.*, 1990). In order to be able to generate a mouse line that is able to transmit the GFP cassette through the germ line, and also to avoid any effects that the bidirectional activity of the PGK promoter (Johnson and Friedmann, 1990) could have on the endogenous *Tbra* locus, the cassette [FRT::PGK/EM7-Neo::MC1-TK::FRT] was removed from the targeted clones. As explained in section 2.1.11 in more detail, clone 34 and 78 were transfected with a Flp recombinase-expressing vector and following negative selection with ganciclovir, colonies were picked and expanded. Clone 78 failed to give rise to growing colonies, therefore the study was carried out working with clone 34. In order to confirm the TK cassette excision, gDNA was extracted from the clones and PCR was performed (figure 5.3, D). To confirm functional loss of the neomycin resistance gene, the clones were plated in ES cell medium in the presence of G418. In these conditions, the clones failed to grow (figure 5.3, D). The correctly excised clone 34.5 was selected for subsequent experiments.

5.3. Tbra-GFP-Pac ES cell and EpiSC line validation.

To assess the correlation between the expression of endogenous Tbra and GFP, 34.5 Tbra reporter ES cells were cultured for 2 days in ES cell medium containing 3μM CHI, in the absence of Lif. Under these conditions that promote *Tbra* expression and mesoderm differentiation (Yamaguchi *et al.*, 1999; Gadue *et al.*, 2006), GFP was induced and immunocytochemistry showed that GFP recapitulates the expression pattern of *Tbra* (figure 5.4, A). Tbra-GFP fluorescence was detectable by microscopy, without immunofluorescence staining being necessary, a property that opens the way

for live imaging microscopy studies in the future. A discussed before, in contrast to ES cells, a major fraction of EpiSC cultures (30-60%) expresses *Tbra* heterogeneously, under self-renewing conditions. Hence, EpiSCs were differentiated *in vitro* from ES cells (Brons *et al.*, 2007; Tesar *et al.*, 2007) and used for further validation of the reporter line. To investigate the correlation of *Tbra* RNA with the GFP levels, EpiSCs were FACS sorted into a GFP negative, a low GFP and a high GFP fraction and QRT-PCR analysis of *Tbra* in the three subpopulations was performed. This analysis demonstrated that the GFP expression levels reflect the RNA levels of endogenous *Tbra* (figure 5.4, B), proving altogether that this cell line constitutes a reliable reporter. As shown in figure 5.1, the targeted clone includes a puromycin N-acetyltransferase gene, which is under the transcriptional control of the *Tbra* promoter. Titration of puromycin in serum-free EpiSC culture conditions showed that a minimum working concentration of 0.55μg/ml puromycin at a plating cell density of 26000/cm^2 was sufficient to eliminate wild type cells within 5-7 days. When Tbra-GFP-Pac EpiSCs were cultured in these conditions, puromycin resistant colonies emerged and the GFP levels increased to 80% (figure 5.4, C), proving that the GFP-Pac cassette is functional, and therefore the cell line can be used for lineage selection studies.

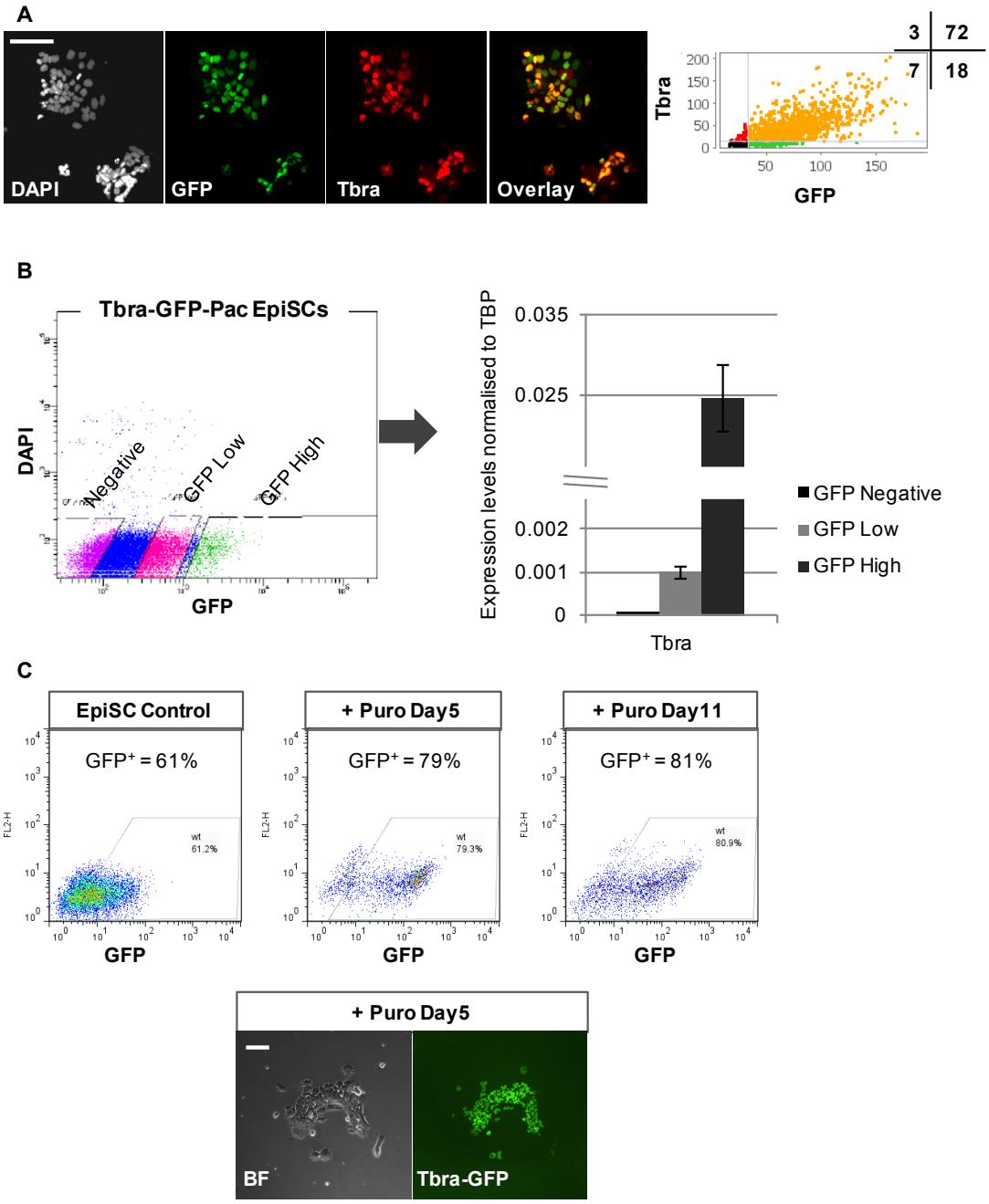


Figure 5.4: Validation of the Tbra-GFP-Pac reporter cell line. Immunocytochemistry for endogenous Tbra and GFP, in ES cells cultured in mesoderm-inducing conditions. The plot was generated as the result of image quantification and shows the correlation between the expression levels of endogenous Tbra and GFP. **B)** Tbra-GFP-Pac EpiSCs were FACS sorted in a $GFP^{negative}$, GFP^{low} and GFP^{high} fractions (plot on the left). RNA was extracted from each subpopulation

and used for QRT-PCR analysis of *Tbra* (plot on the right). N=1. Error bars represent standard deviation of technical triplicates. **C)** FACS analysis of GFP expression in Tbra-GFP-Pac EpiSCs cultured in puromycin for 11 days. The microscopy images on the lower panel show a healthy, GFP positive colony of EpiSCs, surviving in the presence of puromycin for 5 days. Puro, puromycin; BF, bright field. Scale bars: 100μm.

To assess the ability of the Tbra-GFP-Pac line to contribute to embryo development and examine the GFP expression pattern in the mouse embryo, ES cells from clone 34.5 were aggregated with wild-type morulae and embryos were collected at E7.5. Five out of the nine dissected embryos were chimeric. These embryos were morphologically normal and GFP faithfully recapitulated the expression pattern of *Tbra*, being present in the posterior epiblast, along the primitive streak, and in the node (figure 5.5, A and B). To test whether Tbra-GFP cells were compatible with normal development, ES cells were injected into blastocysts and embryos were transferred into pseudopregnant females and carried to term. Out of the 6 mice that were born, 3 exhibited coat colour chimerism, including one male, one female and one hermaphrodite (figure 5.5, C). The chimeric mice were morphologically normal and most importantly, the tails had normal length and did not present any kinks. To test germline transmission, the male chimera was bred with MF1 females, but failed to generate offspring. Therefore, the line was terminated and the procedure will be repeated at a later date.

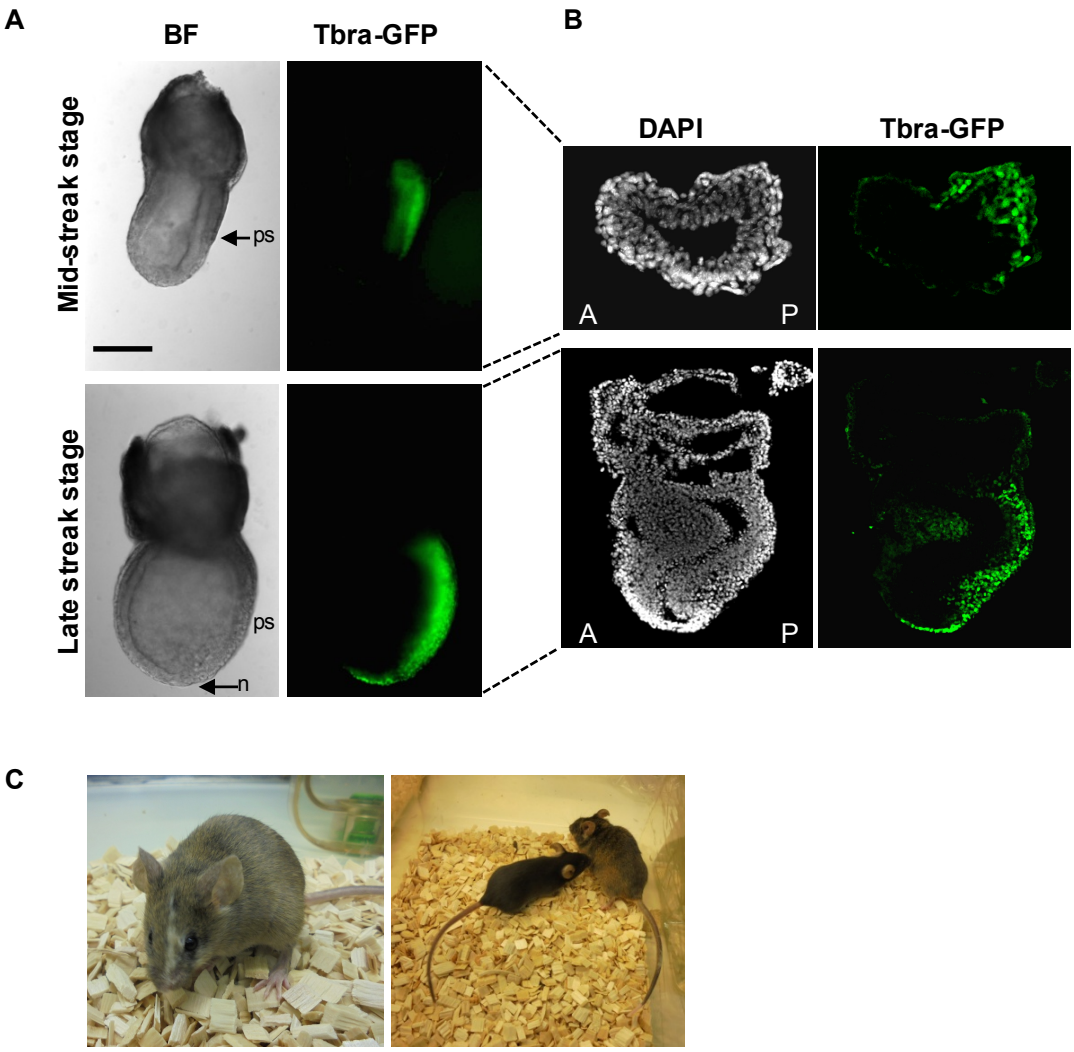


Figure 5.5: Contribution of Tbra-GFP-Pac ES cells to the development of chimeric mice. A) Expression of Tbra-GFP in embryos generated by morula aggregation (dissection and imaging of the embryos was performed with the help of Yali Huang). **B)** Transverse and sagittal sections of the mid-streak and late streak stage embryos shown in A. **C)** Postnatal chimera generated by blastocyst injection. ps, primitive streak; n, node; A, anterior; P, posterior. Scale bar: 200 µm.

5.4. Tbra-GFP expression in *in vitro*-derived NMPs

We have previously shown that when EpiSCs are cultured in NMP-inducing conditions (FGF/CHI), the number of Tbra/Sox2 double positive cells peaks within 48-72h (day 2-3). Thereafter, the colocalisation of the two transcription factors resolves into expression of *Tbra* and *Sox2* in distinct subpopulations in the culture and the number of Tbra positive cells decreases (Gouti *et al.*, 2014; Tsakiridis and Wilson, 2015). In order to quantify the expression levels of *Tbra* under prolonged culture in NMP conditions, Tbra-GFP EpiSCs were cultured in FGF/CHI (see section 2.1.6) for 8 days and GFP expression was monitored by FACS analysis. As shown in figure 5.6, A, on day 2, Tbra-GFP levels increased as compared to EpiSC cultures, with 75% of the population expressing GFP. By day 5, the GFP levels had decreased, with 32% of the culture being GFP positive and by day 8, expression of the fluorescent protein was almost undetectable.

The use of the Tbra-GFP-Pac reporter line allowed us to investigate for the expression profile of the *Tbra*-expressing cells found in NMP cultures. QRT-PCR analysis of sorted GFP-high and GFP-negative subpopulations of 48h NMP cultures (figure 5.6, B) demonstrated that *Tbra* transcript levels were upregulated compared to EpiSCs and as expected, they were much higher in the GFP^+ fraction; however, low levels of *Tbra* mRNA were detected in the GFP negative subpopulation too, suggesting that the later does contain cells with low transcriptional activation of *Tbra*. The expression profile of *Wnt3a* resembled the pattern of its direct transcriptional target, *Tbra,* but *Wnt3a* transcripts were not detected in the GFP negative fraction. Expression levels of *Evx1*, *Nkx1.2* and *Cdx2* were the same in the two fractions. Importantly, expression of *Wnt3a*, *Evx1*, *Nkx1.2* and *Cdx2*, was not detected in the EpiSC control sample. No differences between the two NMP subpopulations were observed in the levels of *Sox2*, whereas *Sox1* was enriched in the GFP negative fractions. This expression profile suggests that the Tbra-GFP^+ fraction contains NMP-like cells, confirming the utility of this cell line for studying NMPs.

