Dissertation

submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Luigi Russo, M.Sc. in Medical Biotechnology
born in Vico Equense, Italy

Oral examination:
December 3rd, 2020

Molecular characterization of *in vitro* specification towards neural and primitive streak fates

Examiners:

Dr. James Hackett

Prof. Dr. Ana Martin-Villalba

This work has been carried out at the European Molecular Biology Laboratory in Heidelberg under the supervision of Dr. Pierre Neveu from September 2016 to September 2020.

Fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza.

Dante, Inferno,
canto XXVI, vv. 119-120.

TABLE OF CONTENTS

SUMMARY

Since their discovery, mouse embryonic stem cells (mESCs) constitute an invaluable system to investigate complex developmental biology processes, such as germ layer specification. In my PhD project, I used fluorescent reporter mESC lines to explore the signaling network underlying fate decisions from Primitive Streak (PS) formation to neuroectoderm induction. I first defined culture conditions allowing the generation of cells transcriptionally resembling the PS and relying on the same signaling as their *in vivo* counterparts. I successfully discriminated the individual contributions of the Wnt and Nodal/TGF-β pathways in this context, the first starting the PS program, and the second enabling to proceed towards the endoderm fate while repressing neuroectoderm formation. Using a Wnt reporter line, I identified cells with active canonical Wnt signaling during PS-like differentiation and characterized by a posterior PS and mesoderm transcriptome. I additionally showed that neural progenitors arise in these conditions, with identity spanning the entire anteroposterior axis, and I uncovered their putative inductive stimuli. A Retinoic Acid (RA) sensor line, generated ad hoc, showed active RA signaling in the PS-like population. Intriguingly, the fraction of neural progenitors obtained was dependent on the activation state of RA signaling, including RA accumulation by deleting its degrading enzyme, Cyp26a1. The combination of RA perturbation and Wnt or TGF-β inhibition, highlighted the involvement of RA signaling in other mechanisms of neural induction. Finally, I mimicked the phenotype of the RA synthesis-defective *Aldh1a2*$^{-/-}$ mice in my sensor line, observing no RA reporter expression in PS-like differentiation, and a seemingly unaffected specification of neuroectoderm cells. However, the neural fate acquisition in the *Aldh1a2*$^{-/-}$ cells was still sensitive to repression of RA receptors, and providing vitamin A led most of the cells towards the neural lineage, and reactivated the RA reporter. This proves a residual ability of *Aldh1a2*$^{-/-}$ cells to synthetize RA, undermining a central argument against a role of RA in neural induction by the PS.

In conclusion, in this project I established a manipulatable system, suitable for investigating processes going from the primitive streak formation to the acquisition of the neuroectoderm fate. The system provided convincing arguments about the necessity to reevaluate the role of retinoids in neuroectoderm specification by the PS.

ZUSAMMENFASSUNG

Seit Ihrer Entdeckung stellen Embryonale Stammzellen der Maus (mESCs) eines der wichtigsten Testsysteme in der Entwicklungsbiologie dar. Sie werden verwendet um komplexe Entwicklungsprozesse wie jene der Keimblatt Spezifizierung zu untersuchen. In vorliegender Dissertation wurden fluoreszierende reporter Zelllinien der Maus dazu verwendet, Zellkommunikationsnetzwerke die zur Spezifizierung des Primitivstreifens (PS) notwendig sind, und schliesslich zur Ausbildung des Neuroektoderms führen, zu untersuchen. In einem ersten Schritt wurden Kulturbedingungen entwickelt um Stammzellen reproduzierbar in PS ähnliche Zellen zu differenzieren. Diese Zellen zeigten das gleiche Genexpressionsprofil wie *in-vivo* Zellen des Primitivstreifens und unterliegen somit dem gleichen Zellkommunikationsnetzwerk. Ergebnisse dieser Arbeit zeigen Eindrucksvoll den Beitrag von Wnt und Nodal/TGF-β Signalkaskaden zur initialen Ausbildung des Primitivstreifens. Ferner konnte gezeigt werden, dass diese Signalwege notwendig sind um Stammzellen zu Zellen des Endoderm-Keimblatts zu differenzieren, während die Differenzierung zu Zellen des Neuroectoderm-Keimblatts unterdrückt werden. Nach Etablierung einer Wnt reporter Zellinie konnte gezeigt werden, dass PS-ähnliche Zellen intrinsisches Wnt signalling aufweisen. Zudem spiegeln diese Zellen das Transkriptom von posterior PS Zellen sowie Mesoderm Zellen wieder. Ferner konnte gezeigt werden, dass in vorliegenden Differenzierungskonditionen neuronale Vorläuferzellen entstehen, wessen Identität die gesamte Spannbreite der anteroposterior Achse abbildet. Interessanterweise konnten die dazu mutmasslich notwendigen Stimulatoren enthüllt werden. Unter Etablierung einer Retinoic Acid (RA) Reporterzelllinie konnten aktive RA-Signalkaskaden, vor allem in PS ähnlichen Zellpopulationen, nachgewiesen werden. Interessanterweise war dabei der Anteil an neuronalen Vorläuferzellen abhängig vom Status der RA-Signalkaskade. Dies involvierte die Eliminierung des RA abbauenden Enzyms Cyp26a1, wodurch die intrazelluläre RA Konzentration anstieg. Die Kombination aus dem Eingriff in die RA-Konzentration und der Inhibierung von Wnt oder TGF-β Signalkaskaden, verdeutlicht den Einfluss des RA-Signalwegs auf weiterreichende Mechanismen als jener der Neuronalen Induktion.

Schließlich konnte in einem Zellmodell der Phänotyp von Retinoic Acid defekten *Aldh1a2*$^{-/-}$ Mäusen nachgestellt werden. Dabei konnte die Abwesenheit der RA-reporter Expression bei sich differenzierenden PS-ähnlichen Zellen nachgewiesen werden. Zudem schien die Neuroektoderm Spezifikation unbeeinflusst. Trotz der Abwesenheit des *Aldh1a2* Gens im Neuronalen Keimblatt waren die Zellen sensitiv gegenüber der Unterdrückung des RA- Rezeptors. Ferner gelang durch Zugabe von Vitamin A die Differenzierung dieser vermeintlich defekten Stammzellen in Richtung des neuronalen Keimblattes und der RA-reporter in diesen Zellen wurde reaktiviert. Dieses Zellverhalten beweist eine Restaktivität von vermeintlich *Aldh1a2*$^{-/-}$ defekten Zellen, welche immer noch Retinoic Acid synthetisieren können. Dies Wiederspricht jedoch dem derzeitigen Stand der Wissenschaft nach welchem RA keine zentrale Rolle zur Induktion neuronaler Zellen durch den Primitivstreifen einnimmt.

In vorliegender Dissertation wurde ein manipulierbares System etabliert welches sich für die Untersuchung von Signalkaskaden und Prozessen in der Entwicklung des Primitivstreifens zum Neuronalen Keimblatt eignet. Das System lieferte überzeugende Argumente die Notwendigkeit der Rolle von Retinoiden in der Neuroektoderm Spezifikation durch den Primitivstreifen neu zu bewerten und weiter zu untersuchen.

LIST OF FIGURES

List of license numbers for reprinting permission

Figure 1.1: permission from Elsevier under license number 4901251270614.

Figure 1.2: permission from Springer Nature under license number 4901271338204 and 4901280053388.

Figure 1.3: (A) permission from Elsevier under license number 4901280403916, and (B) from Springer Nature under license number 4901271338204.

Figure 1.5: permission from Springer Nature under license number 4901271338204.

Figure 1.7: (A) permission from Springer Nature under license number 4901290153993, (B) from Elsevier under license number 4901290301259, and (C) from Elsevier under license number 4901281366857.

Figure 1.9: (B) permission from Springer Nature under license number 4901291255289.

Figure 1.10: reproduced with permission from the Company of Biologists Ltd.

Figure 1.11: permission from Elsevier under license number 4901301067416.

Figure 1.12: permission granted from the Cold Spring Harbor Laboratory Press.

Figure 1.14: permission from Elsevier under license number 4901310832933.

Figure 1.15: permission from Elsevier under license number 4901311099449.

Figure 1.16: (A) permission from Oxford University Press under license number 4901350378857, and (B) from Elsevier under license number 4901320950549.

Figure 4.4: (A) adapted with permission from the Company of Biologists Ltd. (B) permission from John Wiley and Sons under license number 4901360198567.

LIST OF ABRREVIATIONS

(s)gRNA	(single) guide RNA	Chrd	Chordin
2i	Two inhibitors	CK1α/γ	Casein Kinase 1α/γ
3KI	Triple reporter knock-in line	CLE	Caudal Lateral Epiblast
ActRIB/C	Activin A receptor, type 1B/C	CMV	Cytomegalovirus
ADHs	Alcohol Dehydrogenases	CRABP	Cellular Retinoic Acid Binding Proteins
AGN	AGN193109	CRBPs	Cellular Retinol-Binding Proteins
ALDHs	Aldehyde Dehydrogenase	CYP26s	cytochrome P450 26 subfamily
ALK	Activin-receptor like kinase	DAPI	4′,6-diamidino-2-phenylindole
AME	Anterior Mesendoderm	DE	Definitive Endoderm
APC	Adenomatosis Polyposis Coli	DHRS3	Short-chain dehydrogenase/reductase 3
AVE	Anterior Visceral Endoderm	Dkk1	Dickkopf WNT signaling pathway inhibitor 1
BFP	Blue Fluorescent Protein	DMEM	Dulbecco’s modified Eagle medium
bGHpA	Bovine Growth Hormone Polyadenylation Signal	DMSO	Dimethylsulfoxide
bHLH	basic helix-loop-helix	D-PBS	Dulbecco's phosphate-buffered saline
BMP	Bone morphogenetic protein	DR	Direct repeat
BMPRIA/B	Bone Morphogenetic Protein receptor, type 1A/B	DVE	Distal Visceral Endoderm
bp	base pair	Dvl	Dishevelled
BSA	Bovine Serum Albumin	E	Embryonic day
CAG	CMV enhancer, chicken beta-Actin promoter and rabbit beta-Globin splice acceptor site	EB	Elution buffer
Cer1	Cerberus-like1	EBs	Embryoid Bodies
CHIR	CHIR99021	EDTA	Ethylenediamine Tetraacetic Acid

EGO	Early Gastrula Organizer	MEFs	Mitotically-Inactivated Mouse Embryonic Fibroblast
EMT	Epithelial to Mesenchymal Transition	mESCs	mouse Embryonic Stem Cells
EpiSCs	Epiblast-like stem cells	MGO	Mid-Gastrula Organizer
ExE	Extraembryonic Ectoderm	NEB	New England Biolabs
FACS	fluorescence-activated cell sorting	NLS	Nuclear Localization Signal
FCS/FBS	Fetal Calf Serum/Fetal Bovine Serum	NMPs	Neuromesodermal Progenitors
FGF	Fibroblast Growth Factor	Nog	Noggin
FGFR	Fibroblast Growth Factor Receptor	NPs	Neural Progenitors
FRT	FLP recognition target	NSB	Node-Streak Border
FSC	Forward Scatter	Oct4	Octamer binding transcription factor 4
Fzd	Frizzled receptor	PCA	Principal Component Analysis
GFP	Green Fluorescent Protein	PD	PD184352
GSK3β	Glycogen synthase kinase 3β	Pen-Strep	Penicillin-Streptomycin
HTS	High Throughput Sampler	PGK	Phosphoglycerate kinase 1 promoter
ICM	Inner Cell Mass	PI3K	Phosphoinositide 3-kinase
IDE1/2	Inducer of Definitive Endoderm	PKC	Protein Kinase C
IGF	Insulin-like growth factor	PS	Primitive Streak
iRFP	near-infrared fluorescent protein	PSM	Presomitic Mesoderm
KI	Knock-in	RA	Retinoic Acid
KO	Knock-out	RALDHs/	Retinaldehyde Dehydrogenase
LB	Lysogeny broth	RAR	Retinoic Acid Receptor
LEF	Lymphoid enhancer binding factor	RARE	Retinoic Acid Response Elements
LG268	LG100268	RBP4	Retinol Binding Protein 4
LIF	Leukemia Inhibitory Factor	RDHs	Retinol Dehydrogenases
LRP5/6	Low-density lipoprotein receptor-related protein 5/6	rpm	revolutions per minute
MAPK	Mitogen-Activated Protein Kinase	RPMI	Roswell Park Memorial Institute