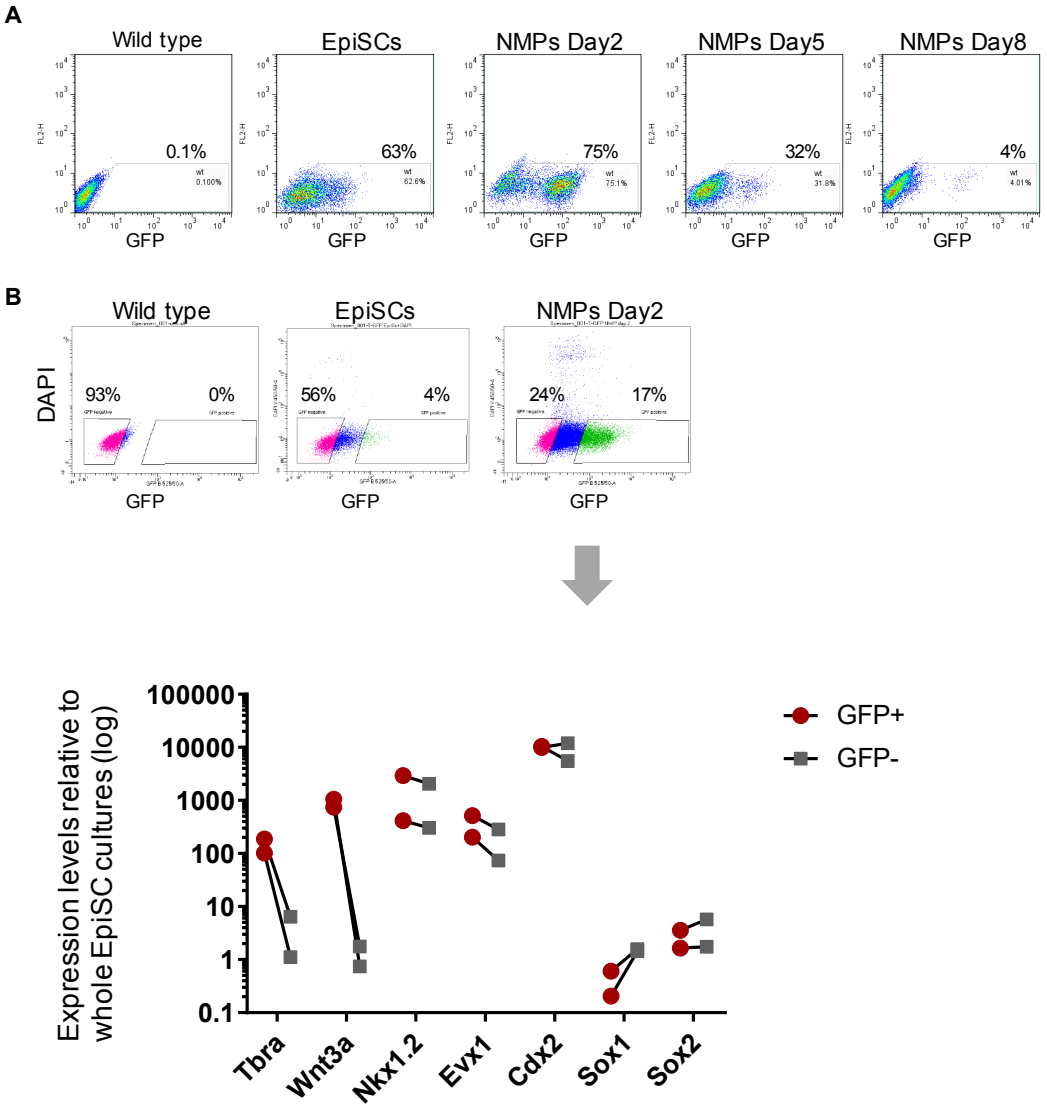


Figure 5.6: Tbra-GFP expression in NMP culture conditions. Tbra-GFP-Pac EpiSCs were cultured in NMP conditions (FGF/CHI) for 8 days and the GFP expression pattern was analysed by FACS. N=3. **B)** FACS sort of Tbra-GFP positive and negative cells of 48h NMP cultures and QRT-PCR analysis of NMP markers in the sorted fractions. N=2.

5.5. Targeting of *Sox2* locus

As mentioned above, we have previously generated a Sox2::T2A::H2B-tdTomato reporter mouse ES cell line (Frederick Wong, unpublished data), but contribution of the transgenic cells to embryonic development resulted to abnormalities in the formation of neural tissues and genitalia. Based on studies reporting that fusion of fluorescent proteins to H2B increases their tendency for oligomerization and therefore it can have cytotoxic effects (Olenych *et al.*, 2007), we speculated that this phenotype could be caused by the fusion of tdTomato to H2B. Therefore, a new version of the vector was made, by Dr. Frederick Wong, containing the tdTomato ORF followed by three repeats of a nuclear localisation signal (tdT-NLS). More specifically, the targeting vector was designed to replace the stop codon of *Sox2* with a cassette consisting of a GSG-T2A::tdT-NLS coding region, which was joined, via an IRES, to a blasticidin-resistance gene (Blasto) [Sox2::GSG-T2A::tdT-NLS::IRES::Blasto]. This sequence was followed by a FRT-flanked cassette containing a G418 resistance gene driven by a merged eukaryotic PGK and bacterial EM7 promoter, as well as an expression cassette for HSV-TK [FRT::PGK/EM7-Neo::MC1-TK::FRT] (figure 5.7, A).

For the generation of the double reporter line, the targeting vector was linearised with the restriction enzyme BstZ17I and transfected into Tbra-GFP-Pac reporter ES cells by elecroporation. G418-resistant clones were picked and expanded, and correct gene targeting was confirmed by southern blot. 22 out of 45 clones (49%) exhibited correct integration of the cassette (figure 5.7, B). To verify that there was only one copy of the targeting vector integrated into the genome, southern blot was performed using a probe that binds to the reporter cassette (figure 5.7, C). The karyotype of some of the correctly targeted clones was assessed; as presented in figure 5.7, D clones 25, 28 and 36 were characterised by a modal chromosome count of 40, therefore they were selected for further validation.

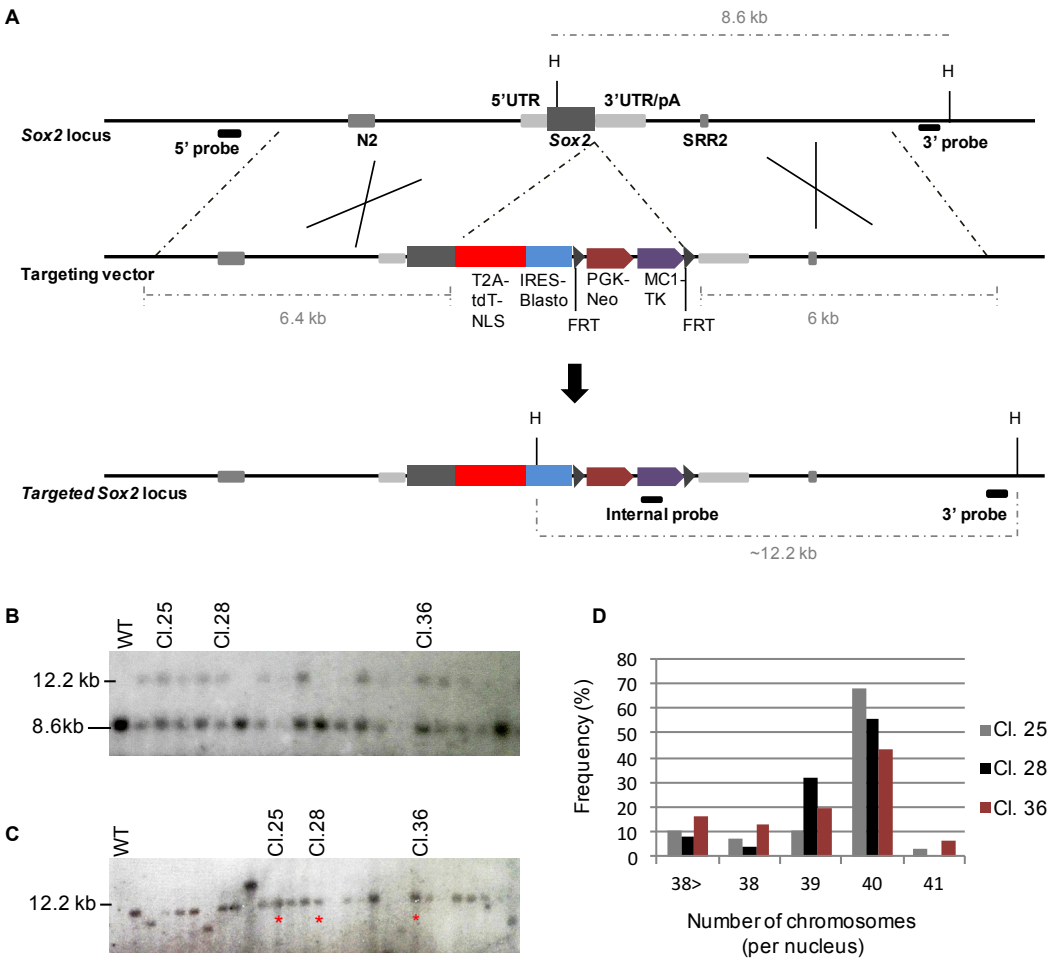


Figure 5.7: Sox2 locus targeting and generation of a Tbra/Sox2 double reporter cell line. **A)** Schematic illustration of the endogenous *Sox2* locus, the targeting construct and the insertion of the reporter cassette through homologous recombination. The doted lines next to the targeting vector indicate the size of the homology region between the endogenous locus and the construct. "H" denotes the cutting site of the restriction enzyme HindIII that was used for southern blot analysis. **B)** Southern blot analysis of G418-resistant clones, using the 3'prime probe shown in A. DNA fragments of wild-type (WT) clones correspond to 8.6kb, whereas targeted clones present bands of 12.2kb. **C)** Southern blot analysis using the internal probe shown in A. **D)** Karyotypic analysis of correctly targeted clones.

When targeted ES cells were observed under the microscope, tdTomato fluorescence signal could not be detected. tdTomato expression was detected by FACS, but surprisingly, it was shown that almost 60% of the cell population was tdTomato negative. To investigate this observation further, the tdTomato positive and negative fractions were FACS sorted (figure 5.8, A). To look into the possibility that the targeted clones could have been contaminated with wild-type cells, genomic DNA was extracted from the tdTomato negative fraction and used for a PCR reaction designed to detect and amplify the MC1-TK cassette. The expected PCR product was generated, proving that at least some of the tdTomato negative cells did contain the reporter cassette into their genome (figure 5.8, B). Cells from the tdTomato positive fraction were replated in ES cell medium supplemented with G418 and in ES cell medium lacking drug selection. FACS analysis demonstrated that in the selection-free medium, 80-94% of the population turned into tdTomato negative within 16 days and even in the presence of G418, tdTomato negative cells accounted for 40-45% of the culture, within a week (figure 5.8, C). Given that the tdTomato negative cells contain the reporter cassette into their genome, it is possible that the loss of tdTomato expression is due to gene silencing.

In order to eliminate any effect that the PGK-Neo::MC1-TK cassette may have on the tdTomato expression levels, and also to allow for germ line transmission of the reporter cassette (explained in section 5.2), the FRT-flanked Neo::TK cassette was excised from the targeted clones (figure 5.9, A). Colonies were picked and expanded in ES cell medium supplemented with blasticidin. Excision of the cassette was confirmed by gDNA PCR (figure 5.9, B) and incubation of the clones in the presence of G418 (figure 5.9, C). FACS analysis of FRT-excised clones grown in the presence of blasticidin demonstrated that the intensity of the tdTomato fluorescence signal remained very low and still not detectable by microscopy. Interestingly, in contrast to what we had observed before the FRT-flanked cassette excision, when ES cells were cultured in the absence of any selection for 3 weeks, the tdTomato levels remained stable (figure 5.9, D).

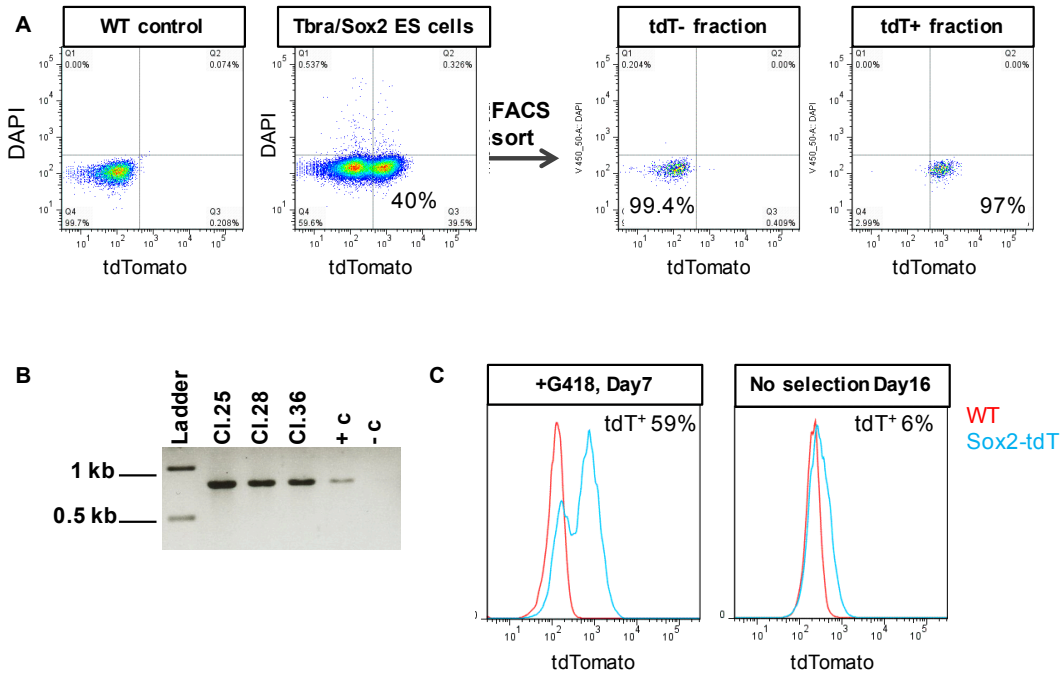


Figure 5.8: Sox2-tdTomato expression in ES cells. A) Following colony picking, clones were expanded in ES cell medium without any selection. FACS analysis of correctly targeted clones demonstrated the presence of a tdTomato negative and a tdTomato positive fraction. The two subpopulations were FACS sorted and used for further analysis. The purity of the sorted fractions is presented on the right. **B)** Genomic DNA was extracted from the sorted tdTomato negative cells and analysed by PCR, using primers that bind to the MC1-TK cassette. The length of the expected PCR product is 800bp. "+c" stands for positive control and "-c" denotes negative control**. C)** The sorted tdTomato positive fractions were replated in ES cell medium with and without G418. The tdTomato expression levels were recorded by FACS.