RT	room temperature
RXR	Retinoid x Receptor
SB43	SB431542
$SCF^{\beta\text{-}TRCP}$	Skp1–cullin-1–F-box β-transducin repeat–containing protein
s-FRP	secreted Frizzled-related Protein
SH2	Src homology 2
Sox1	sex determining region Y-box 1
Sox2	sex determining region Y-box 2
SSC	Side Scatter
Stat3	signal transducer and activator of transcription 3
STRA6	stimulated by retinoic acid gene 6
T	Brachyury
TAE	Tris-Acetate-EDTA
TCF	transcription factor
TGF-β	transforming growth factor-b
Tm	melting temperature
TβRI	Transforming growth factor, beta receptor I
VAD	vitamin A-deficiency
Wnt	wingless-type MMTV integration site family
WT	wild type
XAV	XAV939

1 INTRODUCTION

For almost a century, experimental embryology has investigated the mechanism enabling a single cell to generate and arrange in an orderly fashion the multitude of cell types composing an adult organism. Despite the challenges, inestimable achievements have been made, establishing the set of signaling molecules orchestrating this articulate plot and discovering that the main actors are conserved throughout the development of all vertebrates. Nevertheless, technical hurdles and paucity of experimental material imposed severe limitations, partially overcome only by the possibility to isolate and maintain *in vitro* pluripotent stem cells. Preserving the nature of their source material, these cells provided an accessible, unlimited, and easily manipulatable material to confirm and sometimes anticipate answers to developmental biology questions. In conditions where the reference embryonic process cannot be accessed, as in the case of early stages of human development, the deployment of human embryonic stem cells and the invention of micropattern devices has recently led to the identification of a human equivalent of the organizer (Martyn et al., 2018). On the other hand, only recently the organization of mouse embryonic stem cells in three-dimensional structures, named gastruloids, enabled to mimic even complex processes as somitogenesis *in vitro* (van den Brink et al., 2020).

1.1 Mouse development from blastocyst to germ layer formation

In mammals, the life of every new organism starts from a single cell, the zygote, which undergoes a process of stereotyped cleavages up to the formation of eight virtually identical cells, called blastomeres (Fig1.1). At this point, the first fate segregation of the embryo occurs, as the cells on the outer surface acquire polarization and increase their cell-to-cell contacts, in a process called compaction, whereas the ones remaining inside stay undifferentiated (Leung & Zernicka-Goetz, 2015; Nichols & Smith, 2012). By the embryonic day (E) 3.5, the emergence of a fluid-filled cavity, the blastocoel, marks the transition to the blastocyst stage, and the conceptus now consists of an outer epithelial layer, the trophectoderm, and a cluster of cells anchored in an extremity of the cavity, known as Inner Cell Mass (ICM) (Cockburn & Rossant, 2010). The trophectoderm contributes to the placenta and a portion of the parietal yolk

sac, whereas the ICM gives rise to embryo proper and other extraembryonic tissues. Between E4.0 and E4.5, the second lineage segregation occurs, involving the cells of the ICM. The fraction of these cells exposed to the blastocoelic cavity forms the primitive endoderm or hypoblast, whereas the remaining cells enclosed between the two layers of trophectoderm and primitive endoderm become the Epiblast, from which the entire fetus derives (Fig1.1) (Cockburn et al., 2010; Gadue et al., 2005).

At the time of implantation, between E4.5 and E5.0, the trophectoderm cells that come in contact with the uterus expand forming the Extraembryonic Ectoderm (ExE) and the ectoplacental cone. The primitive endoderm gradually covers the blastocoelic cavity and diverges into visceral and parietal endoderm, surrounding respectively the proliferating epiblast and the Extraembryonic Ectoderm. Upon selective apoptosis of the innermost epiblast cells, the pro-amniotic cavity is created, and the embryo adopts a cup-like shape (Fig1.1). The epiblast consists at the moment of a columnar epithelium, and the most distal part of the visceral endoderm originates a thick structure, termed Distal Visceral Endoderm (DVE). The concerted effort of restricted cell movements, selective proliferation and localized signal domains, causes the translocation of the Distal Visceral Endoderm to the prospective anterior pole of the embryo, becoming Anterior Visceral Endoderm (AVE) (Gadue et al., 2005; Tam & Loebel, 2007).

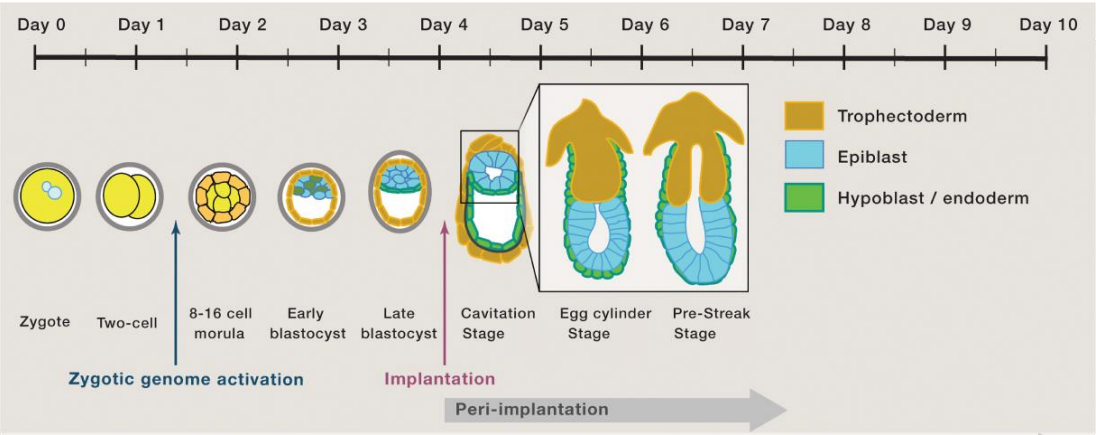


Figure 1.1: Embryonic development from pre-implantation to gastrulation.

Schematic depicting the first steps of mouse embryogenesis from a single cell to gastrulation. For the peri-implantation stages after cavitation, only the embryonic region of the mouse conceptus is displayed. Adapted from Rossant and Tam 2017.

The establishment of the anterior-posterior polarity in the epiblast and visceral endoderm completes the premises for the arguably major milestone of development, the gastrulation, through which the entire embryo body plan is established, and its main axes are defined. During this process, the epiblast single layered epithelium gives rise to the three primary germ layers: endoderm, precursor of the epithelium of the gut tube, lungs and organs including pancreas, liver and thyroid; mesoderm, contributing to hematopoietic, vascular, cardiac and skeletal muscle derivatives; ectoderm, further differentiating in the central and peripheral nervous systems and the epidermis as well (Loebel et al., 2003; Murry & Keller, 2008). The beginning of the gastrulation is classically considered the emergence at E6.25 of a structure called Primitive Streak (PS) at the proximal-posterior pole of the epiblast. The Streak is marked by the expression of the T-box transcription factor Brachyury (T), starting in the region in contact with the Extraembryonic Ectoderm at the opposite side of the Anterior Visceral Endoderm (Beddington et al., 1993; Ramkumar & Anderson, 2011; Rivera-Pérez & Magnuson, 2005; Wilkinson et al., 1990). This structure comprises highly proliferative and motile cells which undergo epithelial to mesenchymal transition (EMT) and intercalate between the epiblast and the visceral endoderm layers. From its onset, the streak expands towards the distal tip of the embryo, including new cells along the way, while others continuously exit laterally and medially and become specified (Ramkumar & Anderson, 2011). The fate acquired by the cells leaving the streak is subject to tight spatiotemporal regulation, dictated by regionally controlled signaling environments (Fig1.2). Accordingly, the first cells exiting the PS originate extraembryonic mesoderm forming allantois and amnion, as well as hematopoietic, endothelial, and vascular smooth muscle cells of the yolk sac. Cells migrating through more anterior parts of the PS give rise to cranial and cardiac mesoderm first, and then to paraxial and axial mesoderm (Fig1.2). The definitive endoderm (DE) arises from the epiblast cells included in the most anterior region of the PS which, around E6.75, ingress the visceral endoderm layer displacing it proximally, while covering the distal side of the embryo (Gadue et al., 2005; Murry et al., 2008). On the other hand, the anterior epiblast cells, which do not enter the PS, originate the ectoderm, as a result of the interactions between signals secreted by all the surrounding tissues (Fig1.2) (Gadue et al., 2005).

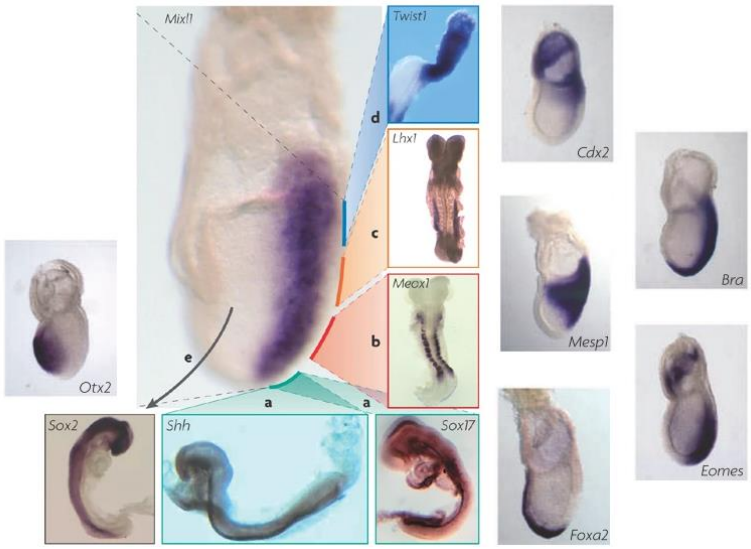


Figure 1.2: Fate allocation during gastrulation.

*Emergence of tissue progenitors from the Primitive Streak during gastrulation revealed by the expression of fate markers, ordered from most anterior to posterior: axial mesoderm (expressing Sonic Hedgehog, Shh) and definitive endoderm (expressing Sox17) (box **a**) derive from the most anterior region of the streak (node), which is marked by Forkhead box A2 (Foxa2) and the T-box gene eomesodermin (Eomes); paraxial mesoderm (expressing Meox1, box **b**), lateral plate mesoderm (expressing Lhx1, box **c**) and cardiac mesoderm (marked by Mesp1), originate from regions in the central portion of the streak; the posterior region of the streak, marked by the expression of Caudal-type homeobox protein 2 (Cdx2), gives rise to progenitor of extraembryonic mesoderm (for instance, allantois, expressing Twist1; box **d**). MIX1 homeobox-like 1 (Mixl1) and the T-box transcription factor Brachyury (Bra; also known as T) are expressed throughout the entire streak, with the latter marking the node and the notochord as well. Cells in the anterior region of the embryo, not included in the primitive streak, become part of the ectoderm, as indicated by the expression of orthodenticle homologue 2 (Otx2) and Sox2. Adapted from Arnold & Robertson, 2009; Tam et al., 2007.*

1.2 Signaling pathways in the context of gastrulation

It is still debated how the distribution of the signaling domains orchestrating gastrulation is achieved, whether is asymmetric gene expression to determine a divergence of fates, leading to differential localization of certain populations, or is the uneven exposure to signals in certain regions of the embryo to originate the heterogeneity, later fixed into patterns. Consequently, is not clear when fates are

irreversibly determined, because this would require considering each population in isolation, away from the influences of the others.

The correct execution of the gastrulation plot and establishing the main embryonic axes relies on cooperative or antagonistic interactions between different signaling domains, determining the formation of gradients of diffusible signals. The key signaling centers in this perspective are the Extraembryonic Ectoderm, the Anterior Visceral Endoderm (AVE), the Primitive Streak (PS) and its distal tip, the node. Across the anterior-posterior axis, the opposite actions of the PS, which secrets Wnt3 and Nodal, and the AVE, antagonizing these signals by secreting respectively Dkk1 and Lefty-1, in addition to the bivalent inhibitor Cerberus-like 1 (Cer1), establishes a gradient of Wnt and Nodal/TGF-β pathways (Fig1.3). On the other hand, a BMP gradient is achieved along the proximal-distal axis, owing to the production of BMP4 from the Extraembryonic Ectoderm, and of its secreted inhibitors Noggin, Chordin and Follistatin from the node (Fig1.3) (Gadue et al., 2005; Tam et al., 2006).

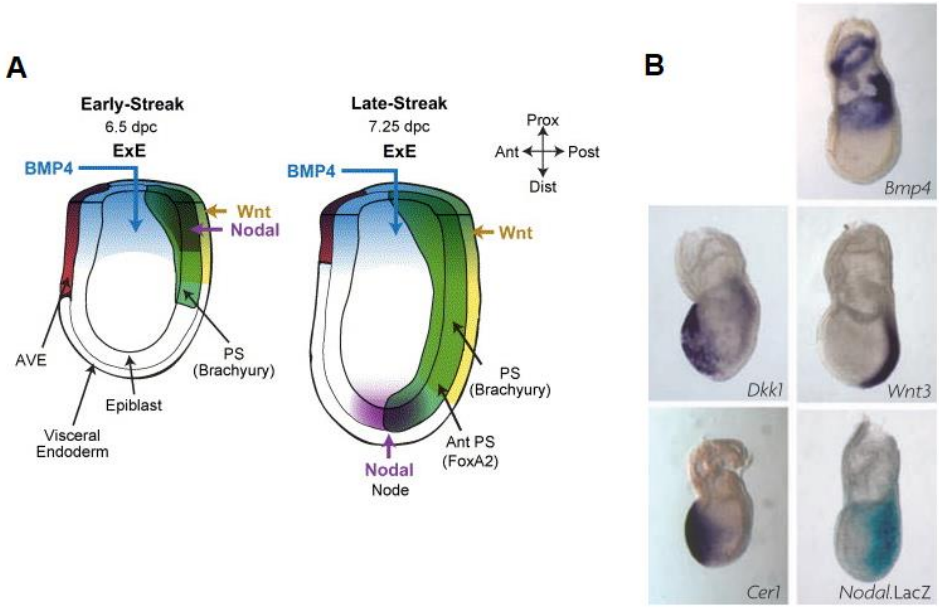


Figure 1.3: Signaling pathways at the basis of gastrulation.

A*) Schematic representation of the signaling gradients acting during the Primitive Streak formation (adapted from Gadue et al., 2005).* ***B****) RNA in situ hybridization of signaling molecules at E7.5. Posterior expression of Bmp4 in the proximal region, and of Wnt3 and Nodal (through LacZ staining) in the streak, versus expression of the inhibitors cerberus-like 1 (Cer1) and Dickkopf homologue 1 (Dkk1) on the anterior side of the embryo. Adapted from Arnold and Robertson 2009.*

Before approaching these pathways as single entities, it is worth mentioning that BMP, Nodal and Wnt signaling constitute an intricated network underlying gastrulation. Briefly, Nodal production in the epiblast activates the expression of *Bmp4* in the Extraembryonic Ectoderm. Bmp4 promotes the expression of *Wnt3* in the adjacent epiblast, which in turn activates the transcription of *Nodal* and its co-receptor *Cripto*, establishing a positive feedback loop (Tam & Loebel, 2007). The specific role of each of these pathways in the context of gastrulation has been investigated *in vivo* by transplantation and mutagenesis studies and more recently *in vitro* in embryonic stem cells.

1.2.1 The Transforming Growth Factor-β Pathway

The Transforming Growth Factor (TGF)-β superfamily of signaling molecules comprises 29 secreted ligands which can be further subdivided into the TGF-β/Activin/Nodal subfamily and the Bone morphogenetic protein (BMP) subfamily. Distinct combinations of receptors and intracellular transducers allow each subgroup to exert specific functions, that ultimately involve regulating the expression of a set of target genes. The general mechanism of transduction requires that the ligand, which acts as disulphide-linked dimer, binds at the membrane specific combinations of two type I and two type II receptors, both having Ser/Thr kinase activity (Fig1.4). In vertebrates, seven type I receptors, Activin-receptor like kinases (ALKs) 1–7, and five type II receptors have been identified. In the newly formed complex, the type II receptors phosphorylate the type I, which in turn amplify the signal by phosphorylation of the receptor-regulated SMADs (R-SMADs). SMAD1, 5 and 8 are activated by the binding of BMP to type I receptors ALK1, ALK2, ALK3/BMPRIA, and ALK6/BMPRIB; whereas SMAD2/3 activation is triggered by the interaction of TGF-β/Activin/Nodal ligands to the receptors ALK4/ActRIB, ALK5/TβRI or ALK7/ActRIC. The activated R-SMADs form heterotrimeric complexes with the common mediator SMAD4 (co-SMAD) and then translocate into the nucleus (Fig1.4). Here, partnering with other DNA-binding transcription factors, the R-SMADs regulate the expression of their target genes (Derynck & Zhang, 2003; Kubiczkova et al., 2012; Massagué, 2012).

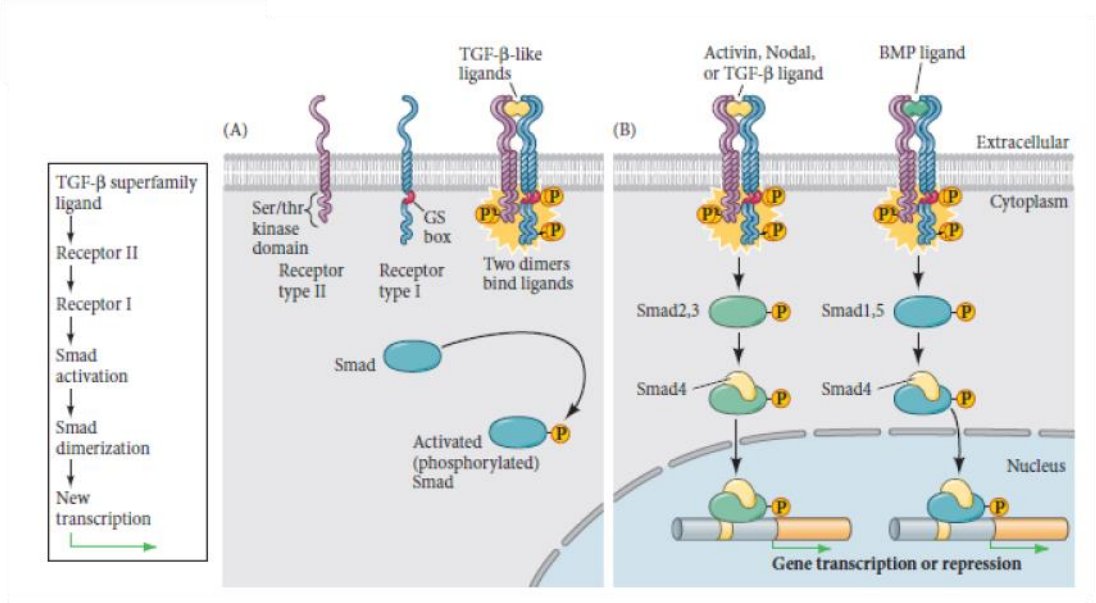


Figure 1.4: The Transforming Growth Factor-β/SMAD pathway.

__A__) Activation of the receptors and SMAD phosphorylation upon binding of TGF-β superfamily ligands. __B__) Smad2/3 activation by Activin/Nodal/TGF-β type of ligands as compared to Smad1/5 activation by BMP family of ligands. Smad4 is a common mediator of both branches of the TGF-β pathway. Adapted from Gilbert S. F., Barresi M. J. F., Developmental Biology.

The anterior-posterior patterning of the epiblast and the correct formation of the PS rely on a fine balance between Nodal and its antagonists produced by the Anterior Visceral Endoderm. Nodal expression is first detected at E5.5 in the epiblast and primitive endoderm, but at E6.5 it becomes restricted to the posterior epiblast and nascent PS (Conlon et al., 1994; Varlet et al., 1997). At late streak stage, Nodal disappears from the PS and its expression is only detected in the node (Collignon et al., 1996). In absence of Nodal signaling the PS does not form, but milder degrees of signaling disruption, combining different *Smad2* and *Smad3* mutant alleles or using hypomorphic *Nodal* alleles, lead to the loss of anterior PS structures such as axial and paraxial mesoderm along with definitive endoderm (Dunn et al., 2004; Lowe et al., 2001; Vincent et al., 2003). For this reason, it is commonly agreed that the formation of anterior PS derivatives requires high doses of Nodal signaling. Nodal is also critical for the specification and migration of the Distal Visceral Endoderm. In fact, in absence of Nodal this structure is not induced, whereas in mice bearing null mutations of the Nodal co-receptor *Cripto*, it is present but fails to progress towards the prospective anterior pole (Brennan et al., 2001; Ding et al., 1998; Waldrip et al., 1998). Nodal

antagonism is as important as its activation, and embryos lacking both the inhibitors Lefty1 and Cerberus1 show an expanded anterior PS domain or in some cases even multiple Primitive Streaks (Perea-Gomez et al., 2002).

Bmp4 expression is first detected between E5.5 and E6.0 in the Extraembryonic Ectoderm, with particularly high levels in the region adjacent the epiblast. At mid-streak stage, it gets localized in the newly formed extraembryonic mesoderm and in the allantois and mesodermal components of amnion, chorion, and visceral yolk sac (Lawson et al., 1999). From E7.5-8.5, *Bmp4* is expressed also in the posterior PS and embryonic mesoderm (Winnier et al., 1995). Disruption of the genes coding for either BMP4, BMPRIA or BMPRII is lethal within E9.5 with defects affecting mostly the mesoderm germ layer. Posterior derivatives of the PS, such as extraembryonic and hematopoietic mesoderm, are in fact absent in these conditions (Beppu et al., 2000; Mishina et al., 1995; Winnier et al., 1995). Tetraploid complementation experiments using *Bmp4*$^{-/-}$ ES cells, in which the mutant cells contribute to the epiblast, whereas the extraembryonic tissues derive from wild type (WT) cells, demonstrated that embryos undergo gastrulation and even manage to develop a PS and a node. The finding proved that the Extraembryonic Ectoderm and not the epiblast is the source of Bmp4 required for the onset of gastrulation and PS formation (Fujiwara et al., 2002). The role of TGF-β and particularly BMP inhibition in neuroectoderm induction is discussed in a separate section.

1.2.2 Canonical Wnt/β -catenin pathway

Mammalian genomes encode for 19 Wnt protein ligands and 10 seven-transmembrane Frizzled receptors (Fzd) (Nusse, 2008). The canonical way of transduction of Wnt signaling is centered on the regulation of the cytoplasmic levels of β-catenin. In the absence of ligand, β-catenin is directed to ubiquitin-mediated proteasomal degradation by the β-catenin destruction complex (Fig1.5). The complex comprises the scaffolding proteins Axin and APC (adenomatosis polyposis coli), and the kinases CK1α (casein kinase 1α) and GSK3β (Glycogen synthase kinase 3β), and catalyzes the serine/threonine phosphorylation of the N-terminal domain of β-catenin, causing its ubiquitination by the E3-ubiquitin ligase, $SCF^{\beta\text{-}TRCP}$. The binding of Wnt ligands to Fzd receptors and the LRP5 or LRP6 co-receptors, recruits Dishevelled (Dvl) to the membrane, and this in turn Axin-GSK3 complexes. The kinases GSK3,

Cyclin Y and CK1γ (casein kinase 1γ) phosphorylate the C-terminal intracellular domain of LRP5/6, which in turn directly repress the GSK3 activity (Fig1.5). The sequestration at the membrane of components of the β-catenin destruction complex, along with direct GSK3 inhibition by phospho-LRP5/6, lead to the increase of the cytoplasmatic levels of β-catenin, and its translocation to the nucleus. Here, β-catenin acts as transcriptional co-activator, interacting with the TCF/LEF DNA-binding proteins, and converting their function from repression to activation of the Wnt target genes (Steinhart & Angers, 2018).