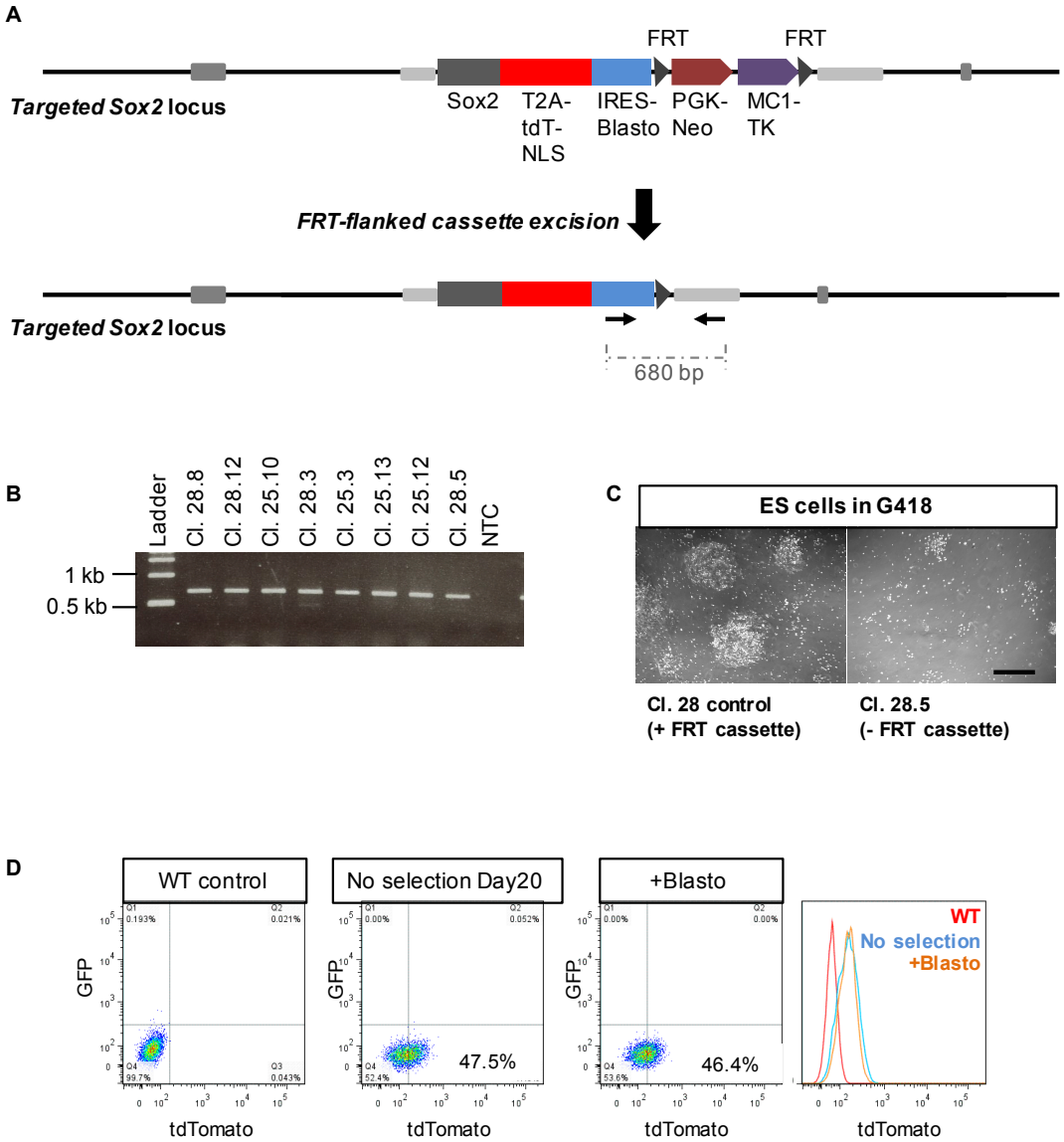


Figure 5.9: Sox2-tdTomato expression in ES cells following Neo::TK cassette excision. A) Schematic illustration of the FRT-flanked cassette excision. **B)** gDNA was extracted from FRT-excised clones and analysed by PCR. The binding sites of the primers are shown on A as black arrows. When Neo::TK cassette has been removed, the PCR reaction gives rise to a product of 680 bp. NTC denotes non-template control. **C)** Sox2 targeted clones before (left) and after (right) the Neo::TK cassette excision, cultured in G418 for 7 days. Scale bar: 600µm. **D)** Following Neo::TK cassette excision, clones were grown in ES cell medium with and without blasticidin (Blasto) for 20 days and the tdTomato expression levels were recorded by FACS. WT, wild-type; Cl, clone; NTC, no template control.

5.6. Tbra-GFP/Sox2-tdT ES cell and EpiSC line validation

To assess the correlation between the levels of *Sox2* transcript and tdTomato fluorescence signal, ES cell cultures were FACS sorted into a tdTomato-negative and a tdTomato-high subpopulation and QRT-PCR was performed. This showed that the tdTomato-high fraction was enriched in *Sox2* transcript levels (figure 5.10, A). Immunofluorescence analysis of whole reporter ES cell cultures using an anti-tdTomato antibody demonstrated a good correlation between tdTomato and Sox2, suggesting altogether that this cell line constitutes a reliable reporter for Sox2 (figure 5.10, A). Image analysis showed that tdTomato/Sox2 double positive cells accounted for 50% of the culture, while 42% of the cells were positive for Sox2 only, which agrees with the FACS analysis presented in figure 5.9, D. Taking into account that the presence of blasticidin in the culture medium does not alter the percentage of the tdTomato positive cells in the population (figure 5.9, D), overall these findings indicate that the tdTomato cassette is expressed in ES cells, but the signal is so low, that a large fraction of the cells appears to be negative. Nevertheless, differences in the half-lives of the tdTomato and Sox2 proteins may also contribute to the formation of the $Sox2^{+}tdTomato^{-}$ fraction.

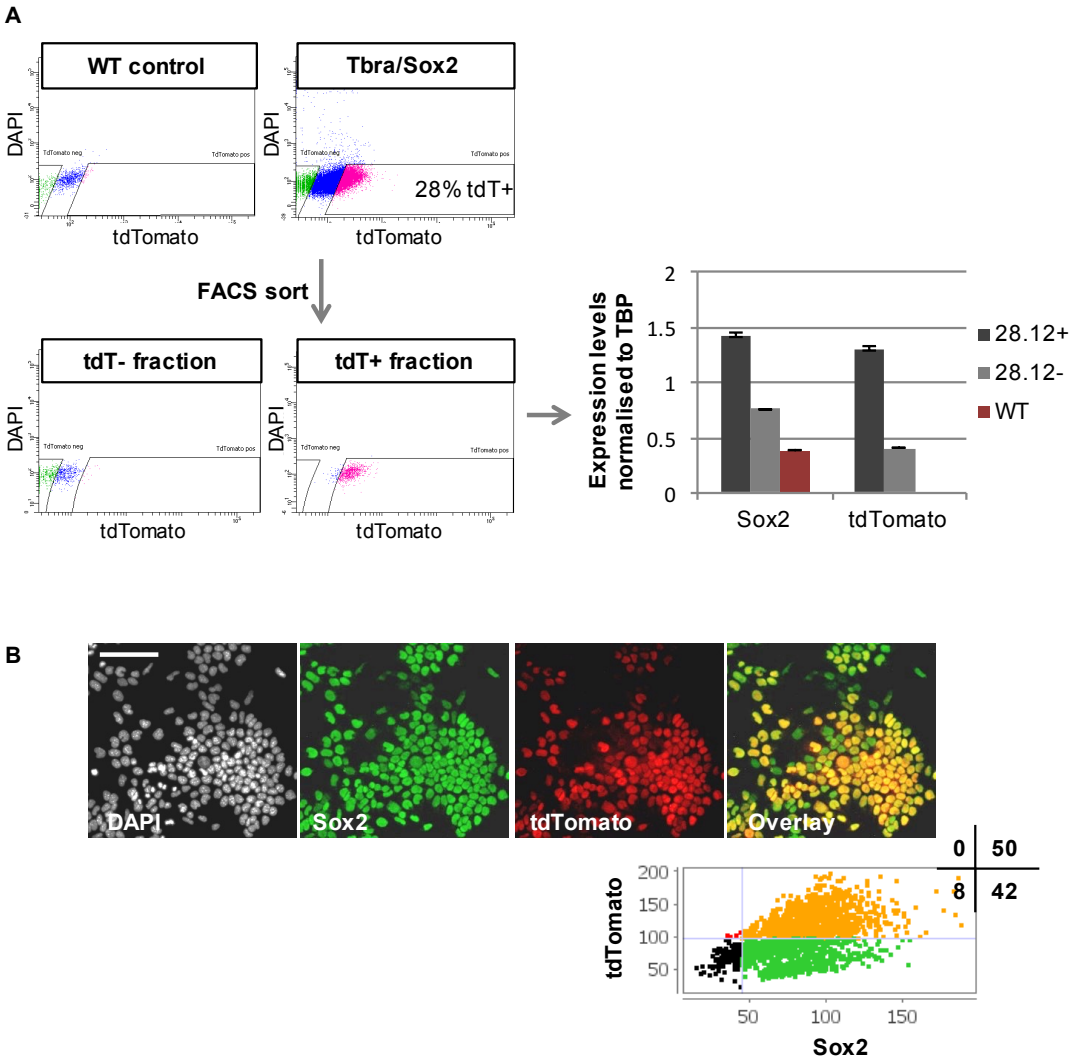


Figure 5.10: Validation of the Tbra-GFP/Sox2-tdT ES cell line. Tbra/Sox2 double reporter ES cells (clone 28.12) were FACS sorted into a tdTomato-negative and a tdTomato-high expressing fractions. The purity of the sorted subpopulations is presented. RNA was extracted from each fraction and used for QRT-PCR analysis of Sox2 and tdTomato. WT denotes wild-type cells. 28.12+/- stand for the tdTomato^{+} and tdTomato^{-} fractions of sorted clone 28.12. Error bars represent standard deviation of three technical replicates in one experiment. **B)** Immunocytochemistry and image quantification of endogenous Sox2 and tdTomato in ES cell cultures. Scale bar: 100µm. The plot presents the correlation between the expression levels of endogenous Sox2 and tdTomato and the numbers show the percentages of cells in each quadrant.

EpiSC were derived *in vitro* from Tbra-GFP/Sox2-tdT double reporter ES cells. FACS analysis of EpiSC cultures demonstrated robust induction of Tbra-GFP, but very low intensity of Sox2-tdTomato (figure 5.11). To verify that EpiSCs express the tdT-Blasto cassette, the minimum working concentration of blasticidin in the serum-free EpiSC culture conditions was identified (3µg/ml) and Tbra-GFP/Sox2-tdT EpiSC were cultured under these conditions for 7 days. These cells survived in the presence of the drug and the tdTomato levels remained very low (figure 5.11), showing that the transgenic cassette is being expressed in EpiSCs, but at lower levels compared to ES cells.

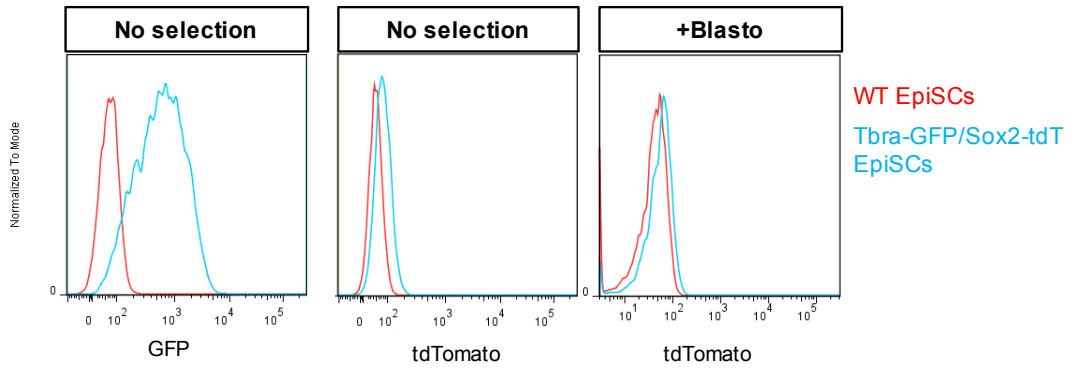


Figure 5.11: Expression of the Sox2-tdTomato reporter in EpiSCs. FACS analysis of GFP and tdTomato in Tbra/Sox2 double reporter EpiSCs. Blasto, blasticidin.

The capacity of the reporter cells to contribute to normal embryonic development and the expression pattern of Sox2-tdTomato *in vivo* were assessed by generating chimeric embryos through morula aggregation. Embryos were collected at E9.5 and E10.5 and chimerism was identified by expression of Tbra-GFP (figure 5.12). 30 out of 34 resulting embryos were chimeric (=88%). Tbra-GFP signal was observed along the notochord and in the tail bud, including the CNH (figure 5.12, A and B), which is where endogenous *Tbra* is known to be expressed at these stages (Kispert and Herrmann, 1994; Cambray and Wilson, 2007; Wymeersch *et al.*, 2016). In contrast to GFP, tdTomato fluorescence could not be detected in any of the embryos. Since in ES

cell cultures tdTomato signal was detected only by FACS and not by microscopy, embryos were subjected to whole mount immunohistochemistry for tdTomato. Immunofluorescence failed to reveal tdTomato expression. However, it is important to note that the anti-tdTomato antibody used for this experiment had not been used before for whole mount immunofluorescence, and since no positive control was available at that point, we could not extract any information about the tdTomato expression in the chimeras. Crucially, the majority of the chimeric embryos were morphologically abnormal (figure 5.12, A and C). Most of the embryos carried defects in neural tissues, including a kinked neural tube of varying severity and abnormal head formation. In a smaller proportion of the chimeric embryos, a spherical mass of tissue had grown at the region where the umbilical cord develops, indicative of failure of the allantois to fuse with the ectoplacental cone. Normal tail phenotypes were observed in the vast majority of the chimeric embryos, while very few chimeras exhibited aberrant tail bud morphology.

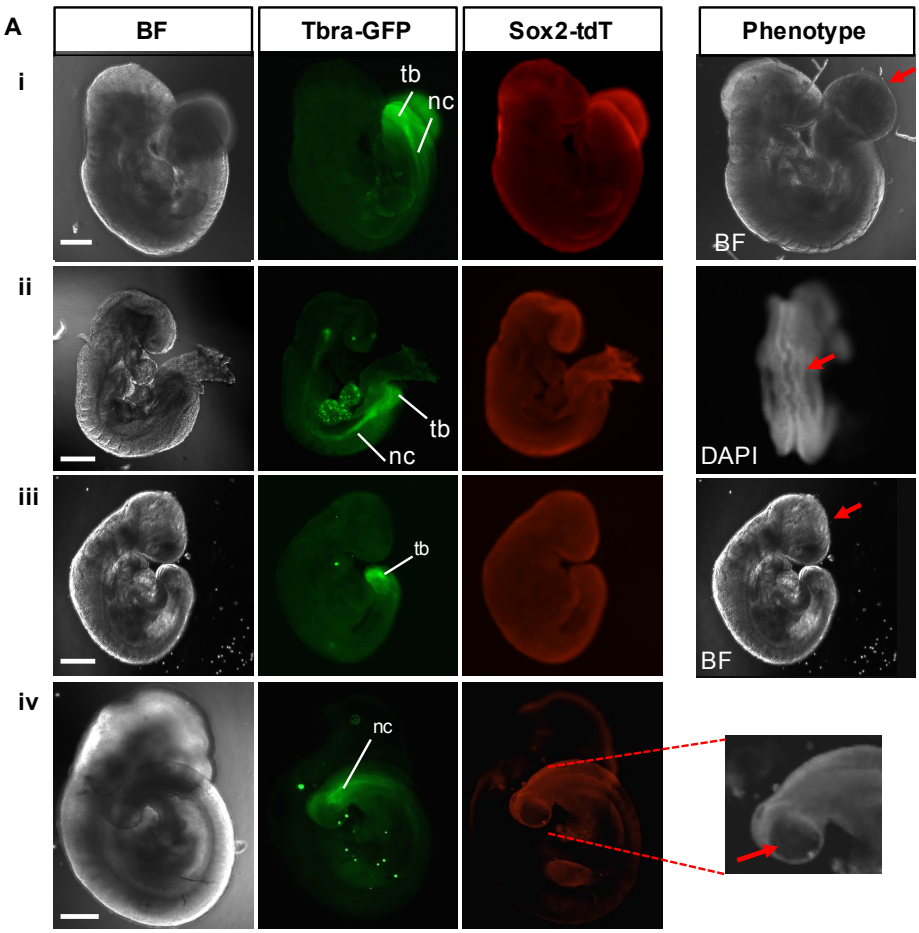

A
BF
Tbra-GFP
Sox2-tdT
Phenotype
i
ii
iii
iv
tb
nc
BF
DAPI
B
DAPI
Tbra-GFP
Sox2-tdTomato
Autofluorescence
psm
nc
CNH
C
Morphology of chimeric embryos
Number of embryos
18
16
14
12
10
8
6
4
2
0
E8.5
E9.5
E10.5
Abnormal
Normal

Figure 5.12: Generation of Tbra-GFP/Sox2-tdT chimeric embryos. A) Immunofluorescence for GFP and tdTomato in chimeric embryos generated through morula aggregation (clones 25.3 and 28.12). The embryos shown here are representative examples of the phenotypes that the majority of the embryos exhibited. Red arrows point to the body structure that appears abnormal, including the allantois (i), the neural tube (ii), the head (iii) and the tail bud (iv). Scale bars: 300μm. **B)** Transverse sections of a Tbra-GFP/Sox2-tdT chimeric E9.5 embryo. In addition to the channels necessary for DAPI, GFP and tdTomato imaging, the far-red channel was also used in order to assess autofluorescence. Circled areas constitute highly autofluorescent blood cells. **C)** Embryos were collected on E9.5 and E10.5 and chimeras were scored based on developmental morphology. Two of the embryos collected on E9.5 were still at the very early stages of somitogenesis, therefore they were classified as E8.5. BF, bright field; tb, tail bud; nc, notochord; psm, presomitic mesoderm; CNH, chordoneural hinge.