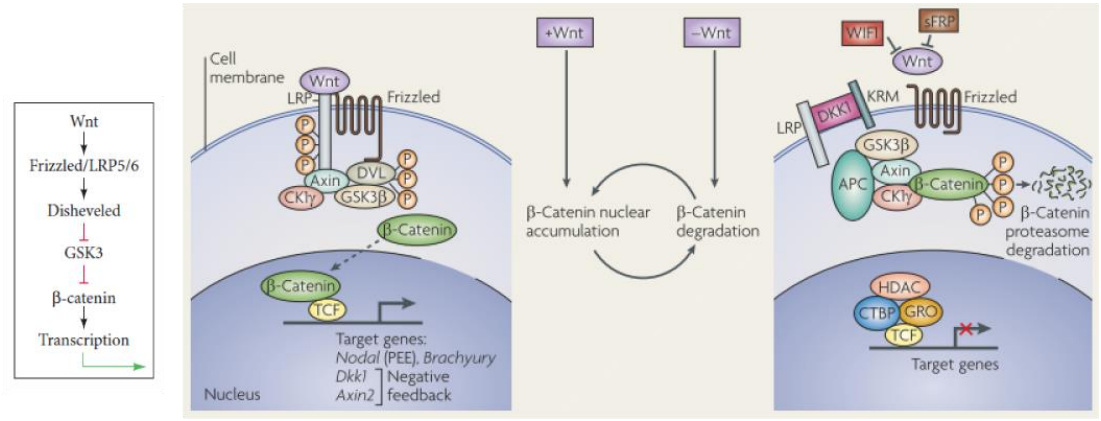


Figure 1.5: The canonical Wnt/β-catenin pathway.

The canonical Wnt pathway is essentially based on the control of the cytoplasmic levels of β-catenin; in absence of Wnt ligand, β-catenin is targeted to proteasomal degradation by the destruction complex, which is instead inhibited when Wnt binds to frizzled receptors. Secreted inhibitors such as the secreted frizzled-related proteins (sFRPs) or the inhibitor of LRP co-receptors, DKK1, cause β-catenin degradation by preventing Wnt binding to the receptors. Adapted from Arnold and Robertson 2009.

Canonical Wnt signaling is first detected around E6.0 upon activation by Wnt3 in the proximal-posterior epiblast, where it marks the emergence of the Primitive Streak and, then, it follows the progression of the streak distally (Wang et al., 2012). Wnt3 expression is downregulated and eventually terminated by E8.5, but from E7.5 another canonical Wnt ligand, Wnt3a, gets upregulated in the PS. One critical target of canonical Wnt activation during this time is the PS marker *T* (*Brachyury*), containing multiple Lef1 binding sites in its regulatory sequence, which have proved necessary for maintaining but not initiating the expression of the gene (Galceran et al., 2001).

The essential role of the Wnt pathway in the generation of the PS is demonstrated by the observation that *Wnt3*$^{-/-}$, *ctnnb1*$^{-/-}$ (β-catenin), or *Lrp5*$^{-/-}$; *Lrp6*$^{-/-}$ mutant mice do not present any PS and the epiblast remains undifferentiated (Haegel et al., 1995; Kelly et al., 2004; Liu et al., 1999). *Wnt3a*$^{-/-}$ mutant embryos, instead, initiate gastrulation but display posterior truncations, lack paraxial mesoderm and form ectopic neural tubes (Takada et al., 1994; Yoshikawa et al., 1997). In *ctnnb1*$^{-/-}$ mutants, the Distal Visceral Endoderm does not migrate proximally to form the Anterior Visceral Endoderm and it is still debated whether this effect might be due to reduced *Cripto* expression and Nodal signaling promoted by Wnt, or to the absence of Wnt activation *per se* (Wang et al., 2012).

Whilst Wnt activation is of paramount importance for the formation of the Primitive Streak and its derivatives in the posterior epiblast, proper patterning of the anterior epiblast and neuroectoderm requires its controlled repression. Dkk1, Wnt inhibitor which acts by binding to the LRP5/6 co-receptors, is produced by the Anterior Visceral Endoderm, and the anterior mesendoderm (AME) derived from the streak (Kimura-Yoshida et al., 2005; Mukhopadhyay et al., 2001). *Dkk1*$^{-/-}$ null mutations cause expansion of the posterior Wnt domain and forebrain truncation by E9.5 (Lewis et al., 2008; Mukhopadhyay et al., 2001).

1.2.3 The Fibroblast Growth Factor pathway

The Fibroblast Growth Factor (FGF) family of extracellular ligands comprises 22 members in the vertebrate genome that interact with four tyrosine kinase receptors (FGFR1-4). The binding of the ligands to their receptors requires in addition heparan sulphate, with a stoichiometry of 2:2:2 FGF:FGFR:heparan (Dorey & Amaya, 2010). The receptors dimerization upon ligand binding, triggers their tyrosine kinase activity, leading to the autophosphorylation of intracellular domains. The autophosphorylation controls the kinase activity of the receptors and the phosphorylated tyrosines act as docking sites for SH2 (Src homology 2) and other phosphotyrosine binding domains of adaptor proteins or enzymes, whose catalytic activity gets activated upon binding. The transduction of FGF signal can then proceed via three main routes: the Ras/MAPK pathway, the PI3K/Akt pathway or the protein kinase C (PKC) pathway (Böttcher & Niehrs, 2005).

Fgf4 and Fgf8 are the two most important members of the family expressed during gastrulation. *Fgf4* is under direct control of Oct4 and Sox2 and its expression starts in the Inner Cell Mass. At the onset of gastrulation, it gets restricted to the PS first and to the anterior region of the streak afterwards (Niswander & Martin, 1992; Sun et al., 1999). *Fgf8* expression, promoted by Wnt3, starts in the posterior epiblast and subsequently in the streak. Also the Fgfr1 receptor, even though being first expressed throughout the epiblast before gastrulation, becomes later restricted to the PS (Yamaguchi et al., 1992). FGF signaling has a major role in the EMT, proliferation and migration of the cells from the streak (Gadue et al., 2005). Mice *fgf4*$^{-/-}$, die shortly after implantation with impaired proliferation of the ICM, whereas in *fgfr1*$^{-/-}$ mutant embryos, epiblast cells do not undergo EMT and do not ingress the streak (Deng et al., 1994; Feldman et al., 1995; Yamaguchi et al., 1994). As for *fgf8* mutation, the cells can ingress the streak but then fail to disperse resulting in the accumulation of mesenchyme (Tam & Loebel, 2007).

1.3 Ectoderm formation: pioneering studies and default model

With their experiments, Spemann and Mangold demonstrated that transplanting the dorsal lip of the amphibian blastopore to the ventral side of a host newt embryo induced the formation of a complete ectopic neural axis. Interestingly, the new axis was composed mostly of cells from the host ectoderm rather than from the graft, suggesting that the donor population possessed an instructive capability, for which reason it was named “organizer” (Fig1.6) (Spemann & Mangold, 1924). It was also noticed that dorsal blastopore lip cells from older embryos, when transplanted, induced the formation of structure missing the forebrain (eyes) (Andoniadou & Martinez-Barbera, 2013). This observation led to hypothesize that the organizer had stage-dependent inducing abilities: early induction by a “head organizer” would promote anterior neural identity, whereas the late activity of a “trunk organizer” induces only posterior neural fates. Subsequent experiments by Nieuwkoop, in the urodele embryo, highlighted the caveats of this theory leading to the definition of the “induction-transformation” model. He implanted pieces of naïve ectoderm in different positions along the rostro-caudal axis of the host embryo and tested whether they formed anterior or posterior structures. Surprisingly, the donor cells matched the neural identity of neighboring tissues where they were inserted. According to the model

formulated from these experiments, the neural tissue has an anterior character at the moment of its induction, and it subsequently acquires posterior character by the action of signals deriving from the organizer (Nieuwkoop, 1952).

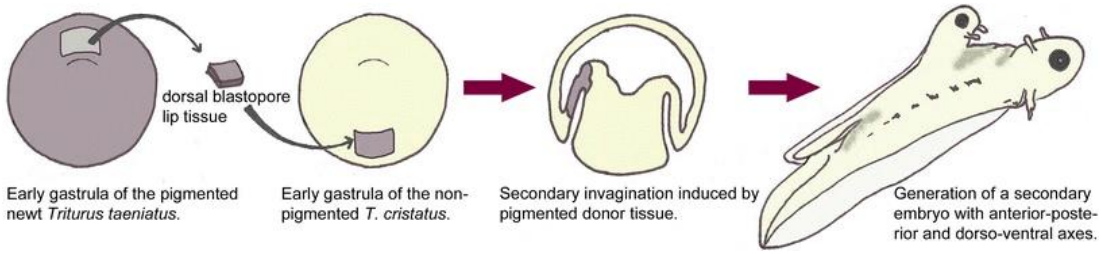


Figure 1.6: The discovery of the organizing activity.

Schematic representation of the classical studies in the newt embryo that led to the discovery of the structure, termed organizer, able to induce the formation of a second neural axis mainly composed by the host tissues. From Andoniadou and Martinez-Barbera 2013.

Studies in *Xenopus* shed lights on the activity of the organizer and the molecular underpinnings of the anterior neural induction. It was observed that intact animal caps in culture originated surface ectoderm, whereas disaggregation converted their fate into anterior neural ectoderm. Similar experiments suggested that the formation of neural tissue might require the removal of some unknown inhibitory substances, and the dissociation of the cells allowed their dilution (Grunz & Tacke, 1989). These observations constituted the pillar of the "default model" of neural induction, proposing that the default fate of the cells within the ectoderm layer is neural, in absence of substances converting them into surface ectoderm (Hemmati-Brivanlou et al., 1994; Levine & Brivanlou, 2007; Stern, 2005). It was later understood that BMP4 was responsible for the conversion of the explants into epidermis (Wilson & Hemmati-Brivanlou, 1995), and that the organizer owed its neural induction property to the production of three BMP antagonists: Noggin (Lamb et al., 1993; Smith et al., 1993; Smith & Harland, 1992), Chordin (Sasai et al., 1994, 1995) and Follistatin (Hemmati-Brivanlou et al., 1994). The simultaneous depletion of the three antagonists in *Xenopous* by morpholino injection causes ventralization of the embryo and almost complete loss of neural plate (Khokha et al., 2005).

1.3.1 Mouse organizers and anterior neural induction

Since these studies, regions equivalent to the amphibian organizer have been identified in most vertebrate classes, including mouse, where it is located at the distal

tip of the Primitive Streak, and takes the name of node (De Robertis et al., 2000; Storey et al., 1992). Nevertheless, transplantation of the mouse node failed in inducing an ectopic axis including a forebrain. It was soon clear that the organizer/node does not represent a stable entity, but it is rather contributed by constantly evolving populations with different inducing properties. In fact, based on morphogenetic capabilities and cellular composition one can distinguish an early gastrula organizer (EGO), comprising progenitors of anterior axial mesoderm, cranial and heart mesoderm and foregut mesoderm; a mid-gastrula organizer (MGO) including progenitors of the whole anterior axial mesoderm and notochord; and the node, whose cells contribute primarily to the notochord formation (Robb & Tam, 2004). None of these structures was able alone to induce an entire neural axis, but they could do so when associated with the Anterior Visceral Endoderm (AVE) (Thomas & Beddington, 1996). Originating before the formation of the PS, the AVE plays a critical role in the induction of the anterior neural plate, expressing genes such as *Otx2*, *Cer1*, *Lefty1*, *Dkk1*, *Foxa2* and *Lhx1*. The disruption of each of these genes specifically in the AVE demonstrated its requirement for the anterior forebrain development and anteroposterior patterning (Kimura-Yoshida et al., 2005; Perea-Gomez et al., 2002; Rhinn et al., 1998; Yamamoto et al., 2004). As already mentioned, *Cer1*$^{-/-}$; *Lefty1*$^{-/-}$ mutant mice present an expansion of the PS domain towards anterior regions (Perea-Gomez et al., 2002). Therefore, the AVE is considered necessary to establish anterior-posterior polarity positioning the PS, and to "prime" the epiblast for the anterior neural induction.