5.7. Discussion

The Tbra-GFP-Pac reporter cell line opens the way to new experimental approaches for future NMP studies

In this study, the CRISPR/Cas9-mediated generation of a Tbra-GFP-Pac reporter mES cell line is presented. The targeting construct was designed so as not to disrupt the sequences of the two alleles, therefore the cell line is homozygous for *Tbra*. Expression analysis demonstrated that GFP faithfully recapitulates the endogenous expression of Tbra *in vivo* and *in vitro,* while chimeric mice, including embryos and adults, appeared physiologically normal. Given that *Tbra* heterozygosity disrupts gastrulation and axis elongation (Wilson *et al.*, 1995), the potential of the new reporter line to give rise to physiologically normal chimeras makes this cell line a very useful tool for *in vivo* studies. The Tbra promoter-driven puromycin resistance is another important feature of these cells that offers the opportunity for lineage selection. In addition, GFP can readily be detected by microscopy, without immunostaining being necessary, suggesting that in the future, live imaging experiments could possibly be performed.

As discussed in 1.4.3, stimulation of Fgf and Wnt signalling induces NMP identity in ES cells and EpiSCs (Gouti *et al.*, 2014; Turner *et al.*, 2014). However, the generation of *in vitro* NMPs is transient, as prolonged treatment with FGF/CHI results in reduction of Tbra/Sox2 double positive cells and differentiation of the bipotent progenitors into the neural and mesoderm lineages (Gouti *et al.*, 2014; Tsakiridis and Wilson, 2015). Crucially, the culture conditions that promote the self-renewal of *in vitro* NMPs are not known yet. Here, FACS analysis indicated that Tbra-GFP expression peaks in EpiSC cultures treated with FGF/CHI for 2-3 days and thereafter it decreases, until day 8 when it becomes almost undetectable. Since Tbra is expressed in NMPs (and in early mesoderm), but not in neural or late mesoderm differentiation products, self-renewal of the bipotent progenitors would be indicated by maintenance of the Tbra-GFP levels for an extended period of time (>3 days at least). Hence, Tbra-

GFP expression and puromycin resistance could be used as a quick read-out for the identification of culture condition that could promote the self-renewal of NMPs. Such conditions would include the inhibition and/or stimulation of signalling pathways that are known to be active in the axial progenitor regions, such as the Fgf, Wnt, Notch and Shh pathways (reviewed in Wilson *et al.*, 2009), but also different substrates, such as gelatin or laminin.

Based on the same principle, i.e Tbra-GFP expression decreases progressively as NMPs exit the bipotent state and differentiate, the Tbra-GFP-Pac reporter line could be a valuable tool for the investigation of the gene network underlying the regulation of self-renewal of NMPs. For this purpose, a genome-wide screen could be performed, employing piggyBac transposition to activate endogenous gene expression at random (Guo and Smith, 2010). More specifically, the genomic screen would involve transfection of Tbra-GFP-Pac EpiSCs with a piggyBac gene-trap vector containing the murine stem cell virus (MSCV) enhancer/promoter and a splice donor site (Dupuy *et al.*, 2005). This construct can activate nearby genes in either orientation, although the splice donor sequence can affect only genes found downstream of the insertion site. Hence, it has been hypothesised that activation of genes that promote the proliferation of NMPs would result in Tbra-GFP positive and puromycin resistant colonies that would persist under prolonged treatment with FGF/CHI. The resistant colonies would be then used for the identification of the targeted genes, through Splinkerette PCR. Alternatively, genetic screening could be performed using CRISPR/Cas9 technology to introduce loss-of-function mutations into genomic DNA (Shalem *et al.*, 2014; Zhou *et al.*, 2014). CRISPR knockout libraries are commercially available and consist of thousands of plasmids, each containing a gRNA towards a different target gene. Hence, in the future, a genome-scale CRISPR-Cas9 knockout screening in Tbra-GFP-Pac cell line could help us to identify previously unknown genes that function as negative regulators of NMP self-renewal.

In summary, the expression profile of Tbra-GFP in EpiSC cultures under prolonged exposure to FGF/CHI offers the opportunity to develop new strategies for the study of

the molecular mechanisms underlying the self-renewal and differentiation of NMPs. What does the Tbra-GFP$^+$ fraction that emerges in *in vitro* NMP cultures represent though? Could Tbra-GFP be used as a single marker for the identification and isolation of NMP-like cells? In chapter 4 and in the study of Gouti *et al.* (2014), it is presented that *in vitro* NMP cultures are heterogeneous and not all of the Tbra$^+$ cells express *Sox2*. In addition, QRT-PCR analysis in sorted Tbra-GFP positive and negative fractions of 48h NMP cultures demonstrated that the two fractions are characterised by the same expression levels of *Evx1, Nkx1.2*, *Cdx2* and *Sox2. In vivo* studies have shown that although Tbra/Sox2 coexpression overlaps extensively with the NMP status, all of the markers used for the identification of NMPs, including Tbra and Sox2, are found in tissues outside the caudal progenitor regions too. Importantly, the Tbra$^+$Sox2$^+$ cells found in NMP regions of the mouse embryo are characterized by low-to-medium expression levels of both genes, while high expression levels of *Tbra* and *Sox2* were detected in regions fated to mesodermal and neural differentiation respectively (Cambray and Wilson, 2007; Wymeersch *et al.*, 2016). Hence, according to the *in vivo* data, it is most likely that within the FGF/CHI-treated EpiSC cultures, NMP-like cells are those with relatively low levels of *Tbra* mRNA, whereas cells that express *Tbra* at very high levels may represent a fraction of the culture that is committed to mesoderm differentiation. This hypothesis is further supported by the observation that the Tbra-GFP$^+$ fraction was enriched in transcripts of *Wnt3a*, but exhibited lower expression levels of *Sox1*. Since Sox2 is not present in mesoderm derivatives, it is suggested that a *Tbra/Sox2* double reporter line would be an optimum tool for tracking and isolating a more homogeneous NMP-like population from *in vitro* cultures.

Targeting of *Sox2* locus: challenges and next steps

In order to generate a *Tbra/Sox2* double reporter line, *Sox2* locus was targeted in the Tbra-GFP-Pac ES cells, using a vector designed (by Frederick Wong) to replace the stop codon with a T2A::tdT-NLS::IRES::Blasto cassette. Although tdTomato is a very bright fluorescent protein, correctly targeted ES cell clones exhibited very faint red

fluorescent signal, that could be detected by FACS but not microscopy. When ES cells were differentiated into EpiSCs, the levels of tdTomato decreased and although reporter cells exhibited blasticidin resistance, the FACS plot of Sox2/tdTomato EpiSCs appeared similar to the plot of wild-type cells. The reduced tdTomato levels in EpiSC cultures agree with unpublished data of our group indicating that *Sox2* levels decrease as ES cells turn into EpiSCs. However, detection of such a weak tdTomato fluorescent signal in ES cells is very surprising. A former member of our group has previously introduced a T2A::H2B-tdTomato::IRES::Neo cassette downstream of the *Sox2* exon (Frederick Wong, unpublished data), generating a reporter cell line characterized by bright red fluorescence. This suggests that the regulatory elements around the *Sox2* gene do not affect the levels of fluorescence. Could the addition of the 3xNLS sequence (around 30 amino acids) have altered the folding and consequently the properties of tdTomato? In order to address this question, the tdTomato-NLS and tdTomato without NLS DNA sequences could be cloned under a eukaryotic, constitutively active promoter, such as PGK and the vectors could be transferred into ES cells. If the NLS sequence affects the properties of tdTomato, then the presence of NLS should result in reduced (if any) fluorescent signal.

Use of Tbra/Sox2 double reporter ES cells for morula aggregations gave rise to chimeras with defects in the CNS, including kinked neural tubes and head malformations. Interestingly, these phenotypes are very similar to the developmental abnormalities that have been observed in chimeric embryos, generated by a Sox2::T2A::H2B-tdT::IRES::Neo single reporter ES cell line (Frederick Wong, unpublished data.). More specifically, in the latter case, kinked neural tubes, failure of neural tube closure and smaller head folds were observed, while adult mice exhibited defects in development of genitalia too. The results of both studies are very intriguing, as the targeting constructs were designed so as not to delete any part of the *Sox2* exon and the 3' UTR; in both cases, the only modification that cells carry in one of the two *Sox2* alleles is the insertion of a ~2.5kb cassette between the end of the *Sox2* exon and the beginning of the 3' UTR. Comparison of the *Sox2* expression levels in sorted tdTomato ES cell subpopulations and wild type cells demonstrated that the reporter cells express *Sox2* at higher levels, which is probably due to culture of Sox2-tdTomato

ES cells in the presence of blasticidin (figure 5.10, A). In the future, in order to assess whether the presence of the tdTomato cassette affects the expression levels of *Sox2*, it will be important to compare the *Sox2* RNA and protein levels in targeted and wild-type cells at the population level, in cell cultures grown in the absence of blasticidin.

How could this insertion of exogenous DNA potentially interfere with *Sox2* expression though? A number of *Sox2* regulatory elements have been identified downstream of the *Sox2* exon (Uchikawa *et al.*, 2003; Kamachi *et al.*, 2009). SRR2 is one of them; however, its activity is limited to ES cells and possibly pre-implantation embryos (Tomioka *et al.*, 2002), therefore it is unlikely that it is relevant to the phenotype we observed. N1 has been shown to drive *Sox2* expression in the posterior domains of the forming neural primordium (Takemoto *et al.*, 2006). Two more *Sox2* enhancers, N5 and N4, have been identified in the same region of the locus and they seem to induce *Sox2* expression in the developing brain and spinal cord (Kamachi *et al.*, 2009). Hence, one possible explanation is that the presence of the ~2.5kb exogenous DNA at the 3' end of the *Sox2* exon may interfere with the activity of the regulatory elements, e.g. enhancers, that are found downstream of the *Sox2* coding region.

Could disruption of the expression of only one of the two *Sox2* alleles have such an effect on the development of the CNS though? The overlap in the expression profile of *Sox1*, *Sox2*, *Sox3* in the CNS and evidence provided by functional assays have suggested potential redundancy within the *Sox* gene family (Pevny and Lovell-Badge, 1997; Uchikawa *et al.*, 2011). However, a gene-dosage study on neural retinal cells, where *Sox1* and *Sox3* become downregulated, whereas *Sox2* expression is maintained, has shown that the levels of *Sox2* expression are critical for the regulation of self-renewal and differentiation of neural progenitors (Taranova *et al.*, 2006). No phenotype has been reported for $Sox2^{eGFP}$ and $Sox2^{\beta geo}$ heterozygous mice, apart from reduced fertility in $Sox2^{\beta geo+/-}$ adult males (Avilion *et al.*, 2003; Ellis *et al.*, 2004). *Sox2* null embryos die shortly after implantation due to defects of the epiblast (where Sox2 acts as a pluripotency factor and neither *Sox1* nor *Sox3* are expressed) (Avilion *et al.*, 2003). Interestingly, inducible excision of *Sox2* in E6.5 to E7.0 embryos resulted in

the formation of hydrocephalus (11/20 embryos) and kinked neural tube (2/20 embryos) (DeVeale *et al.*, 2013). Since only one of the two *Sox2* alleles contains the tdTomato cassette, the fact that the phenotype we have observed is similar to the one that emerges after inducible deletion of both *Sox2* alleles is very intriguing. Based on the fact that Sox2-tdTomato chimeras did not exhibit epiblast defects, it seems that the presence of the tdTomato cassette does not interfere with the role of Sox2 as a pluripotency factor, but in a way that is not understood yet, it disrupts the normal function of neural progenitor cells.

The CNS was not the only region of the chimeric mice that exhibited abnormalities though. In some Tbra-GFP/Sox2-tdT embryos, the allantois failed to fuse with the ectoplacental cone, while very few chimeras exhibited aberrant morphology of the caudal extreme of the tail. Irregular development of the allantois has been observed in the *Tbra* null mouse embryo and it is considered the physiological cause of embryonic death that occurs at mid-gestation (around E10.5). Abnormalities in the allantois and defects in the distal tail region have also been observed in chimeras containing *Tbra* null cells. Interestingly, Tbra has only been detected in a very early basal component of the allantois (Inman and Downs, 2006), therefore it has been speculated that Tbra may be required for the initial specification of allantoic precursors within the streak region (Beddington *et al.*, 1992; Wilson *et al.*, 1995; Wilson and Beddington, 1997). Hence, it is most likely that the defects in allantois and the few cases of tail tip abnormalities we have observed in the present study are due to targeting of *Tbra*, and not *Sox2* gene.

In summary, the two main issues that arose from introducing the T2A::tdT-NLS::Blasto cassette in the *Sox2* locus is the very faint red fluorescent signal that targeted ES cells show and the CNS defects that chimeras exhibit. One possible way to generate a reporter line that will give rise to physiologically normal embryos is to minimize the disruption of *Sox2* locus, by introducing a cassette of smaller size. Hence, we could sacrifice the blasticidin resistance gene and we could replace the *Sox2* stop codon with a T2A::mOrange cassette. mOrange is a fluorescent protein that has half

the size of tdTomato and is very bright (Shaner *et al.*, 2004). Fluorescent proteins have a tendency for oligomerization, particularly when they are attached to proteins such as actin or histones, which form oligomers as part of their natural function. Importantly, protein oligomerization can lead to the formation of aggregates that can disrupt the normal function of cells (Olenych *et al.*, 2007). In addition, a study in yeast cells has shown that the presence of localization signals in GFP proteins expressed at high levels affected cellular growth in the same way as misfolded proteins (Kintaka *et al.*, 2016). Although this study did not detect any toxic effect of NLS on the growth of yeast cells, we cannot exclude the possibility that this effect varies amongst different organisms, different cell types or even different fluorescent proteins. Altogether, these observations indicate that forced localization of highly expressed proteins can potentially cause cellular defects, therefore in the future, it may be beneficial to avoid the fuse of the fluoresce protein with H2B or NLS sequence.

Although the present Tbra-GFP/Sox2-tdT reporter cell line is not optimal, it could still be used for the screening experiments described above. Given the very faint tdTomato fluorescent signal in ES cell and EpiSC cultures, it is predicted that FACS analysis will not allow us to distinguish Sox2 positive from Sox2 truly negative cells in NMP cultures. However, EpiSCs do exhibit blasticidin resistance, that could be used as a marker of *Sox2* expression. Hence, since coexpression of Tbra/Sox2 is a more reliable NMP marker than the presence of Tbra alone, screening experiments could be performed applying simultaneous puromycin (*Tbra*-driven resistance) and blasticidin (*Sox2*-driven resistance) selection or combining blasticidin treatment with FACS analysis for Tbra-GFP.

Chapter 6: General discussion

In this study, I first examine whether Tbra^{+} EpiSCs correspond to any known embryonic cell types, in particular NMPs. I show that a big fraction of the EpiSC cultures is Tbra/Sox2 double positive. However, Nanog is detected in the vast majority of this fraction suggesting that these cells are pluripotent rather than bipotent NMPs. In contrast to *Sox2*, my data agrees with published information showing that in self-renewing conditions, *Tbra* expression is mutually exclusive with expression of *Sox1-GFP*, that appears to mark cells with neural-like characteristics (Tsakiridis *et al.*, 2014). Using a previously published Tbra-GFP reporter cell line (Fehling *et al.*, 2003), I demonstrate that the Tbra-GFP^{+} fraction exists in a dynamic equilibrium with GFP negative cells and exhibits an early primitive streak-like character. As the coexpression of *Tbra* with pluripotency markers suggested, this subpopulation is not committed to the mesoderm/endoderm lineages, but it undergoes neural differentiation more efficiently than Tbra-GFP negative cells. The detection of mesendoderm-like and node/notochord-like subpopulations (in particular the latter, which has not previously been described in EpiSC populations), is of interest. *In vivo* the mesendoderm and the source of Shh, node/notochord, have an important role in neural differentiation and patterning (reviewed Arnold and Robertson, 2009). Hence, in the future, it will be very interesting to characterize these subpopulations further and look into how they interact with epiblast-like cells within the culture, not only under self-renewing conditions but also when differentiation is promoted. The characterization of the distinct subsets of cells that are included in the EpiSC populations will help us to dissect the mechanisms underlying the heterogeneity that stem cell cultures exhibit and it could help us to model tissue interactions *in vitro*, helping us in this way to improve current differentiation protocols.