The onset of the neural induction is generally set around early/mid-streak stage, marked by the restriction of *Otx2* expression to the anterior epiblast, followed by the expression in the same location of the anterior markers *Hesx1* and *Six3*. Highlighting the importance of fine tuning the BMP pathway for correct anterior ectoderm formation, *Bmpr1a*$^{-/-}$ mice embryos suffer of premature neural induction at pre-gastrula stage, marked by the expression of *Hesx1* and *Six3* (Di-Gregorio et al., 2007). On the contrary, simultaneous mutation of both the BMP inhibitors Noggin and Chordin (*Nog*$^{-/-}$;*Chrd*$^{-/-}$) cause severe forebrain truncations, with defects in the anterior patterning arising as early as the late-streak stage and absence of *Hesx1* and *Six3* expression (Fig1.7a). The absence of Noggin and Chordin in the Anterior Visceral Endoderm, points out that their production by the node is essential to induce proper

anterior structures (Bachiller et al., 2000). However, the scenario is complicated by the analysis of conditions where one structure between node and AVE is absent, and the other one could still be partially functional. In fact, mouse embryos lacking Wnt3 or β-catenin, preserve the expression of AVE markers, but do not form the PS and fail to induce neural specification (Huelsken et al., 2000; Liu et al., 1999). On the other hand, surgical removal of the node in the mouse embryo has little effect on the anteroposterior patterning, whereas it disrupts dorso-ventral patterning and left-right asymmetry, probably owing to the absence of the notochord derived from the node, which is the source of the Sonic Hedgehog signaling required for patterning of the ventral tube (Klingensmith et al., 1999).

While Anterior Visceral Endoderm, gastrula organizers and node collaborate to set up neural induction, the protection of the newly induced neural plate from the posteriorizing signals is considered a prerogative of the anterior axial mesendoderm (AME). The AME derives from the cells composing organizers and node and includes definitive endoderm and mesodermal component of the prechordal plate. It exerts its functions by secreting the same set of inhibitors adopted by the node and AVE for the neural induction: *Dkk1*, *Cer1*, *Nog* and *Chrd* (Del Barco Barrantes et al., 2003; Mukhopadhyay et al., 2001; Yang & Klingensmith, 2006). The AME plays also an important role in the establishment of new signaling centers within the neural plate, allowing for expansion and further patterning of the forebrain (Andoniadou & Martinez-Barbera, 2013). In *Xenopus*, the overexpression of *Cer1* (Cerberus), secreted inhibitor of Nodal, BMP and WNT, is able to induce an ectopic full axis, highlighting the importance for the head formation of repressing both BMP and Nodal/TGF-β families and the WNT pathway (Bouwmeester et al., 1996; Glinka et al., 1997; Piccolo et al., 1999). In this context, Wnt signaling is regarded as one of the caudalizing activities, together with the retinoic acid pathway, from which the newly originated neural tissue must be protected. In fact, the loss of forebrain structures caused by *Dkk1* mutations, as in the case of Noggin and Chordin compound mutants, it is considered a consequence of the absence of Wnt inhibition by the node-derived Anterior Mesendoderm (Fig1.7b) (Mukhopadhyay et al., 2001). Taken together these finding point out that the Anterior Visceral Endoderm, gastrula organizers, node and Anterior Mesendoderm synergize, using secreted inhibitors against the same pathways, to allow the formation of neuroectoderm cell types in the anterior epiblast, and subsequently

protect it against posteriorizing signals, a concept known as posterior protection hypothesis (Fig1.7c) (Andoniadou & Martinez-Barbera, 2013).

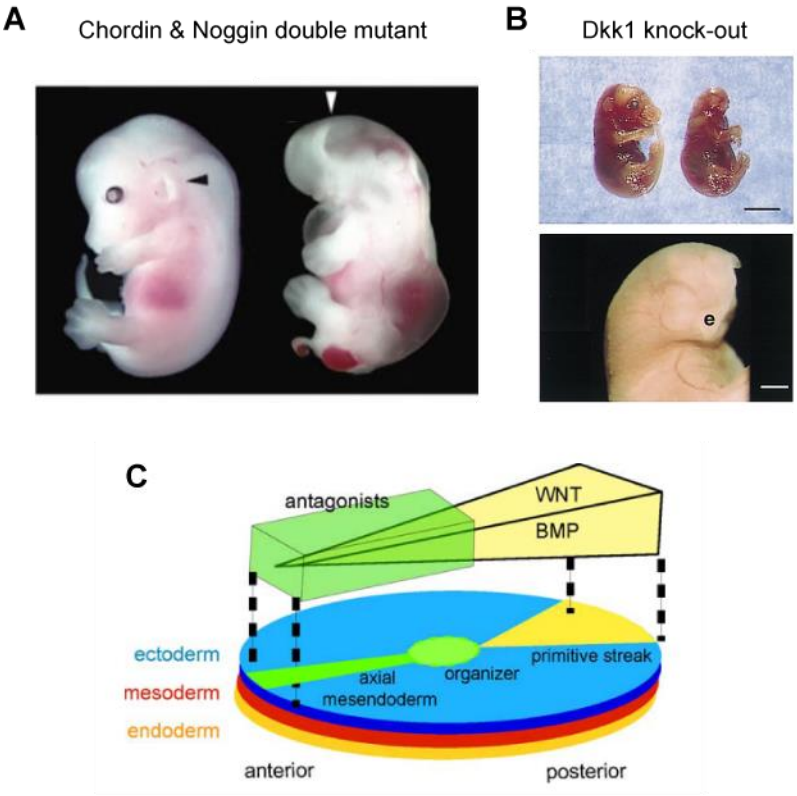


Figure 1.7: Concerted induction of anterior neuroectoderm by Wnt and BMP inhibitors.

__A__) Truncation of the anterior neuroectoderm caused by combined mutation of the two BMP inhibitors, Chordin and Noggin. Adapted from Bachiller et al. 2000. __B__) Forebrain truncation following mutation of the Wnt inhibitor, Dkk1. Adapted from Mukhopadhyay et al. 2001. __C__) Flattened representation of mouse gastrula, showing the streak and its signaling gradients starting from the posterior region of the embryo in yellow, whereas in green are the organizer, the axial mesoderm and their activity inhibiting Wnt and BMP gradient in the anterior epiblast, allowing the formation of ectoderm in this region. From Robb and Tam 2004.

1.3.2 Formation of the posterior neuroectoderm

In a broad range of species, the only structures formed at the end of gastrulation are the head and the anterior trunk. The embryonic development proceeds then progressively from anterior to posterior, as exemplified by the formation of the somites, at the basis of the extension of the anterior-posterior axis (Kimelman, 2016). Soon after gastrulation, the ectoderm diverges in surface ectoderm forming the epidermis, and neuroectoderm, consisting at this stage of a single-layered pseudostratified epithelium called neural plate (Andoniadou & Martinez-Barbera,

2013). The plate later folds into the neural tube, which originates all the rostral structures of the nervous system (forebrain, midbrain, and anterior hindbrain). The posterior hindbrain, precursor of the spinal cord, originates from a bipotent population called neuromesodermal progenitors (NMPs) responsible for the formation of trunk and tail (Kimelman, 2016; Ozair et al., 2013). The earliest known specific neuroectoderm marker is the transcription factor of the SoxB1 subfamily, Sox1 (Wood & Episkopou, 1999). First expressed in the neural plate, it is subsequently maintained in neuroepithelial cells along the entire neural axis and is later downregulated during neuronal and glial differentiation.

As anticipated, although the term ectoderm usually refers to a single germ layer, the structures that will compose the central nervous system, namely brain and spinal cord, have different developmental origins. The conventional view of the segregation of the three germ layers at the gastrulation stage, generating progenitors with restricted competence, is particularly challenged by the formation of the spinal cord during body axis extension (Tzouanacou et al., 2009). In fact, clonal analysis studies using transgenic mouse embryos showed that epidermis, endoderm and cranial and cardiac mesoderm diverge early on as independent lineages, but somites and spinal cord originate from neuromesodermal progenitors (NMPs), located in the caudal lateral epiblast (CLE) region, and node-streak border (NSB) (Tzouanacou et al., 2009). Consequently, the precursors of the spinal cord are genealogically closer related to somite-forming paraxial mesoderm than to neural progenitors of the brain and surface ectoderm. The linage tracing experiments also argued against the presence of a distinct mesendodermal precursor in mouse, but they highlighted a closer relationship between endoderm and early ingressing mesoderm than between endoderm and ectoderm or posterior mesoderm (Tzouanacou et al., 2009).

The NMPs are characterized by the expression of a set of transcription factors, such as T, Sox2 and Nkx1-2 (Martin & Kimelman, 2012; Olivera-Martinez et al., 2012; Schubert et al., 1995; Tsakiridis et al., 2014). The fate of the cells derived from the NMPs seems to be dependent on the relative levels of T and Sox2, with high level of the former promoting specification towards mesoderm and higher level of the latter favoring instead neural fate. T(Brachyury) is the founder member of the T-box transcription factor family and, as previously mentioned, is the main PS marker. Disruption of its gene in different vertebrates causes the truncation of posterior body

parts, with absence of notochord and somitic mesoderm (Chesley, 1935; Gluecksohn-Schoenheimer, 1944). It has been proposed that in the context of the NMPs T would exert its function in a non-cell autonomous way, by creating a niche that allows a pool of NMPs to remain progenitors, while producing more differentiated derivatives during axis elongation (Martin & Kimelman, 2010). Given that some members of the Wnt family are direct targets of T (Martin & Kimelman, 2008), and that *T* itself is regulated by Wnt (Arnold et al., 2000; Vonica & Gumbiner, 2002; Yamaguchi et al., 1999), it is proposed that T and Wnt establish an autoregulatory loop at the basis of NMPs maintenance. This loop can be disrupted by increase of the relative levels of the transcription factor Tbx6, promoted by elevation of Wnt signaling and T itself, determining the specification to mesoderm and the repression of *Sox2*, or by the activation of retinoic acid signaling, which instead favors the acquisition of neural fate (Gouti et al., 2017).

1.4 Retinoic acid signaling and metabolism

Retinoids are lipophilic molecules provided with the diet that cannot be synthesized de novo. During embryonic development their only source is transfer of maternal retinol (vitamin A) through the placenta (Rhinn & Dolle, 2012). The uptake of retinoids by target tissues occurs either by diffusion through the membrane or can be facilitated by active transport of retinol, bound to the carrier protein RBP4, by the transmembrane protein STRA6 (Kawaguchi et al., 2007). Once inside the cell, vitamin A is bound by the cellular retinol-binding proteins (CRBPs) and it can be converted into its bioactive metabolite, retinoic acid (RA), in a two-step process (Fig1.8a). The first step of the synthesis is retinol oxidation to retinaldehyde by two classes of enzymes, the alcohol dehydrogenases (ADHs) and the retinol dehydrogenases (RDHs). This conversion can be reverted *in vivo*, to prevent excessive RA synthesis, by the enzyme retinaldehyde reductase, DHRS3 (Billings et al., 2013; Feng et al., 2010). The subsequent oxidation of retinaldehyde to RA is catalyzed by the retinaldehyde dehydrogenase enzymes (RALDH1-3/ALDH1A1-3) and is irreversible. In the cytoplasm, RA can bind to the cellular retinoic acid binding proteins (CRABP1 and 2), while its clearance is carried out by enzymes of the cytochrome P450 26 subfamily CYP26A1, CYP26B1 and CYP26C1, which further oxidize RA to more polar metabolites (4-hydroxy-RA and 4-oxo-RA) (Fig1.8a) (Hernandez et al., 2007;

Pennimpede et al., 2010). Although 4-oxo-RA is able to bind to RA receptors and even interfere with embryonic patterning when administered exogenously (Pijnappel et al., 1993), elegant genetic studies demonstrated incontrovertibly that the products of CYP26s enzymatic reaction are inactive metabolites (Niederreither et al., 2002). Similarly, it has long been debated which of the several isomeric forms of RA, such as all-*trans*-RA, 9-*cis*-RA, 13-*cis*-RA, all-*trans*-3,4-dehydro-RA, has actual functions during development, before a clear consensus was reached in considering all-*trans*-RA the main active retinoid during vertebrate development (Cunningham & Duester, 2015).