While testing the hypothesis that *Sox1/Tbra* expression is a more stringent marker for NMP than *Sox2*, which is also expressed in pluripotent cells, I found that *Sox1-GFP* is expressed in a fraction of the Tbra/Sox2 double positive cells, in FGF/CHI-treated EpiSC cultures. The Sox1-GFP^{+} fraction is characterized by expression of NMP

related markers and appears to be enriched in *Hox* genes, including genes from the 5' paralogous groups 8-10, suggesting altogether that Sox1-GFP marks cells that are the *in vitro* equivalent of the NMPs found at the tail bud stages. However, the presence of Sox1/Sox2 double positive cells and the upregulation of the posterior neural marker *Gbx2* in the Sox1-GFP$^+$ cell suggests that this fraction may also contain neural committed cells.

The apparent NMP phenotype of Sox1-GFP cells in NMP conditions thus contrasts with their apparent inability to give rise to mesoderm. Does that mean that Sox1-GFP positive cells are neural committed? The same experiment has been performed by a former member of our group investigating the differentiation potential of Tbra-GFP$^+$ cells of *in vitro* NMP cultures (Tsakiridis and Wilson, 2015). Interestingly, the Tbra-GFP$^+$ cells exhibited strong bias towards neural differentiation too (55% Sox2$^+$ colonies); mesoderm like colonies accounted for 9% of the total number of clones, while mixed colonies consisting of both mesoderm-like and neural-like differentiation products, which are indicative of NM bipotency, formed 12% of the clones (Tsakiridis and Wilson, 2015). Since cells that express the mesoderm marker *Tbra* exhibited bias towards neural differentiation too, it is possible that prolonged exposure to FGF/CHI favours neural over mesoderm differentiation. The notion that these culture conditions possibly affect the fate of progenitor cells is further supported by comparing these *in vitro* data with the grafting experiments presented by Gouti *et al.* (2014). Analysis of the contribution of the grafted NMPs in host embryos demonstrates the opposite pattern of differentiation bias, with cells contributing to paraxial mesoderm in 10/10 embryos, but incorporating in the neural tube of only 4/10 hosts (Gouti *et al.*, 2014). Although grafting experiments are not performed at the single cell level, as the *in vitro* clonal analysis, these data suggest that prolonged exposure of cells to FGF/CHI does not mimic the environment that *in vivo* NMPs are exposed to and therefore may not allow the full potential of Sox1-GFP$^+$ cells to develop. Hence in the future, the mesoderm differentiation capacity of Sox1-GFP$^+$ cells would be tested by culturing them in mesoderm inducing conditions or ideally, transplanting them into the NSB of E8.5 embryos. The identification of Sox1-GFP as a marker of late NMPs could help us to model the maturation of NMPs *in vitro* and since late NMPs are the precursors

of more posterior tissues, it could contribute to the development of refined differentiation protocols that aim to generate muscle cells or spinal cord neurons of lower axial levels.

As mentioned before, there are no unique markers for NMPs, but the genes that are active in NMP regions *in vivo* are also expressed in tissues fated to mesoderm or neural differentiation. That makes the isolation of a pure NMP population challenging, as FACS sorting based on single markers results to the isolation of fractions that contain bipotent progenitors but also lineage committed cells. This phenomenon is illustrated by the QRT-PCR analysis in sorted Tbra-GFP (chapter 5) and Sox1-GFP (chapter 4) fractions, showing that the NMP related genes *Nkx1.2*, *Evx1* and *Cdx2* are expressed in positive and negative fractions at exactly the same levels, while the Tbra-GFP$^+$ fraction seems to be enriched in *Wnt3a* transcripts and the Sox1-GFP$^+$ fraction enriched in *Gbx2* RNA. In addition, QRT-PCR analysis and immunocytochemistry indicate that Tbra correlates with lower levels of Sox1-GFP, which is consistent with a recent study reporting the same correlation between Tbra and Sox2 in NMP regions *in vivo* (Wymeersch et al., 2016). Altogether, these data suggest that double reporter lines for a mesodermal and a neural specific marker are required in order to dissect the properties of the subpopulations that exist in *in vitro* NMP cultures and be able to distinguish between bipotent progenitors and lineage committed cells. Based on that, I generated a reliable Tbra-GFP reporter ES cell line that in contrast to the one published before, contains both endogenous *Tbra* loci intact and gives rise to physiologically normal chimeras. Since *Tbra* heterozygosity disrupts the normal migration of mesoderm from the primitive streak and leads to axis truncations (Beddington *et al.*, 1992; Wilson *et al.*, 1995), this cell line constitutes an optimum tool for axis elongation studies. By targeting the *Sox2* locus in the Tbra-GFP ES cells, I generated a Tbra-GFP/Sox2-tdT double reporter ES cell line; although the use of this line has certain limitations (discussed in detail in chapter 5), it opens the way to new experimental approaches that would allow us to dissect the molecular mechanisms underlying the self-renewal and differentiation of NMPs. In the future, a Tbra/Sox1 double reporter line could be generated too, targeting the Tbra-GFP ES cell line with

a Sox1-tdT construct that has already been prepared (modified from Aubert *et al.*, 2003).

In summary, in this study, I describe early lineage specification events that take place in EpiSC and *in vitro* NMP cultures and I identify subpopulations with distinct properties. In the future, the use of the tools I developed during this project will help us to take us these findings further and investigate the molecular mechanisms underlying the generation, the maintenance and the differentiation of these subpopulations, contributing overall to the development of differentiation protocols of clinical or basic science interest.

References

Abdelkhalek, H. B., Beckers, A., Schuster-Gossler, K., Pavlova, M. N., Burkhardt, H., Lickert, H., Rossant, J., Reinhardt, R., Schalkwyk, L. C., Muller, I., Herrmann, B. G., Ceolin, M., Rivera-Pomar, R. and Gossler, A. (2004). The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation. *Genes Dev, 18*(14), 1725-1736.

Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D. and Bekman, E. (2009). Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo. *PLoS One, 4*(7), e6286.

Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A. and Brulet, P. (1995). Forebrain and Midbrain Regions Are Deleted in Otx2(-/-) Mutants Due to a Defective Anterior Neuroectoderm Specification during Gastrulation. *Development, 121*(10), 3279-3290.

Alshawi, R., Burke, J., Jones, C. T., Simons, J. P. and Bishop, J. O. (1988). A Mup Promoter-Thymidine Kinase Reporter Gene Shows Relaxed Tissue-Specific Expression and Confers Male-Sterility Upon Transgenic Mice. *Molecular and Cellular Biology, 8*(11), 4821-4828.

Andre, P., Song, H., Kim, W., Kispert, A. and Yang, Y. (2015). Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development, 142*(8), 1516-1527.

Arkell, R. M. and Tam, P. P. (2012). Initiating head development in mouse embryos: integrating signalling and transcriptional activity. *Open Biol, 2*(3), 120030.

Arnold, S. J. and Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol, 10*(2), 91-103.

Aubert, J., Stavridis, M. P., Tweedie, S., O'Reilly, M., Vierlinger, K., Li, M., Ghazal, P., Pratt, T., Mason, J. O., Roy, D. and Smith, A. (2003). Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. *Proc Natl Acad Sci U S A, 100 Suppl 1*, 11836-11841.

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B. and Herrmann, B. G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev Cell, 4*(3), 395-406.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev, 17*(1), 126-140.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., McMahon, J. A., McMahon, A. P., Harland, R. M., Rossant, J. and De Robertis, E. M. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature, 403*(6770), 658-661.

Bancroft, J. D. and Gamble, M. (2002). *Theory and Practice of Histological Techniques* (5th ed.). Edinburgh: Churchill Livingstone.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K. and Surani, M. A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature, 461*(7268), 1292-1295.

Bastian, H. and Gruss, P. (1990). A Murine Even-Skipped Homolog, Evx-1, Is Expressed during Early Embryogenesis and Neurogenesis in a Biphasic Manner. *Embo Journal, 9*(6), 1839-1852.

Beck, F., Erler, T., Russell, A. and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn, 204*(3), 219-227.

Beddington, R. S. (1983). Histogenetic and neoplastic potential of different regions of the mouse embryonic egg cylinder. *J Embryol Exp Morphol, 75*, 189-204.

Beddington, R. S., Rashbass, P. and Wilson, V. (1992). Brachyury--a gene affecting mouse gastrulation and early organogenesis. *Dev Suppl*, 157-165.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. and Constam, D. B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. *Dev Cell, 11*(3), 313-323.

Bernemann, C., Greber, B., Ko, K., Sterneckert, J., Han, D. W., Arauzo-Bravo, M. J. and Scholer, H. R. (2011). Distinct developmental ground states of epiblast stem cell lines determine different pluripotency features. *Stem Cells, 29*(10), 1496-1503.

Blauwkamp, T. A., Nigam, S., Ardehali, R., Weissman, I. L. and Nusse, R. (2012). Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nat Commun, 3*, 1070.

Bonnerot, C., Rocancourt, D., Briand, P., Grimber, G. and Nicolas, J. F. (1987). A Beta-Galactosidase Hybrid Protein Targeted to Nuclei as a Marker for Developmental Studies. *Proceedings of the National Academy of Sciences of the United States of America, 84*(19), 6795-6799.

Boroviak, T., Loos, R., Bertone, P., Smith, A. and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol, 16*(6), 516-528.

Boulet, A. M. and Capecchi, M. R. (2012). Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. *Dev Biol, 371*(2), 235-245.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature, 309*(5965), 255-256.

Braun, R. E., Lo, D., Pinkert, C. A., Widera, G., Flavell, R. A., Palmiter, R. D. and Brinster, R. L. (1990). Infertility in male transgenic mice: disruption of sperm development by HSV-tk expression in postmeiotic germ cells. *Biol Reprod, 43*(4), 684-693.

Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. and Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature, 448*(7150), 191-195.

Brown, J. M. and Storey, K. G. (2000). A region of the vertebrate neural plate in which neighbouring cells can adopt neural or epidermal fates. *Current Biology, 10*(14), 869-872.

Burtscher, I. and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development, 136*(6), 1029-1038.

Bylund, M., Andersson, E., Novitch, B. G. and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci, 6*(11), 1162-1168.

Cajal, M., Lawson, K. A., Hill, B., Moreau, A., Rao, J., Ross, A., Collignon, J. and Camus, A. (2012). Clonal and molecular analysis of the prospective anterior neural boundary in the mouse embryo. *Development, 139*(2), 423-436.

Cambray, N. and Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development, 129*(20), 4855-4866.

Cambray, N. and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development, 134*(15), 2829-2840.

Camus, A., Perea-Gomez, A., Moreau, A. and Collignon, J. (2006). Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev Biol, 295*(2), 743-755.

Canham, M. A., Sharov, A. A., Ko, M. S. and Brickman, J. M. (2010). Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol, 8*(5), e1000379.

Casarosa, S., Fode, C. and Guillemot, F. (1999). Mash1 regulates neurogenesis in the ventral telencephalon. *Development, 126*(3), 525-534.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature, 450*(7173), 1230-1234.

Chambers, I. and Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. *Development, 136*(14), 2311-2322.

Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell, 10*(5), 615-624.

Chenoweth, J. G., McKay, R. D. and Tesar, P. J. (2010). Epiblast stem cells contribute new insight into pluripotency and gastrulation. *Dev Growth Differ, 52*(3), 293-301.

Chung, S., Shin, B. S., Hedlund, E., Pruszak, J., Ferree, A., Kang, U. J., Isacson, O. and Kim, K. S. (2006). Genetic selection of sox1GFP-expressing neural precursors removes residual tumorigenic pluripotent stem cells and attenuates tumor formation after transplantation. *J Neurochem, 97*(5), 1467-1480.

Ciruna, B. and Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Developmental Cell, 1*(1), 37-49.

Clements, D., Taylor, H. C., Herrmann, B. G. and Stott, D. (1996). Distinct regulatory control of the Brachyury gene in axial and non-axial mesoderm suggests separation of mesoderm lineages early in mouse gastrulation. *Mech Dev, 56*(1-2), 139-149.

Cohen, M., Kicheva, A., Ribeiro, A., Blassberg, R., Page, K. M., Barnes, C. P. and Briscoe, J. (2015). Ptch1 and Gli regulate Shh signalling dynamics via multiple mechanisms. *Nat Commun, 6*, 6709.

Cong, L., Ran, F. A., Cox, D., Lin, S. L., Barretto, R., Habib, N., Hsu, P. D., Wu, X. B., Jiang, W. Y., Marraffini, L. A. and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819-823.

Conlon, F. L., Lyons, K. M., Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B. and Robertson, E. J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development, 120*(7), 1919-1928.

Copp, A. J., Brook, F. A., Estibeiro, J. P., Shum, A. S. and Cockroft, D. L. (1990). The embryonic development of mammalian neural tube defects. *Prog Neurobiol, 35*(5), 363-403.

Crossley, P. H. and Martin, G. R. (1995). The Mouse Fgf8 Gene Encodes a Family of Polypeptides and Is Expressed in Regions That Direct Outgrowth and Patterning in the Developing Embryo. *Development, 121*(2), 439-451.

Dasen, J. S., Liu, J. P. and Jessell, T. M. (2003). Motor neuron columnar fate imposed by sequential phases of Hox-c activity. *Nature, 425*(6961), 926-933.

Davis, R. L. and Kirschner, M. W. (2000). The fate of cells in the tailbud of Xenopus laevis. *Development, 127*(2), 255-267.

De Los Angeles, A., Loh, Y. H., Tesar, P. J. and Daley, G. Q. (2012). Accessing naive human pluripotency. *Curr Opin Genet Dev, 22*(3), 272-282.

del Barco Barrantes, I., Davidson, G., Grone, H. J., Westphal, H. and Niehrs, C. (2003). Dkk1 and noggin cooperate in mammalian head induction. *Genes Dev, 17*(18), 2239-2244.

Deng, C. X., Wynshawboris, A., Shen, M. M., Daugherty, C., Ornitz, D. M. and Leder, P. (1994). Murine Fgfr-1 Is Required for Early Postimplantation Growth and Axial Organization. *Genes & Development, 8*(24), 3045-3057.

Denham, M., Hasegawa, K., Menheniott, T., Rollo, B., Zhang, D., Hough, S., Alshawaf, A., Febbraro, F., Ighaniyan, S., Leung, J., Elliott, D. A., Newgreen, D. F., Pera, M. F. and Dottori, M. (2015). Multipotent caudal neural progenitors derived from human pluripotent stem cells that give rise to lineages of the central and peripheral nervous system. *Stem Cells, 33*(6), 1759-1770.

Deschamps, J. and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development, 132*(13), 2931-2942.

Dessaud, E., McMahon, A. P. and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development, 135*(15), 2489-2503.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P. W., Nagy, A. and van der Kooy, D. (2013). Oct4 is required ~E7.5 for proliferation in the primitive streak. *PLoS Genet, 9*(11), e1003957.

Diez del Corral, R. and Storey, K. G. (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *Bioessays, 26*(8), 857-869.

Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. and Jenkins, N. A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature, 436*(7048), 221-226.

Echelard, Y., Epstein, D. J., Stjacques, B., Shen, L., Mohler, J., Mcmahon, J. A. and Mcmahon, A. P. (1993). Sonic-Hedgehog, a Member of a Family of Putative Signaling Molecules, Is Implicated in the Regulation of Cns Polarity. *Cell, 75*(7), 1417-1430.

Economou, C., Tsakiridis, A., Wymeersch, F. J., Gordon-Keylock, S., Dewhurst, R. E., Fisher, D., Medvinsky, A., Smith, A. J. and Wilson, V. (2015). Intrinsic

factors and the embryonic environment influence the formation of extragonadal teratomas during gestation. *BMC Dev Biol, 15*, 35.