RA acts by interacting with nuclear receptors belonging to two different families: the retinoic acid receptors (RARα,β,γ) and the retinoid X receptors (RXRα,β,γ). The classical mechanism of action sees the receptors binding as RAR-RXR heterodimers to DNA sequences present in target genes known as RA response elements (RAREs) (Ghyselinck & Duester, 2019). As the target sequences of many other nuclear receptors, the RAREs consist of a direct repeat of the hexameric consensus sequence 5'-(A/G)G(G/T)TCA-3' with characteristic spacings of 1, 2 or 5 base pairs, known as DR1, DR2 or DR5 respectively (Balmer & Blomhoff, 2002). More recently, new RAREs have been proposed with "unusual" spacing, such as DR0 and DR8; however, the function of these new responsive elements is still under investigation (Moutier et al., 2012). In any case, the interaction of RA with the receptor dimers can either cause a conformational change of complexes already bound to the target sequences, or trigger de novo binding (Mahony et al., 2011), and the final result of this interaction is the activation or repression of transcription, recruiting respectively nuclear receptors coactivators (NCoA) or nuclear receptor corepressors (NCoR) (Fig1.8b) (Cunningham & Duester, 2015). Among the RA target genes are members of the retinoid pathway and metabolism such as *Rarb*, *Crbp1/2*, *Crabp1/2* and *Cyp26a1*, along with several members of the Hox gene family, including *Hoxa1*, *Hoxb1*, *Hoxb4* and *Hoxd4* (Marshall et al., 1996).

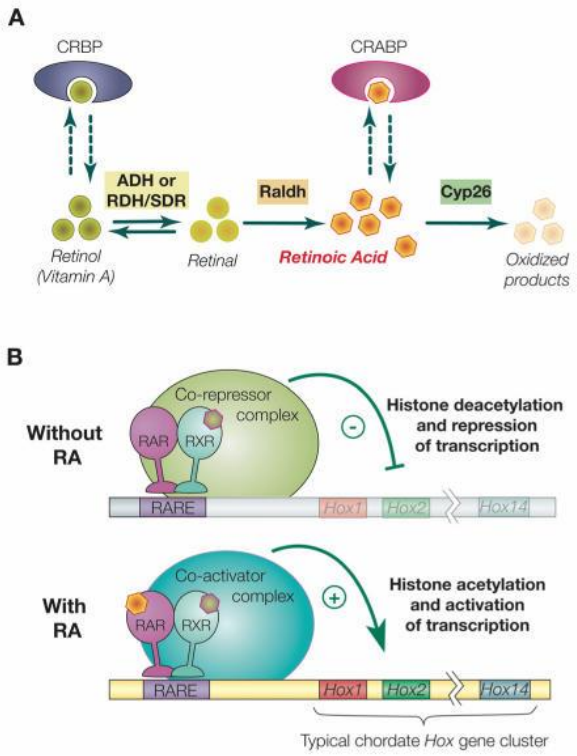


Figure 1.8: Metabolism and mechanism of action of retinoic acid.

__A__) Two steps of oxidation, catalyzed respectively by ADHs/RDHs and RALDHs enzymes, convert retinol (vitamin A) in retinoic acid, which can be further oxidized into inactive metabolites by CYP26s enzymes. __B__) Classical mechanism of retinoic acid signaling activation. RA binding to the RAR-RXR heterodimers induces conformational changes of the receptors, determining the release of co-repressors and the interaction to co-activators, which activates the transcription of target genes. Adapted from Marlétaz et al. 2006.

1.4.1 RA functions during embryonic development

Vitamin A and RA play a critical role during embryogenesis, and the necessity to keep their levels under tight control is demonstrated by the dramatic effects caused by either their deficiency or excess (Duester, 2008). The teratogenic consequences of administering excessive doses of RA to pregnant mice, or the studies on the deficiency of vitamin A have highlighted the involvement of retinoids signaling in the development of several organs including hindbrain, spinal cord, forelimb buds, skeleton, heart, eye, pancreas, lung and genitourinary tract (Clagett-Dame & DeLuca, 2002). The combination of the congenital malformations affecting all these organs and tissues, consequent to vitamin A deficiency, is collectively referred to as the fetal VAD (vitamin A-deficiency) syndrome (Wilson et al., 1953).

The expression of the enzymes responsible for the first step of RA synthesis, ADHs and RDHs, span throughout the embryo often overlapping (Ang et al., 1996; Sandell et al., 2007; Zhang et al., 2001). Null mutations in their coding genes in most of the cases do not result in a clear phenotype, but rather increase the susceptibility to toxicity caused by excess of retinol, suggesting their involvement in the control of retinol levels rather than in normal RA signaling (Molotkov et al., 2002). Nonetheless, *Rdh10* mutants suffer of serious reduction of RA signaling, followed by embryonic lethality within E13 , nominating the activity of this enzyme as critical for the first step of RA synthesis (Sandell et al., 2007). Transgenic mouse lines, bearing a *lacZ* reporter under the control of RAREs reveal activation of RA signaling in the posterior pole of the embryo, where the Primitive Streak lies, from E7.5, and in the trunk between E7.5 and E8.5 (Fig1.9a) (Rossant et al., 1991). The onset of RA activity coincide with the beginning of the expression of *Aldh1a2* (*Raldh2*) at E7.5 in the trunk paraxial mesoderm, and at E8.5 in paraxial and lateral plate mesoderm (Molotkova et al., 2005; Sirbu et al., 2005). Null mutations of *Aldh1a2* (*Raldh2*$^{-/-}$), provoking the absence of RA signaling detection in the reporter mouse line, led to the conclusion that this enzyme is in charge of all RA synthesis at least till E8.5 (Fig1.9b) (Molotkova et al., 2005; Sirbu et al., 2005). After this stage, however, Aldh1a1 and Aldh1a3 also contribute to RA synthesis in the eye and olfactory pit (Molotkov et al., 2006; Molotkova et al., 2007). Throughout the mouse conceptus, the retinoic acid degrading enzymes, Cyp26s, display expression domains complementary to the Aldh1a1-3, with Cyp26a1 being expressed in the tail bud and rostral hindbrain, and Aldh1a2 instead in the trunk. In fact, more than the spatial distributions of synthesizing enzymes, the balance between RA synthesis and degradation is the basis of the uneven distribution of RA within the embryo and of the concentration of active RA in a given tissue (Sakai et al., 2001).

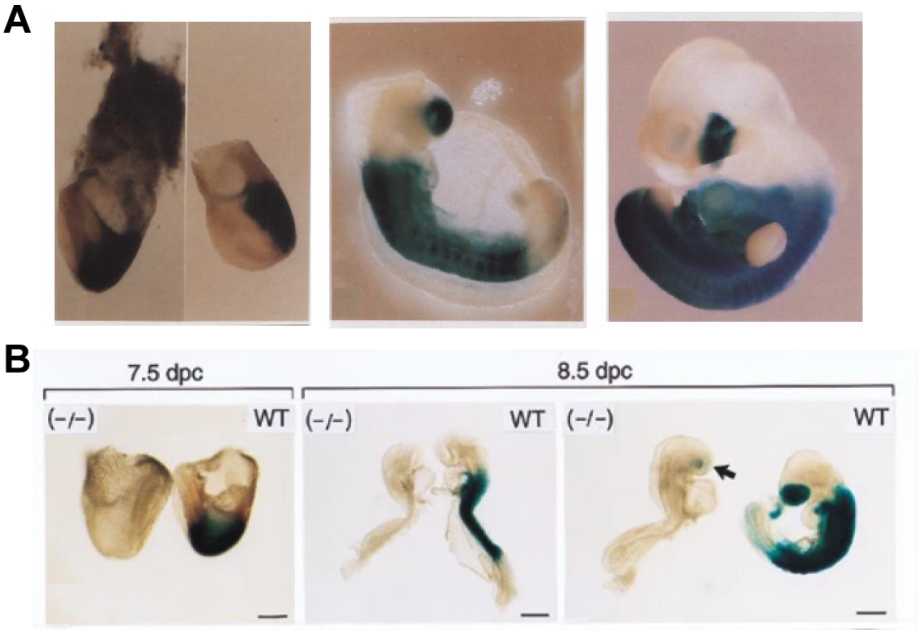


Figure 1.9: RARE_hsplacZ retinoic acid reporter mouse line and Aldh1a2 null embryos.

__A__) Transgenic RARE_hsplacZ embryos stained with X-gal display: RA activity in the posterior region of the embryo at E7.5 (left panel), and in the mid-region of the embryo and in the optic eminence at late E8.5 (9-10 somites) (middle panel); absence of staining anterior to the brain-spinal cord boundary, in the limb buds or tail region at E10.5 (right panel). Adapted from Rossant et al. 1991. __B__) Comparison of wild type and Aldh1a2 KO RARE_hsplacZ embryos at E7.5, early and late E8.5. β-gal staining reveals the absence of RA signaling in the Aldh1a2$^{-/-}$ embryos. Adapted from Niederreither et al. 1999.

1.4.1.1 Functions of the RA receptors during development

An additional layer of regulation of the RA response is the availability of its receptors, the effectors of RA actions during development. Unfortunately, studying the functions of individual RA receptors during development has proven particularly challenging. Their expression, in fact, is widespread and overlapping in early gastrulating vertebrate embryos, and genetic studies showed that combined mutations of multiple receptors are necessary to cause developmental defects typical of retinoid deficiency, implying a high degree of redundancy of their functions (Dollé, 2009; Samarut & Rochette-Egly, 2012). RAR-null compound mutant mice recapitulate some of the developmental abnormalities caused by either vitamin A deprivation or genetically-impaired RA synthesis. Moreover, *Rxrα* mutations alone cause some of the defects typical of the vitamin A deficiency syndrome, whereas mice lacking both *Rxrβ* and *Rxrγ* and even one allele of *Rxrα* do not display any evident morphogenetic

defects, demonstrating that RXRα is functionally the most important RXR during embryogenesis (Krezel et al., 1996; Mark et al., 2009). Simultaneous mutations of RARs and RXRs cause more severe phenotypes than RARs mutations alone (Kastner et al., 1997; Mark et al., 2009), highlighting reduced redundancy between RARs in a RXRα mutant background, and suggesting that heterodimers composed of RAR and RXR are the functional units transducing RA signal during development. Finally, the notion that RARs are the major conductors of RA signaling during embryonic development came by elegant studies conducted in *Aldh1a2*$^{-/-}$ mutant mice, which are characterized by developmental defects typical of RA deficiency and die within embryonic day E8.5. Investigating the possibility to rescue this lethal phenotype, the authors found that only all-*trans*-RA or RAR specific agonists, but not RXRs selective ligands, could revert the defects of *Aldh1a2*$^{-/-}$ mice (Mic et al., 2003).

Although RA receptors are traditionally considered as ligand-activated transducers, gene repression by unliganded receptors was demonstrated as an additional mechanism of action at least in the head formation in *Xenopus*, and skeletal development in mice, processes characterized by the presence of differentiation events inhibited by RA (Weston et al., 2003). The possibility of a repressive role for the RA receptors in other processes is still under investigation, but multiple indications suggest that this might be a much more common mechanism than what currently believed. For instance, the expression of the retinoic acid receptors anticipates any production of RA in the embryo and, globally, their pattern of expression does not always overlap to the spatial and temporal availability of RA. Thus, it is more than legitimate to consider that these receptors play functions beyond the transduction of retinoid signaling, and the teratogenic effects of excessive RA might also be a consequence of de-repression of typically silenced genes.

The more classical functions and well understood mechanisms of action of RA in the hindbrain patterning, trunk, and tail formation are examined in the following chapters.