Ellis, P., Fagan, B. M., Magness, S. T., Hutton, S., Taranova, O., Hayashi, S., McMahon, A., Rao, M. and Pevny, L. (2004). SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev Neurosci, 26*(2-4), 148-165.

Etheridge, S. L., Ray, S., Li, S., Hamblet, N. S., Lijam, N., Tsang, M., Greer, J., Kardos, N., Wang, J., Sussman, D. J., Chen, P. and Wynshaw-Boris, A. (2008). Murine dishevelled 3 functions in redundant pathways with dishevelled 1 and 2 in normal cardiac outflow tract, cochlea, and neural tube development. *PLoS Genet, 4*(11), e1000259.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature, 292*(5819), 154-156.

Fedon, Y., Bonnieu, A., Gay, S., Vernus, B., Bacou, F. and Bernardi, H. (2012). Role and Function of Wnts in the Regulation of Myogenesis: When Wnt Meets Myostatin. *Skeletal Muscle - From Myogenesis to Clinical Relations, Dr. Julianna Cseri (Ed.), InTech, DOI: 10.5772/48330.*

Fehling, H. J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G. and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development, 130*(17), 4217-4227.

Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. and Goldfarb, M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science, 267*(5195), 246-249.

Ferrer-Vaquer, A., Piliszek, A., Tian, G., Aho, R. J., Dufort, D. and Hadjantonakis, A. K. (2010). A sensitive and bright single-cell resolution live imaging reporter of Wnt/ss-catenin signaling in the mouse. *BMC Dev Biol, 10*, 121.

Forlani, S., Lawson, K. A. and Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development, 130*(16), 3807-3819.

Gadue, P., Huber, T. L., Paddison, P. J. and Keller, G. M. (2006). Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc Natl Acad Sci U S A, 103*(45), 16806-16811.

Gagliardi, A., Mullin, N. P., Ying Tan, Z., Colby, D., Kousa, A. I., Halbritter, F., Weiss, J. T., Felker, A., Bezstarosti, K., Favaro, R., Demmers, J., Nicolis, S. K., Tomlinson, S. R., Poot, R. A. and Chambers, I. (2013). A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J, 32*(16), 2231-2247.

Gao, B., Song, H., Bishop, K., Elliot, G., Garrett, L., English, M. A., Andre, P., Robinson, J., Sood, R., Minami, Y., Economides, A. N. and Yang, Y. Z. (2011). Wnt Signaling Gradients Establish Planar Cell Polarity by Inducing Vangl2 Phosphorylation through Ror2. *Developmental Cell, 20*(2), 163-176.

Gardner, R. L. (1998). Contributions of blastocyst micromanipulation to the study of mammalian development. *Bioessays, 20*(2), 168-180.

Garriock, R. J., Chalamalasetty, R. B., Kennedy, M. W., Canizales, L. C., Lewandoski, M. and Yamaguchi, T. P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development, 142*(9), 1628-1638.

Goetz, R. and Mohammadi, M. (2013). Exploring mechanisms of FGF signalling through the lens of structural biology. *Nat Rev Mol Cell Biol, 14*(3), 166-180.

Gouti, M., Tsakiridis, A., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biol, 12*(8), e1001937.

Graf, T. and Stadtfeld, M. (2008). Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell, 3*(5), 480-483.

Graham, V., Khudyakov, J., Ellis, P. and Pevny, L. (2003). SOX2 Functions to Maintain Neural Progenitor Identity. *Neuron, 39*(5), 749-765.

Greber, B., Wu, G., Bernemann, C., Joo, J. Y., Han, D. W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P. and Scholer, H. R. (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell, 6*(3), 215-226.

Greco, T. L., Takada, S., Newhouse, M. M., McMahon, T. A., McMahon, A. P. and Camper, S. A. (1996). Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes & Development, 10*(3), 313-324.

Guillemot, F. and Joyner, A. L. (1993). Dynamic expression of the murine Achaete-Scute homologue Mash-1 in the developing nervous system. *Mech Dev, 42*(3), 171-185.

Guo, G. and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development, 137*(19), 3185-3192.

Guo, G., Yang, J., Nichols, J., Hall, J. S., Eyres, I., Mansfield, W. and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development, 136*(7), 1063-1069.

Hamblet, N. S., Lijam, N., Ruiz-Lozano, P., Wang, J. B., Yang, Y. S., Luo, Z. G., Mei, L., Chien, K. R., Sussman, D. J. and Wynshaw-Boris, A. (2002). Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. *Development, 129*(24), 5827-5838.

Han, D. W., Tapia, N., Joo, J. Y., Greber, B., Arauzo-Bravo, M. J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J. T. and Scholer, H. R. (2010). Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell, 143*(4), 617-627.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A. W., Cassady, J. P., Staerk, J., Carey, B. W., Lengner, C. J., Foreman, R., Love, J., Gao, Q., Kim, J. and Jaenisch, R. (2009). Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell, 4*(6), 513-524.

Hart, A. H., Hartley, L., Ibrahim, M. and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev Dyn, 230*(1), 187-198.

Hatano, S. Y., Tada, M., Kimura, H., Yamaguchi, S., Kono, T., Nakano, T., Suemori, H., Nakatsuji, N. and Tada, T. (2005). Pluripotential competence of cells associated with Nanog activity. *Mech Dev, 122*(1), 67-79.

Hayashi, K., Lopes, S. M., Tang, F. and Surani, M. A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell, 3*(4), 391-401.

Hayashi, K. and Surani, M. A. (2009). Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development, 136*(21), 3549-3556.

Henrique, D., Abranches, E., Verrier, L. and Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development, 142*(17), 2864-2875.

Huang, Y., Osorno, R., Tsakiridis, A. and Wilson, V. (2012). In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Rep, 2*(6), 1571-1578.

Inman, K. E. and Downs, K. M. (2006). Localization of Brachyury (T) in embryonic and extraembryonic tissues during mouse gastrulation. *Gene Expr Patterns, 6*(8), 783-793.

Iulianella, A., Beckett, B., Petkovich, M. and Lohnes, D. (1999). A molecular basis for retinoic acid-induced axial truncation. *Dev Biol, 205*(1), 33-48.

Iwafuchi-Doi, M., Matsuda, K., Murakami, K., Niwa, H., Tesar, P. J., Aruga, J., Matsuo, I. and Kondoh, H. (2012). Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development, 139*(21), 3926-3937.

Johnson, P. and Friedmann, T. (1990). Limited bidirectional activity of two housekeeping gene promoters: human HPRT and PGK. *Gene, 88*(2), 207-213.

Jurberg, A. D., Aires, R., Novoa, A., Rowland, J. E. and Mallo, M. (2014). Compartment-dependent activities of Wnt3a/beta-catenin signaling during vertebrate axial extension. *Dev Biol, 394*(2), 253-263.

Kamachi, Y., Iwafuchi, M., Okuda, Y., Takemoto, T., Uchikawa, M. and Kondoh, H. (2009). Evolution of non-coding regulatory sequences involved in the developmental process: Reflection of differential employment of paralogous genes as highlighted by Sox2 and group B1 Sox genes. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences, 85*(2), 55-68.

Kanda, T., Sullivan, K. F. and Wahl, G. M. (1998). Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr Biol, 8*(7), 377-385.

Kang, M., Piliszek, A., Artus, J. and Hadjantonakis, A. K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development, 140*(2), 267-279.

Kanki, J. P. and Ho, R. K. (1997). The development of the posterior body in zebrafish. *Development, 124*(4), 881-893.

Kimmel, C. B., Warga, R. M. and Schilling, T. F. (1990). Origin and organization of the zebrafish fate map. *Development, 108*(4), 581-594.

Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A. K., Nagy, A. and Tam, P. P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development, 126*(21), 4691-4701.

Kinder, S. J., Tsang, T. E., Wakamiya, M., Sasaki, H., Behringer, R. R., Nagy, A. and Tam, P. P. L. (2001). The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. *Development, 128*(18), 3623-3634.

Kintaka, R., Makanae, K. and Moriya, H. (2016). Cellular growth defects triggered by an overload of protein localization processes. *Sci Rep, 6*, 31774.

Kispert, A. and Herrmann, B. G. (1994). Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev Biol, 161*(1), 179-193.

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A., Jones, V., Hor, A., de Alencastro, G., Logan, G. J., Teber, E. T., Tam, O. H., Stutz, M. D., Alexander, I. E., Pickett, H. A. and Tam, P. P. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell, 14*(1), 107-120.

Komiya, Y. and Habas, R. (2008). Wnt signal transduction pathways. *Organogenesis, 4*(2), 68-75.

Kubo, A., Shinozaki, K., Shannon, J. M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H. J. and Keller, G. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development, 131*(7), 1651-1662.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S. and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development, 134*(16), 2895-2902.

Kurek, D., Neagu, A., Tastemel, M., Tuysuz, N., Lehmann, J., van de Werken, H. J., Philipsen, S., van der Linden, R., Maas, A., van, I. W. F., Drukker, M. and ten Berge, D. (2015). Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports, 4*(1), 114-128.

Lawson, K. A., Meneses, J. J. and Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development, 113*(3), 891-911.

Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N. and Weintraub, H. (1995). Conversion of Xenopus Ectoderm into Neurons by Neurod, a Basic Helix-Loop-Helix Protein. *Science, 268*(5212), 836-844.

Lippmann, E. S., Williams, C. E., Ruhl, D. A., Estevez-Silva, M. C., Chapman, E. R., Coon, J. J. and Ashton, R. S. (2015). Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. *Stem Cell Reports, 4*(4), 632-644.

Liu, J. P., Laufer, E. and Jessell, T. M. (2001). Assigning the positional identity of spinal motor neurons: rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids. *Neuron, 32*(6), 997-1012.

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R. and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat Genet, 22*(4), 361-365.

Lu, X. W., Borchers, A. G. M., Jolicoeur, C., Rayburn, H., Baker, J. C. and Tessier-Lavigne, M. (2004). PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. *Nature, 430*(6995), 93-98.

Mallo, M., Wellik, D. M. and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. *Dev Biol, 344*(1), 7-15.

Martello, G. and Smith, A. (2014). The nature of embryonic stem cells. *Annu Rev Cell Dev Biol, 30*, 647-675.

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

Martin, B. L. and Kimelman, D. (2012). Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev Cell, 22*(1), 223-232.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., Ko, M. S. and Niwa, H. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol, 9*(6), 625-635.

Mathis, L. and Nicolas, J. F. (2000). Different clonal dispersion in the rostral and caudal mouse central nervous system. *Development, 127*(6), 1277-1290.

Matsuda, K. and Kondoh, H. (2014). Dkk1-dependent inhibition of Wnt signaling activates Hesx1 expression through its 5' enhancer and directs forebrain precursor development. *Genes Cells, 19*(5), 374-385.

Mazzoni, E. O., Mahony, S., Peljto, M., Patel, T., Thornton, S. R., McCuine, S., Reeder, C., Boyer, L. A., Young, R. A., Gifford, D. K. and Wichterle, H. (2013). Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci, 16*(9), 1191-1198.

McGrew, M. J., Sherman, A., Lillico, S. G., Ellard, F. M., Radcliffe, P. A., Gilhooley, H. J., Mitrophanous, K. A., Cambray, N., Wilson, V. and Sang, H. (2008). Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. *Development, 135*(13), 2289-2299.

McMahon, J. A., Takada, S., Zimmerman, L. B., Fan, C. M., Harland, R. M. and McMahon, A. P. (1998). Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes & Development, 12*(10), 1438-1452.

Melby, A. E., Warga, R. M. and Kimmel, C. B. (1996). Specification of cell fates at the dorsal margin of the zebrafish gastrula. *Development, 122*(7), 2225-2237.

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., Dorward, D. W., Glinka, A., Grinberg, A., Huang, S. P., Niehrs, C., Belmonte, J. C. I. and Westphal, H. (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. *Developmental Cell, 1*(3), 423-434.

Naiche, L. A., Holder, N. and Lewandoski, M. (2011). FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proceedings of the National Academy of Sciences of the United States of America, 108*(10), 4018-4023.

Najm, F. J., Chenoweth, J. G., Anderson, P. D., Nadeau, J. H., Redline, R. W., McKay, R. D. and Tesar, P. J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell, 8*(3), 318-325.

Nichols, J. and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell, 4*(6), 487-492.

Nicolas, J. F., Mathis, L., Bonnerot, C. and Saurin, W. (1996). Evidence in the mouse for self-renewing stem cells in the formation of a segmented longitudinal structure, the myotome. *Development, 122*(9), 2933-2946.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R. O., Ogonuki, N., Makita, R., Kurihara, H., Morin-Kensicki, E. M., Nojima, H., Rossant, J., Nakao, K., Niwa, H. and Sasaki, H. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev Cell, 16*(3), 398-410.

Niswander, L. and Martin, G. R. (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development, 114*(3), 755-768.

Nordstrom, U., Maier, E., Jessell, T. M. and Edlund, T. (2006). An early role for WNT signaling in specifying neural patterns of Cdx and Hox gene expression and motor neuron subtype identity. *PLoS Biol, 4*(8), e252.

Olenych, S. G., Claxton, N. S., Ottenberg, G. K. and Davidson, M. W. (2007). The fluorescent protein color palette. *Curr Protoc Cell Biol, Chapter 21*, Unit 21 25.

Olivera-Martinez, I., Harada, H., Halley, P. A. and Storey, K. G. (2012). Loss of FGF-dependent mesoderm identity and rise of endogenous retinoid signalling determine cessation of body axis elongation. *PLoS Biol, 10*(10), e1001415.

Onishi, K. and Zandstra, P. W. (2015). LIF signaling in stem cells and development. *Development, 142*(13), 2230-2236.

Osorno, R. and Chambers, I. (2011). Transcription factor heterogeneity and epiblast pluripotency. *Philos Trans R Soc Lond B Biol Sci, 366*(1575), 2230-2237.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P. J., Chambers, I. and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development, 139*(13), 2288-2298.

Perea-Gomez, A., Vella, F. D., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H., Behringer, R. R. and Ang, S. L. (2002). Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Dev Cell, 3*(5), 745-756.

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

Pevny, L. H., Sockanathan, S., Placzek, M. and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. *Development, 125*(10), 1967-1978.

Piccolo, S., Agius, E., Leyns, L., Bhattacharyya, S., Grunz, H., Bouwmeester, T. and De Robertis, E. M. (1999). The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature, 397*(6721), 707-710.

Pieters, T. and van Roy, F. (2014). Role of cell-cell adhesion complexes in embryonic stem cell biology. *J Cell Sci, 127*(Pt 12), 2603-2613.

Radice, G. L., Rayburn, H., Matsunami, H., Knudsen, K. A., Takeichi, M. and Hynes, R. O. (1997). Developmental defects in mouse embryos lacking N-cadherin. *Dev Biol, 181*(1), 64-78.

Ralston, A., Cox, B. J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J. S. and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development, 137*(3), 395-403.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc, 8*(11), 2281-2308.

Rhinn, M., Lun, K., Werner, M., Simeone, A. and Brand, M. (2004). Isolation and expression of the homeobox gene Gbx1 during mouse development. *Dev Dyn, 229*(2), 334-339.

Robb, L., Hartley, L., Begley, C. G., Brodnicki, T. C., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. and Elefanty, A. G. (2000). Cloning, expression analysis, and chromosomal localization of murine and human homologues of a Xenopus mix gene. *Dev Dyn, 219*(4), 497-504.

Rossant, J. (2008). Stem cells and early lineage development. *Cell, 132*(4), 527-531.

Rossant, J., Chazaud, C. and Yamanaka, Y. (2003). Lineage allocation and asymmetries in the early mouse embryo. *Philos Trans R Soc Lond B Biol Sci, 358*(1436), 1341-1348; discussion 1349.

Saga, Y., Hata, N., Kobayashi, S., Magnuson, T., Seldin, M. F. and Taketo, M. M. (1996). MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development, 122*(9), 2769-2778.

Saiz, N. and Plusa, B. (2013). Early cell fate decisions in the mouse embryo. *Reproduction, 145*(3), R65-80.