1.4.1.2 RA and hindbrain patterning

The development of the hindbrain requires the generation of seven to eight neuroepithelial structures called rhombomeres, each with a precise identity dictated by their position along the anteroposterior axis (Kiecker & Lumsden, 2005). Genes

belonging to the Hox family of homeobox transcription factors have pivotal roles in instructing the growth and positional identity of specific rhombomers (Rijli et al., 1998). RA controls multiple aspects of the hindbrain pattering, and exposure to excessive RA at late gastrula stages causes expansion of the hindbrain at the expenses of more anterior regions (forebrain and midbrain) (Avantaggiato et al., 1996). Conversely, vitamin A deficiency (VAD) or *Aldh1a2* KO cause misspecification or absence of rhombomeric boundaries and, overall, a reduction of the posterior hindbrain (Niederreither et al., 2000). In normal conditions, RA generated by Aldh1a2 in the somitic mesoderm diffuses towards the hindbrain, where it regulates the expression of posterior rhombomeric determinants, such as the Hox genes (Rhinn & Dolle, 2012). It was initially hypothesized that the establishment of the correct identity of the rhombomeres along the antero-posterior axis was dependent on the establishment of a RA gradient, higher in the caudal hindbrain and decreasing anteriorly. However, studies in Cyp26s mutant mice made clear that, by limiting the RA activity, the differential rhombomere-specific expression pattern of these enzymes is essential in defining the boundaries of the rhombomeres (Hernandez et al., 2007; Sirbu et al., 2005). In other words, it was noticed that the posterior limit of Cyp26s activity determines the anterior boundary of expression of the genes specific of that rhombomere, conferring to Cyp26s a cell-autonomous mechanism of action in establishing the RA the gradient (Hernandez et al., 2007; Sirbu et al., 2005).

The hindbrain patterning is considered a well understood paradigm of the instructive mechanism of action of RA, operating by regulation of the *Hox* genes, but RA may also act in a permissive fashion in non-cell autonomous way, influencing trunk development and axial elongation by the NMPs (Martin & Kimelman, 2010).

1.4.1.3 RA in trunk and tail

RA participates in the trunk formation by delimiting *Fgf8* expression and generating an Fgf8-free zone, between the caudal epiblast *Fgf8* and the *Fgf8* in the posterior mesoderm of the cardiac field, where the trunk can develop (Sirbu et al., 2008; Sirbu & Duester, 2006). The repression of Fgf8 by RA ensures proper differentiation of the spinal cord and somitogenesis during body axis extension (Del Corral et al., 2003; Sirbu & Duester, 2006). RA in this context has been considered as a trunk organizer, antagonizing Fgf8 signaling along the anteroposterior axis and

allowing proper differentiation of the trunk progenitors. Accordingly, vitamin A deficient or *Aldh1a2*$^{-/-}$ mice exhibit reduced somite size and axis shortening, with concomitant anterior expansion of the caudal epiblast *Fgf8* domain, and posterior expansion of cardiac *Fgf8* domain (Fig1.10a) (Duester, 2008).

In the tail bud, instead, Brachyury (T) induces Cyp26a1 expression to prevent RA produced in the somites from spreading posteriorly and disrupting the Brachyury-Wnt autoregulatory loop underlying the NMP niche (Martin & Kimelman, 2010). When RA signaling is not limited, as in the absence of Cyp26a1 activity (*Cyp26a1*$^{-/-}$), the embryos are affected by severe caudal truncation with mermaid-like deformity (sirenomelia), closely resembling the effects of either administration of teratogenic dose of RA, or of the disruption of T gene (Fig1.10b). This phenotype confirms the importance of limiting the exposition of the NMP niche to RA for correct axial elongation (Iulianella et al., 1999; Sakai et al., 2001).

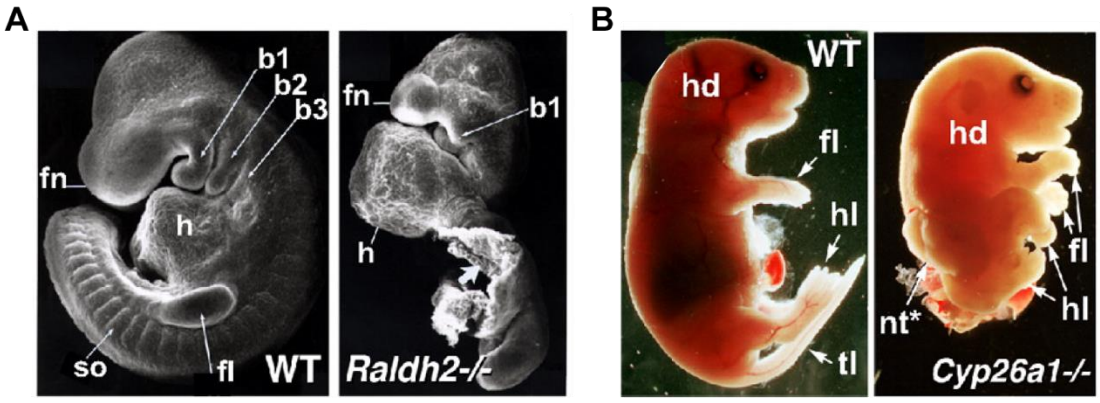


Figure 1.10: Phenotypes of Aldh1a2 (Raldh2) and Cyp26a1 null mice.

***A**) E9.5 Aldh1a2*$^{-/-}$ *mutant embryos show several abnormalities, including hypoplasia of the frontonasal region (fn), absence of posterior branchial arches (b1-b3), impaired heart (h) morphogenesis, shortened trunk (bold arrow) with compacted somites (so), and lack of forelimb buds (fl). **B**) E18.5 Cyp26a1*$^{-/-}$ *mutant embryos exhibit an open neural tube (nt*) and severe truncation of the posterior part of the body with abnormal positioning of the hindlimbs (hl). Adapted from Rhinn and Dollé 2012.*

To investigate the non-cell autonomous mechanism of action of RA in zebrafish, a transgenic RA sensor line was transplanted in either a WT host or into a host expressing a RA-dependent hyperactive RAR, which causes upregulation of Cyp26a1 in every cell. In the first case the donor cells exhibited robust RA reporter expression, whereas in the second Cyp26a1 expression prevented any RA signaling activation.

However, when the donor cells were transplanted into a host bearing hyperactive RAR, but devoid of Cyp26a1 enzyme, the reporter was again expressed at normal levels. The finding argues in favor of a non-autonomous control of Cyp26 on RA availability, by acting as a sink that removes RA from the local environment and limiting the exposure of neighboring cells to RA (Rydeen et al., 2015).

The treatment of pluripotent stem cells with RA levels above 100nM is a well consolidated and effective way to induce neural fate *in vitro* (Cai & Grabel, 2007). However, indications coming from different studies led to retain that RA is not required for neural induction in the mouse embryos. Among the main arguments in this direction is the fact that RA synthesis is considered to begin at E7.5, just after induction of forebrain and midbrain fates (Sirbu et al., 2005). Moreover, RA produced in the somitic mesoderm (E7.5-8.5) is believed to impact the development only of posterior neural tissues, with loss of RA signaling not affecting the expression of neural induction markers *Sox1* and *Sox2* (Molotkova et al., 2005). In particular, the investigation of the phenotypes of *Aldhs* mutant embryos, led to conclude that the role of RA till around E10.5 is in the posterior neuroectoderm (hindbrain and spinal cord) and eye, but not the forebrain itself (Fig1.10a) (Molotkova et al., 2005). Owing to these and other evidences, RA is considered a posteriorizing signal together with Wnt, as hypothesized by the "activation-transformation" model of neural induction (Duester, 2008).

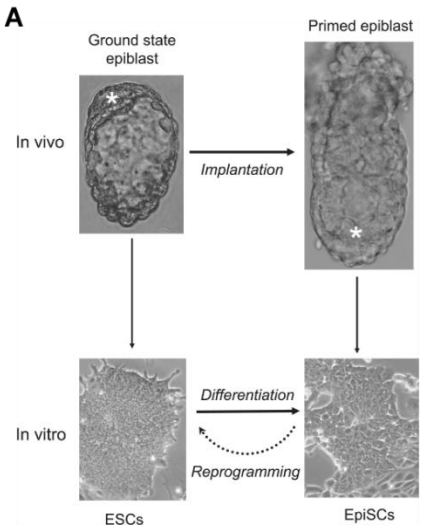
1.5 Mouse Embryonic Stem Cells as a model system

Although *in vivo* genetic loss-of-function studies provided invaluable insights into the molecular underpinnings of gastrulation, and the technology advancements in the last years are returning more and more complete pictures of the evolution of the transcriptome during early stages of development by single-cell RNA sequencing, some limitations have still to be overcome, such as the accessibility and scarcity of the populations of interest within the embryo. On a more theoretical level, given that multiple cues in the embryonic microenvironment can control different cell fates often also interdependent, defining the role of single pathways or determining "sufficient" conditions for a given process is extremely challenging (Nishikawa et al., 2007). For these reasons, since their establishment in culture by Martin and Evans (Evans & Kaufman, 1981; G. R. Martin, 1981), mouse embryonic stem cells (mESCs) have

represented an alternative system to address developmental biology questions. Derived from the ICM of E3.5 blastocyst, mESCs retain, in fact, key features of pre-implantation naïve epiblast. They can be expanded and maintained indefinitely in an undifferentiated state, a property defined as self-renewal, while preserving their pluripotency, the ability to give rise to cells of any germ layer. Other hallmarks of naïve pluripotency of mESCs are the possibility to contribute to chimera formation when re-injected into pre-implantation blastocyst, genome-wide DNA hypomethylation and reactivation of the X chromosome in females (Martello & Smith, 2014). However, the possibility to proliferate while keeping a pluripotent nature is maintained only transiently *in vivo* and, therefore, mESCs requires artificial conditions to preserve this nature, as culturing the cells on mitotically-inactivated mouse embryonic fibroblast (MEFs) and supply of fetal calf serum (FCS) to the media. The discovery that the secretion of an IL-6 like cytokine, the Leukemia Inhibitory Factor (LIF), was the main function of the feeder layer, and that BMP4 was a main component of serum inhibiting differentiation (Ying et al., 2003), allowed the formulation of synthetic versions of the media. Additionally, it was demonstrated that it was possible to maintain the cells in a "ground state" of pluripotency by the inhibition of the MAPK cascade through a MEK inhibitor (PD184352), and Wnt activation by blocking GSK3β with CHIR99021, a condition defined as 2i (two inhibitors) (Ying et al., 2008). Recent studies highlighted, however, that prolonged culture of the cells in this cocktail of inhibitors increases chromosomal aberrations (Choi et al., 2017).

Many years after the establishment of the Inner Cell Mass in culture, it became also possible to stabilize *in vitro* cells from post-implantation egg cylinder epiblast at E5.5 (EpiSCs) (Brons et al., 2007; Tesar et al., 2007). These cells exhibited some features of mESCs as self-renewal and pluripotency, giving rise to any of the three germ layers derivatives, but with some limitations (Fig1.11). In fact, EpiSCs cannot contribute to chimera formation when injected in pre-implantation blastocysts, highlighting that they represent a later developmental stage compared to mESCs (Guo et al., 2009; Rossant, 2008; Tesar et al., 2007). Additionally, they present globally higher levels of DNA methylation and silencing of one of the X chromosomes in females (Martello & Smith, 2014; Nichols & Smith, 2009). The pluripotent state of EpiSCs is defined "primed" to reflect its predisposition for differentiation, due to the expression of lineage priming factors, as compared to the broader "naïve" pluripotency

of mESCs (Fig1.11) (Nichols & Smith, 2009). Proof of the fundamentally different nature of the "primed" pluripotency is the distinct signaling environment on which it relies, including activation of Nodal/TGF-β and FGF pathways, respectively by Activin and FGF2, and independence from LIF signaling. Noteworthy, human ESCs display the hallmarks of pluripotency and self-renewal, whilst sharing many of the features of mouse EpiSCs just mentioned, including the requirement for TGF-β and FGF to maintain pluripotency and globally higher DNA methylation. They are thus considered a more advanced stage of pluripotency than mouse ESCs and, generally, they behave as mEpiSCs in response to the same differentiation stimuli (Hanna et al., 2010; James et al., 2005; Vallier et al., 2005, 2009).