Sasaki, H. and Hogan, B. L. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development, 118*(1), 47-59.

Schier, A. F. (2003). Nodal signaling in vertebrate development. *Annu Rev Cell Dev Biol, 19*, 589-621.

Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J, 9*(7), 2185-2195.

Schubert, F. R., Fainsod, A., Gruenbaum, Y. and Gruss, P. (1995). Expression of the novel murine homeobox gene Sax-1 in the developing nervous system. *Mech Dev, 51*(1), 99-114.

Selleck, M. A. and Stern, C. D. (1991). Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development, 112*(2), 615-626.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G. and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science, 343*(6166), 84-87.

Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E. and Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol, 22*(12), 1567-1572.

Shawlot, W. and Behringer, R. R. (1995). Requirement for Lim1 in head-organizer function. *Nature, 374*(6521), 425-430.

Shawlot, W., Deng, J. M. and Behringer, R. R. (1998). Expression of the mouse cerberus-related gene, Cerr1, suggests a role in anterior neural induction and somitogenesis. *Proc Natl Acad Sci U S A, 95*(11), 6198-6203.

Shi, Y. and Massagué, J. (2003). Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. *Cell, 113*(6), 685-700.

Shum, A. S., Poon, L. L., Tang, W. W., Koide, T., Chan, B. W., Leung, Y. C., Shiroishi, T. and Copp, A. J. (1999). Retinoic acid induces down-regulation of Wnt-3a, apoptosis and diversion of tail bud cells to a neural fate in the mouse embryo. *Mech Dev, 84*(1-2), 17-30.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. and Boncinelli, E. (1992). Nested expression domains of four homeobox genes in developing rostral brain. *Nature, 358*(6388), 687-690.

Slusarski, D. C., Yang-Snyder, J., Busa, W. B. and Moon, R. T. (1997). Modulation of embryonic intracellular Ca2+ signaling by Wnt-5A. *Dev Biol, 182*(1), 114-120.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature, 336*(6200), 688-690.

Smith, J. L., Gesteland, K. M. and Schoenwolf, G. C. (1994). Prospective Fate Map of the Mouse Primitive Streak at 7.5 Days of Gestation. *Developmental Dynamics, 201*(3), 279-289.

Song, H., Hu, J., Chen, W., Elliott, G., Andre, P., Gao, B. and Yang, Y. (2010). Planar cell polarity breaks bilateral symmetry by controlling ciliary positioning. *Nature, 466*(7304), 378-382.

Song, L., Chen, J., Peng, G., Tang, K. and Jing, N. (2016). Dynamic Heterogeneity of Brachyury in Mouse Epiblast Stem Cells Mediates Distinct Respond to Extrinsic BMP Signaling. *J Biol Chem*.

Souilhol, C., Gonneau, C., Lendinez, J. G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S., Antonchuk, J., Zhao, S. and Medvinsky, A. (2016). Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. *Nat Commun, 7*, 10784.

Stern, C. D. (2005). Neural induction: old problem, new findings, yet more questions. *Development, 132*(9), 2007-2021.

Sumi, T., Oki, S., Kitajima, K. and Meno, C. (2013). Epiblast ground state is controlled by canonical Wnt/beta-catenin signaling in the postimplantation mouse embryo and epiblast stem cells. *PLoS One, 8*(5), e63378.

Sun, L. T., Yamaguchi, S., Hirano, K., Ichisaka, T., Kuroda, T. and Tada, T. (2014). Nanog co-regulated by Nodal/Smad2 and Oct4 is required for pluripotency in developing mouse epiblast. *Dev Biol, 392*(2), 182-192.

Sun, X., Meyers, E. N., Lewandoski, M. and Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev, 13*(14), 1834-1846.

Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F. and Vignali, D. A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat Biotechnol, 22*(5), 589-594.

Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A. and McMahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev, 8*(2), 174-189.

Takemoto, T., Uchikawa, M., Kamachi, Y. and Kondoh, H. (2006). Convergence of Wnt and FGF signals in the genesis of posterior neural plate through activation of the Sox2 enhancer N-1. *Development, 133*(2), 297-306.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E. and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature, 470*(7334), 394-398.

Tam, P. P. (1989). Regionalisation of the mouse embryonic ectoderm: allocation of prospective ectodermal tissues during gastrulation. *Development, 107*(1), 55-67.

Taranova, O. V., Magness, S. T., Fagan, B. M., Wu, Y. Q., Surzenko, N., Hutton, S. R. and Pevny, L. H. (2006). SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes & Development, 20*(9), 1187-1202.

Tarkowski, A. K. and Wróblewska, J. (1967). Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *Development 18*(1), 26.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature, 448*(7150), 196-199.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M. and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Research, 30*(14), 3202-3213.

Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S. and Wilson, V. (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development, 141*(6), 1209-1221.

Tsakiridis, A. and Wilson, V. (2015). Assessing the bipotency of in vitro-derived neuromesodermal progenitors. *F1000Res, 4*, 100.

Turner, D. A., Hayward, P. C., Baillie-Johnson, P., Rue, P., Broome, R., Faunes, F. and Martinez Arias, A. (2014). Wnt/beta-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development, 141*(22), 4243-4253.

Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. and Nicolas, J. F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev Cell, 17*(3), 365-376.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. and Kondoh, H. (2003). Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Developmental Cell, 4*(4), 509-519.

Uchikawa, M., Yoshida, M., Iwafuchi-Doi, M., Matsuda, K., Ishida, Y., Takemoto, T. and Kondoh, H. (2011). B1 and B2 Sox gene expression during neural plate

development in chicken and mouse embryos: universal versus species-dependent features. *Dev Growth Differ, 53*(6), 761-771.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L. E., Trotter, M. W., Cho, C. H., Martinez, A., Rugg-Gunn, P., Brons, G. and Pedersen, R. A. (2009a). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development, 136*(8), 1339-1349.

Vallier, L., Touboul, T., Chng, Z., Brimpari, M., Hannan, N., Millan, E., Smithers, L. E., Trotter, M., Rugg-Gunn, P., Weber, A. and Pedersen, R. A. (2009b). Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS One, 4*(6), e6082.

van Amerongen, R. and Berns, A. (2006). Knockout mouse models to study Wnt signal transduction. *Trends Genet, 22*(12), 678-689.

van de Ven, C., Bialecka, M., Neijts, R., Young, T., Rowland, J. E., Stringer, E. J., Van Rooijen, C., Meijlink, F., Novoa, A., Freund, J. N., Mallo, M., Beck, F. and Deschamps, J. (2011). Concerted involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone. *Development, 138*(16), 3451-3462.

van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B. I. and Deschamps, J. (2002). Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development, 129*(9), 2181-2193.

Vossen, R. H., Aten, E., Roos, A. and den Dunnen, J. T. (2009). High-resolution melting analysis (HRMA): more than just sequence variant screening. *Hum Mutat, 30*(6), 860-866.

Walther, C. and Gruss, P. (1991). Pax-6, a Murine Paired Box Gene, Is Expressed in the Developing Cns. *Development, 113*(4), 1435-1449.

Wang, Y., Wang, F., Wang, R., Zhao, P. and Xia, Q. (2015). 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Scientific Reports, 5*, 16273.

Wellik, D. M. and Capecchi, M. R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science, 301*(5631), 363-367.

Wilkinson, D. G., Bhatt, S. and Herrmann, B. G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature, 343*(6259), 657-659.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature, 336*(6200), 684-687.

Wilson, V. and Beddington, R. (1997). Expression of T protein in the primitive streak is necessary and sufficient for posterior mesoderm movement and somite differentiation. *Developmental Biology, 192*(1), 45-58.

Wilson, V. and Beddington, R. S. (1996). Cell fate and morphogenetic movement in the late mouse primitive streak. *Mech Dev, 55*(1), 79-89.

Wilson, V., Manson, L., Skarnes, W. C. and Beddington, R. S. P. (1995). The T-Gene Is Necessary for Normal Mesodermal Morphogenetic Cell Movements during Gastrulation. *Development, 121*(3), 877-886.

Wilson, V., Olivera-Martinez, I. and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development, 136*(10), 1591-1604.

Winnier, G., Blessing, M., Labosky, P. A. and Hogan, B. L. M. (1995). Bone Morphogenetic Protein-4 Is Required for Mesoderm Formation and Patterning in the Mouse. *Genes & Development, 9*(17), 2105-2116.

Winzi, M. K., Hyttel, P., Dale, J. K. and Serup, P. (2011). Isolation and characterization of node/notochord-like cells from mouse embryonic stem cells. *Stem Cells Dev, 20*(11), 1817-1827.

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

Wymeersch, F. J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F. C. and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife, 5*.

Yamaguchi, T. P., Harpal, K., Henkemeyer, M. and Rossant, J. (1994). Fgfr-1 Is Required for Embryonic Growth and Mesodermal Patterning during Mouse Gastrulation. *Genes & Development, 8*(24), 3032-3044.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. Y. and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes & Development, 13*(24), 3185-3190.

Ybot-Gonzalez, P., Savery, D., Gerrelli, D., Signore, M., Mitchell, C. E., Faux, C. H., Greene, N. D. and Copp, A. J. (2007). Convergent extension, planar-cell-polarity signalling and initiation of mouse neural tube closure. *Development, 134*(4), 789-799.

Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Scholer, H. R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development, 122*(3), 881-894.

Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol, 21*(2), 183-186.

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature, 453*(7194), 519-523.

Yoshikawa, Y., Fujimori, T., McMahon, A. P. and Takada, S. (1997). Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Developmental Biology, 183*(2), 234-242.

Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J. N., Beck, F., Mallo, M. and Deschamps, J. (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell, 17*(4), 516-526.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y. and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature, 509*(7501), 487-491.

Appendix

CRISPR | GENOME ENGINEERING TOOLBOX

ZhangLab

Target Sequence Cloning Protocol

(standard de-salted oligos are sufficient)

pX260 (or pX334) – hSpCas9 (or hSpas9n nickase) + CRISPR array + tracrRNA:

In order to clone the target sequence into the pX260 backbone, synthesize two oligos of the form:

```
5′ – AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT     – 3′
3′ –     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAAT – 5′
```

pX330 (or pX335) – hSpCas9 (or hSpCas9n nickase) + chimeric guideRNA:

In order to clone the target sequence into the pX260 backbone, synthesize two oligos of the form:

```
5′ – CACCGNNNNNNNNNNNNNNNNNNN     – 3′
3′ –     CNNNNNNNNNNNNNNNNNNNCAAA – 5′
```

* * * * *

Oligo annealing and cloning into backbone vectors:

1. Digest 1ug of pX260 or pX330 with *Bbs*I for 30 min at 37C:

1 ug	pX260 or pX330
1 ul	FastDigest *Bbs*I (Fermentas)
1 ul	FastAP (Fermentas)
2 ul	10X FastDigest Buffer
X ul	ddH_2O
20 ul	total

2. Gel purify digested pX260 or pX330 using QIAquick Gel Extraction Kit and elute in EB.

3. Phosphorylate and anneal each pair of oligos:

1 ul	oligo 1 (100mM)
1 ul	oligo 2 (100mM)
1 ul	10X T4 Ligation Buffer (NEB)
6.5 ul	ddH_2O
0.5 ul	T4 PNK (NEB)
10 ul	total

Anneal in a thermocycler using the following parameters:

37°C	30 min
95°C	5 min and then ramp down to 25°C at 5°C/min

4. Set up ligation reaction and incubate at room temperature for 10 min:

X ul	*Bbs*I digested pX260 or pX330 from **step 2** (50ng)
1 ul	phosphorylated and annealed oligo duplex from **step 3** (1:200 dilution)
5 ul	2X Quickligation Buffer (NEB)
X ul	ddH_2O
10 ul	subtotal
1 ul	Quick Ligase (NEB)
11 ul	total

5. (optional but highly recommended) Treat ligation reaction with PlasmidSafe exonuclease to prevent unwanted recombination products:

11 ul	ligation reaction from **step 4**
1.5 ul	10X PlasmidSafe Buffer
1.5 ul	10mM ATP
1 ul	ddH_2O
15 ul	total

Incubate reaction at 37C for 30 min.

6. Transformation

v20130112

Figure S2.1: Protocol for cloning sgRNA sequences into CRISPR/Cas9 vectors. Available online from ZhangLab, Genome Engineering Resources.

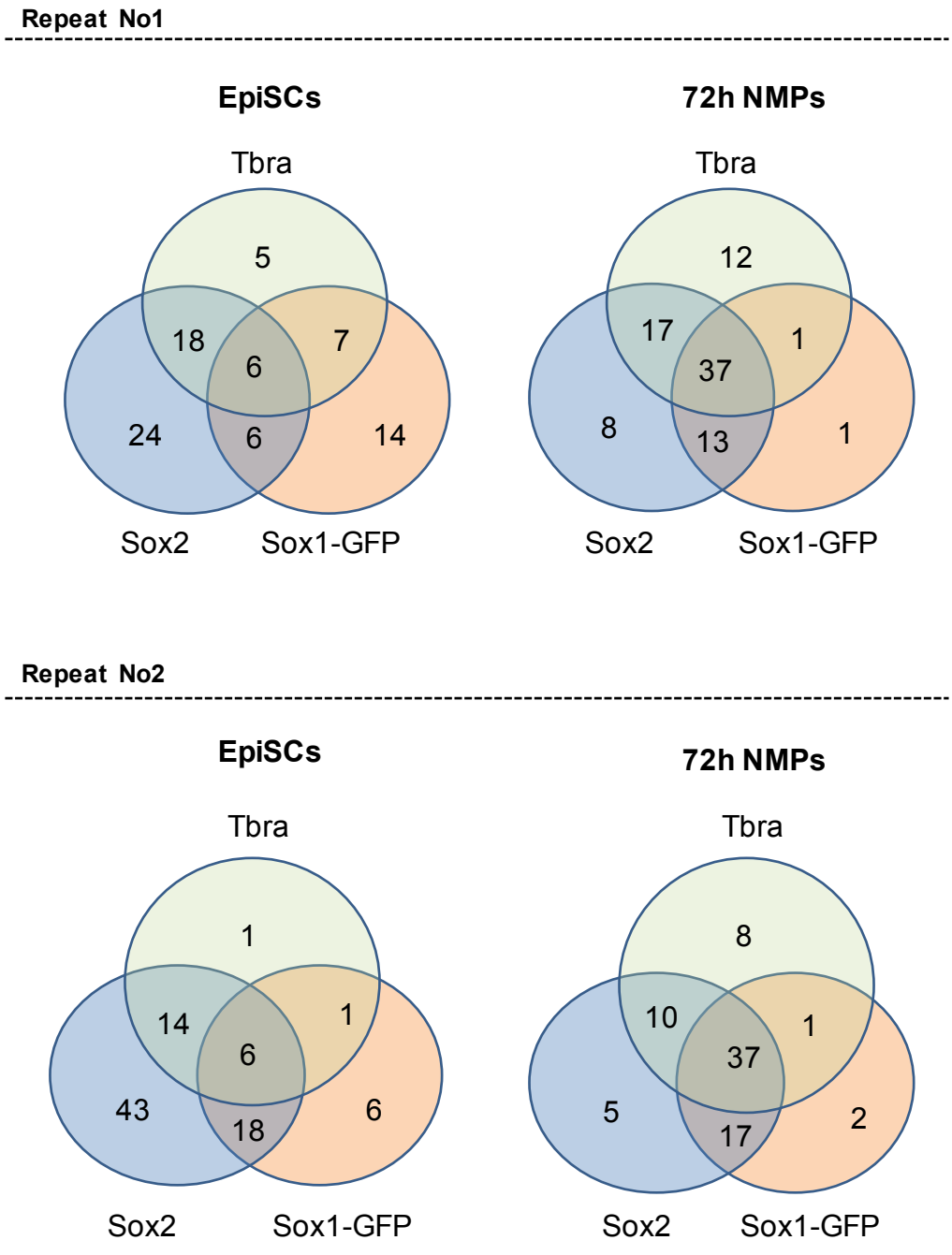
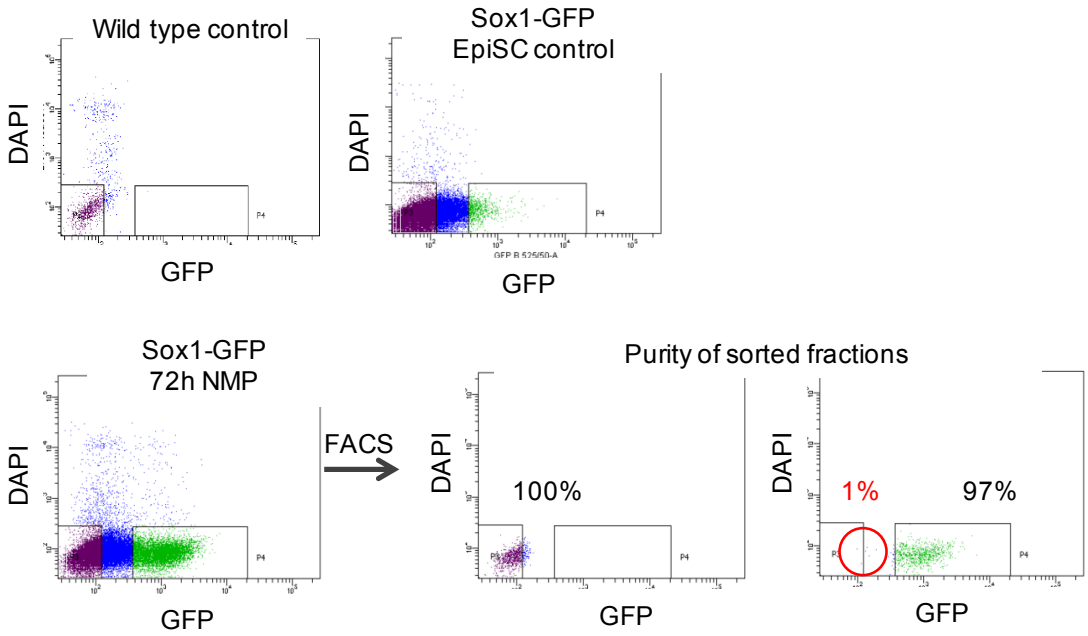


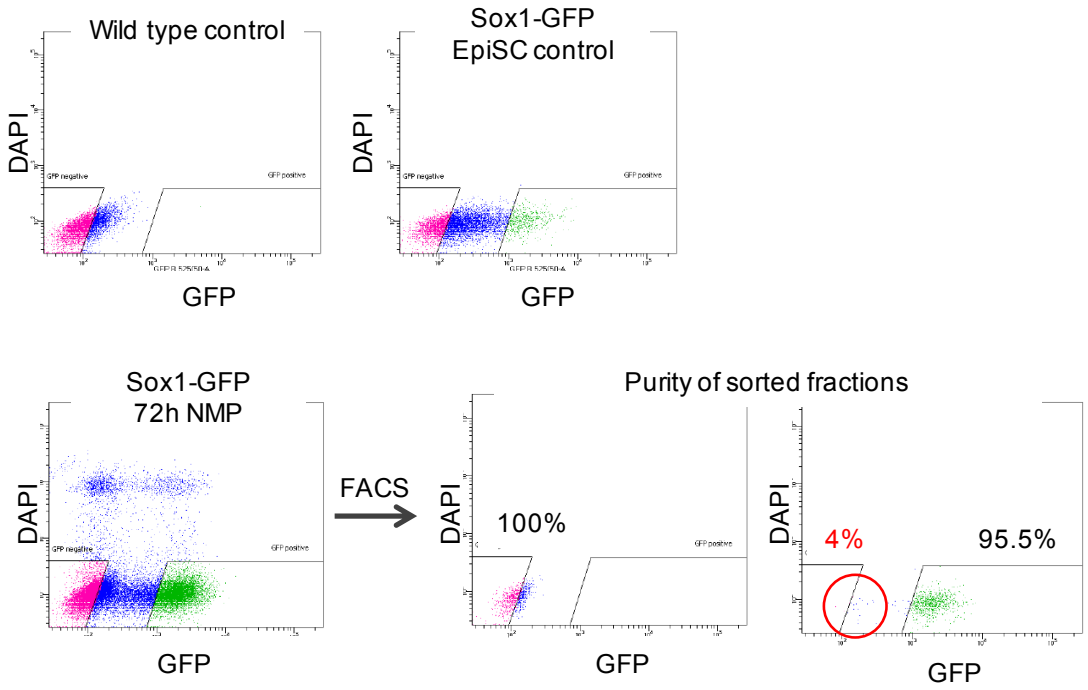
Figure S4.1: Sox1-GFP, Sox2 and Tbra expression in EpiSC and NMP cultures. Quantification following single cell image analysis of triple immunocytochemistry for Sox1-GFP, Sox2 and Tbra in EpiSC and 72h NMP cultures. The Venn diagrams present the percentage of cells that express any combination of the three proteins. The experiment was performed twice. The analysis of each replicate is shown separately.

A

Repeat No1



Repeat No2



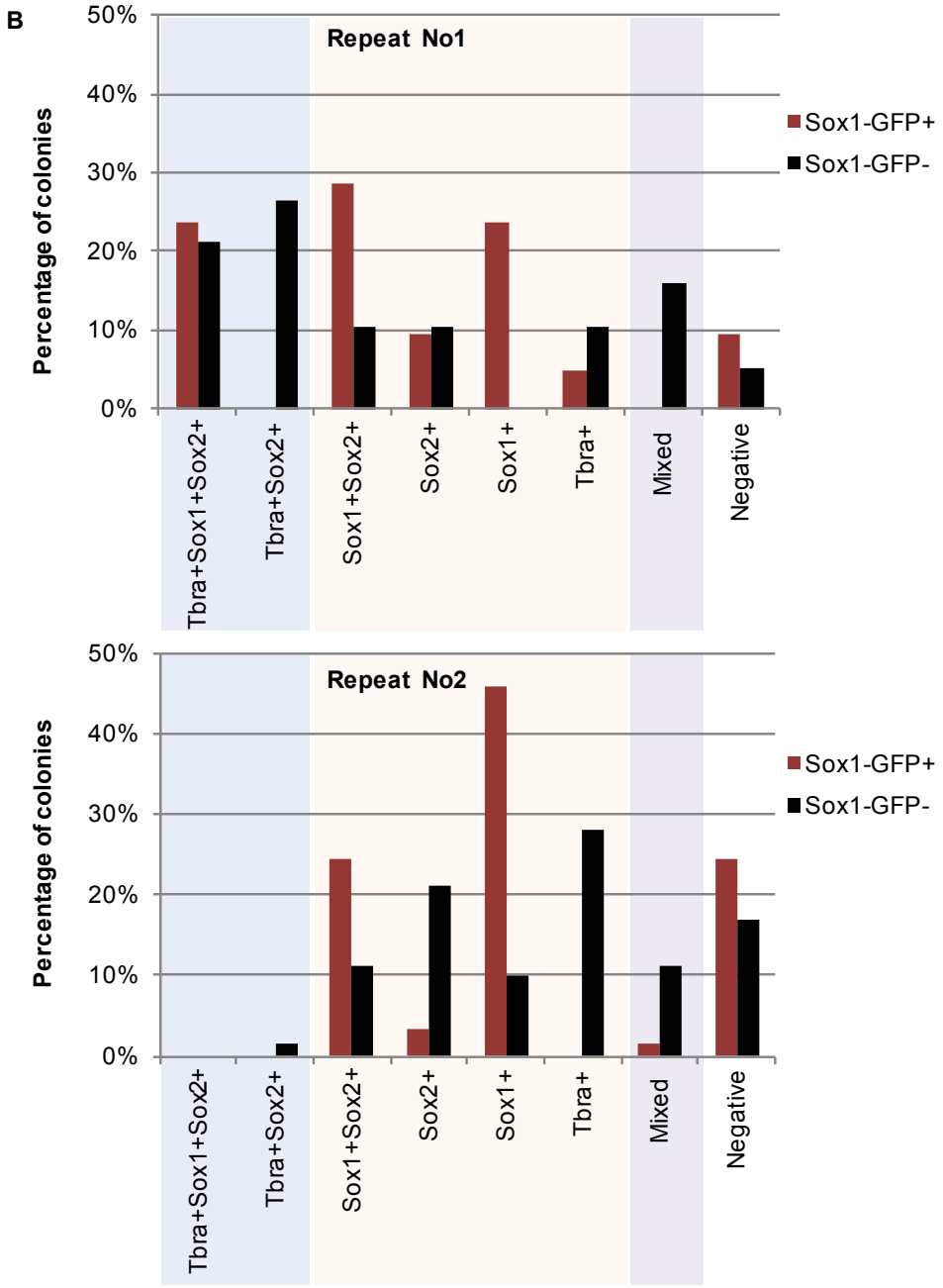
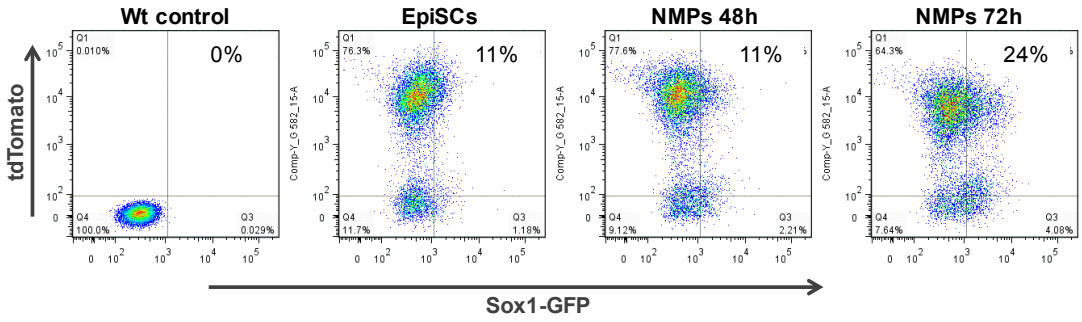


Figure S4.2: Lineage specification of sorted Sox1-GFP positive and negative cells under prolonged culture in FGF/CHI; analysis of individual replicates. A) FACS sorting of 72h NMP cultures into a Sox1-GFP positive and negative fraction. The purity of the sorted subpopulations was tested; the red circles on the plots of Sox1-GFP positive fractions point out a small percentage of contaminating GFP negative cells. **B)** Percentages of the different types of colonies obtained after clonal plating of sorted Sox1-GFP^{+} and Sox1-GFP^{-} cells. Repeat No1: N=20. Repeat No2: N=60 (GFP^{+}) and 70 (GFP^{-}).

A



B

Grafted cells	Embryo ID	Embryo morphology	Contribution		Length of contribution (somite number)	Total Score
			PXM	NT		
48h NMPs	T1E1	Aberrant tail	-	-	-	**4/5**
	T1E2	Normal	+	+	15, 22-TB	
	T2E1	Aberrant tail	+	-	24-TB	
	T2E2	Aberrant tail	+	+*	17-TB	
	T2E3	Normal	+	+	7-21	
72h NMPs	T1E1	Normal	-	-	-	**1/5**
	T1E2	Normal	-	-	-	
	T2E1	Normal	-	-	-	
	T2E2	Abnormal	+	+	13-TB	
	T3E1	Normal	-	-	-	

*also ectopic tube ventral of NT

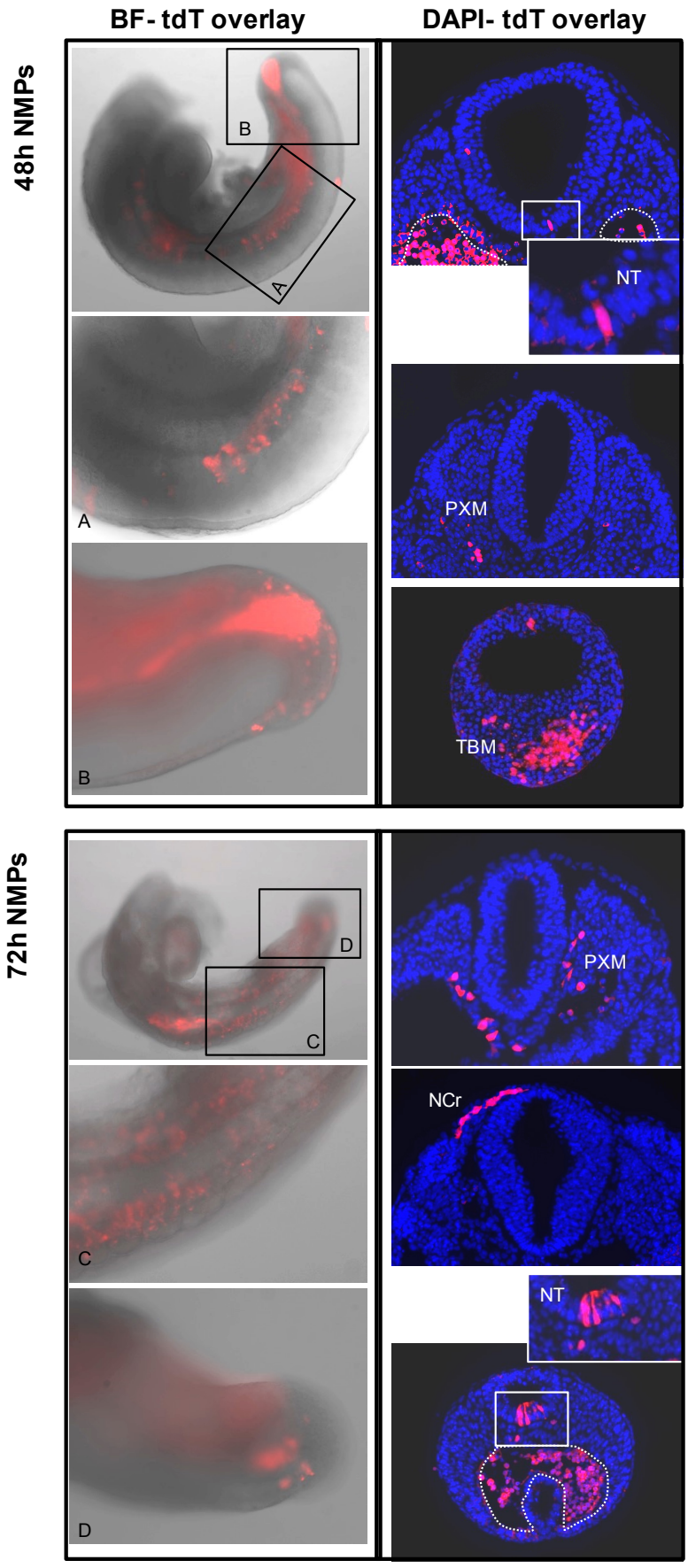

BF- tdT overlay
DAPI- tdT overlay
48h NMPs
72h NMPs
A
B
C
D
NT
PXM
TBM
NCr

Figure S4.3: Grafting of 48h and 72h *in vitro* NMPs in the NSB of E8.5 mouse embryos. Sox1-GFP EpiSCs were transfected with a plasmid containing a CAG-tdT-Puro cassette and expanded in the presence of puromycin, prior to differentiation into 48h and 72h NMPs. **A)** FACS analysis of tdTomato and GFP in EpiSC, 48h and 72h NMP cultures. **B)** Cells from 48h and 72h NMP cultures were transplanted in the NSB of E8.5 mouse embryos. Following *ex vivo* culture of the hosts for 48 hours, the contribution of the grafted cells was assessed by fluorescence microscopy in cryostat sections of the E10.5 embryos. The cell transplantation and and the *ex vivo* culture of the host embryos were performed by Yali Huang. The table presents the summary of the incorporation of grafted 48h and 72h NMPs within host embryos. **C)** Representative examples of donor cell incorporation (red, tdTomato); GFP was not detected therefore photos from the green fluorescence channel are not shown. Circled areas constitute highly autofluorescent blood cells. Wt, wild type; PXM, paraxial mesoderm; NT, neural tube; TB, tail bud; TBM, tail bud mesoderm; Ncr, neural crest; BF, bright field; tdT, tdTomato.