B

Property	Ground State	Primed State
Embryonic tissue	early epiblast	egg cylinder or embryonic disc
Culture stem cell	rodent ESCs	rodent EpiSCs; primate "ESCs"
Blastocyst chimaeras	yes	no[a]
Teratomas	yes	yes
Differentiation bias	none	variable
Pluripotency factors	Oct4, Nanog, Sox2, Klf2, Klf4	Oct4, Sox2, Nanog
Naive markers[b]	Rex1, NrOb1, Fgf4	absent
Specification markers	absent	Fgf5, T
Response to Lif/Stat3	self-renewal	none
Response to Fgf/Erk	differentiation	self-renewal
Clonogenicity	high	low
XX status	XaXa	XaXi
Response to 2i	self-renewal	differentiation/death

[a] Not applied to primate cells.
[b] Representative examples.

Figure 1.11: Naïve and primed state of pluripotency.

__A__) Pre-implantation Inner Cell Mass (E4.5) and post-implantation epiblast (E5.5) are the counterparts respectively of mESCs characterized by naïve pluripotency, and mEpiSCs which represent a pluripotent state primed for lineage specification. __B__) Comparison of the most important features of naïve and primed pluripotency. Adapted from Nichols and Smith 2009.

1.5.1 Naïve Pluripotency Network

The central core fueling pluripotency is the network of transcription factors Oct4, Sox2 and Nanog. Oct3/4 (*Pou5f1*), the first pluripotency factor to be identified, belongs to the family of POU-domain transcription factors (Okamoto et al., 1990; Schöler et al., 1990). In its absence, pluripotency is lost both *in vivo* and *in vitro*, and Inner Cell Mass and mESCs differentiate into trophectoderm (Nichols et al., 1998; Niwa et al., 2000). Surprisingly, its overexpression does not stabilize the self-renewal

and causes instead differentiation (Niwa et al., 2000), suggesting that Oct4 might act as a lineage-priming factor besides being essential to maintain pluripotency, and its expression has to be tightly regulated (Thomson et al., 2011). Indeed, its levels are kept relatively uniform throughout epiblast formation and maturation. The best characterized Oct4 partner is Sox2 (Sex determining region Y-box 2), a member of the SoxB1 transcription factor family. As for Oct4, Sox2 inactivation results in trophoblast formation, and its overexpression causes differentiation of ESCs, underpinning a role as a lineage-priming factor (Masui et al., 2007). Its expression becomes, in fact, restricted to the neural lineage during gastrulation. Sox2 positively regulates the transcription of Oct4, and the two proteins cooperatively bind regulatory elements of target genes. Among them is the other core pluripotency factor, Nanog. Nanog is almost exclusively expressed in the Inner Cell Mass, where its expression demarks the cells that will become naïve epiblast as opposite to the primitive endoderm cells expressing GATA6 and, after being transiently downregulated, it becomes apparent again in the egg cylinder epiblast (Silva et al., 2009). Differently from the other two factors, its forced expression allows ESCs to self-renew in absence of LIF (Chambers et al., 2003), whereas its inactivation impairs epiblast formation in the embryo but not *in vitro* (Chambers et al., 2007; Silva et al., 2009). Beside self-sustaining the expression of each other with positive feedback loops, the three core factors regulate the expression of other exclusive members of the naïve pluripotency network such as *Esrrb*, *Klf2*, *Klf4* and *Tbx3* (Fig1.12) (Chen et al., 2008; Marson et al., 2008; Young, 2011).

The main extrinsic cue sustaining the pluripotency network is LIF signaling. Most of the functions exerted by LIF in this context need the activation of Stat3 (Fig1.12). In fact, the forced expression of Stat3 allows ESC to self-renew in the absence of LIF (Matsuda et al., 1999), whereas its inhibition leads to differentiation (Bourillot et al., 2009). The interaction of LIF with its receptor triggers Stat3 phosphorylation, which then forms dimers and translocate into the nucleus where it controls the expression of target genes (Bourillot et al., 2009; Zhong et al., 1994). Among the direct targets of LIF/Stat3, critical for pluripotency maintenance, are *Klf4*, *Gbx2*, *Pim1* and *Tfcp2l1* (Bourillot et al., 2009; Martello et al., 2013; Niwa et al., 2009). The last transcription factor, in particular, is considered a central mediator of LIF function because its expression is able to recapitulate LIF effects on self-renewal,

and its downregulation impairs the response to LIF (Martello et al., 2013; Ye et al., 2013).

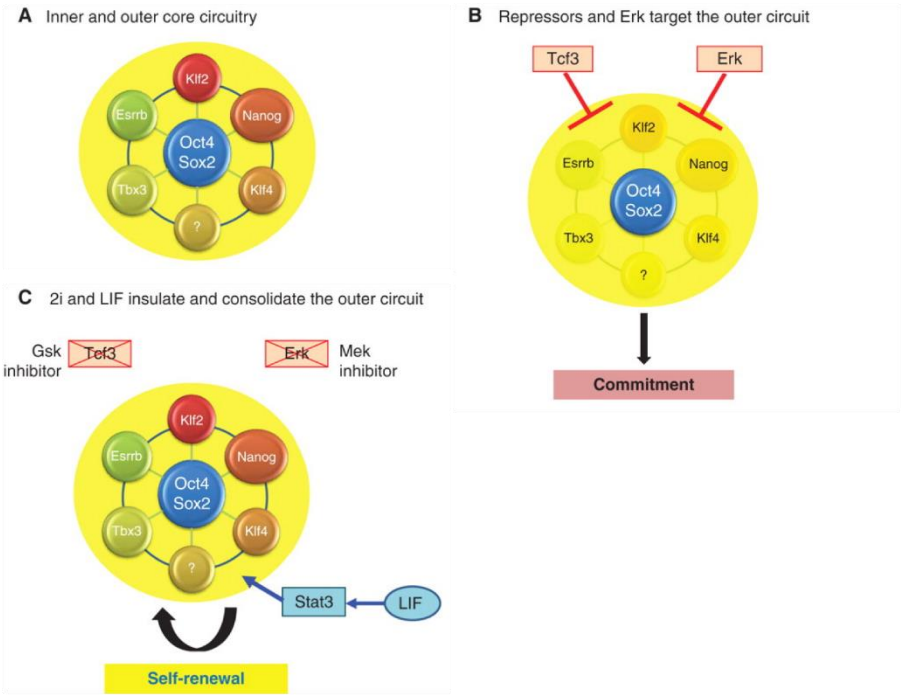


Figure 1.12: The core network of pluripotency factors.

__A__) Network of transcription factors sustaining naïve pluripotency. Oct4 and Sox2 are the central hubs of the network of these factors cross-regulating one another. __B__) Extrinsic cues causing, for example, inhibition of Wnt targets by Tcf3, or MAPK/Erk activation, poison the outer circuit of pluripotency factors leading to lineage commitment. __C__) The 2i prevent external interference on the pluripotency network, whereas LIF/Stat3 enhances the expression of some of the pluripotency factors, both resulting in the stabilization of naïve state. Adapted from Nichols and Smith 2012.

1.5.2 Exit from Pluripotency

At the time of implantation, the naïve pluripotency program is abandoned in less than 24 hours, preparing for the formation of the egg cylinder epiblast and subsequent gastrulation. Although a precise plan for dismantling the pluripotency network would seem more logic, given the rapidity of the process, multiple independent observations propose a stochastic beginning of differentiation, associated with fluctuations in expression of transcription factors like Nanog, Stella, Klf4 and Rex1 (Chambers et al., 2007; Hayashi et al., 2008; Kalmar et al., 2009; Toyooka et al., 2008) or miRNAs as the miR-142 (Sladitschek & Neveu, 2015). These genes exhibit heterogenous

expression in serum and LIF culture, being continuously downregulated in a group of cells and upregulated in others, a state often referred to as metastable (Chambers et al., 2007; Silva et al., 2008; Sladitschek & Neveu, 2015; Macarthur & Lemischka, 2013). The 2i condition, on the contrary, renders the population more uniform, and most of the cells exhibit a robust expression of ground state factors (Dunn et al., 2014) and globally lack lineage markers, likely because of the removal of differentiation-primed cells (Marks et al., 2012; Trott & Martinez Arias, 2013). Nevertheless, cells in 2i do not manifest overall delay in differentiation *in vitro* and *in vivo* (Martello & Smith, 2014). This implicates that the ESC heterogeneity, observed when cells are exposed to serum, might be an epiphenomenon of the response to incoherent environmental inputs rather than being an intrinsic characteristic of pluripotency (Kalkan & Smith, 2014; Smith, 2013). It is currently accepted, however, that culturing the cells in different conditions generates a continuum spectrum of pluripotent states, going from 2i regime to serum and LIF condition, and where the cells sit along this continuum at the beginning of differentiation might influences their lineage propensity afterwards (Hackett et al., 2017).

When the factors sustaining self-renewal are withdrawn, ESC cells progress from naïve pluripotency to lineage specification, driven by autocrine signals. The downregulation of ground state factors starts immediately, but for approximately 24 hours the process is still reversible, and the cells can self-renew if re-exposed to LIF. The loss of the ability to revert between 24 and 72 hours is marked by the disappearance of the marker Rex1 (*Zfp42*) (Fig1.13). It was suggested that cells cultured in the presence of serum might sporadically undergo lineage specification and reach this transitional state, but the presence of LIF forces them to go back (Martello & Smith, 2014). LIF plays, in fact, a critical role in the reversion, most likely in the same way in which it also helps the reprogramming of somatic cells into iPSCs (Yang et al., 2010). A suggested mechanism for its ability to establish/re-establish the naïve pluripotency network is the induction of the expression of the transcription factors *Klf4* and *Tfcp2l1* (Guo et al., 2009; Martello et al., 2013; Ye et al., 2013).

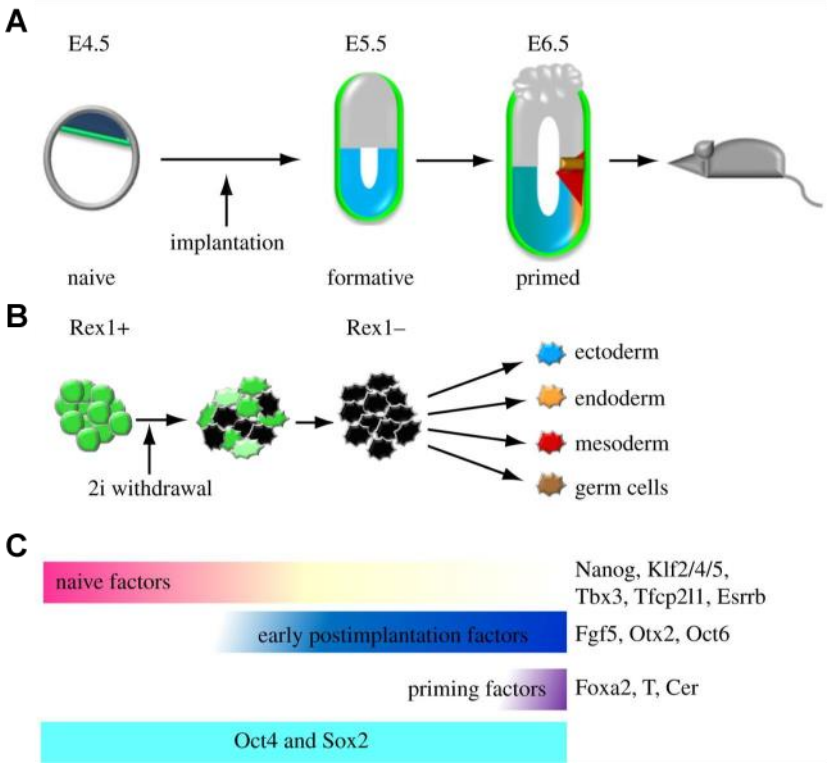


Figure 1.13: The evolution of the pluripotent states.

__A__) Progression of pluripotency in parallel with mouse development before lineage priming. __B__) The exit from naïve pluripotency upon 2i withdrawal, marked by the loss of Rex1 expression in asynchronous fashion, might reflect the restriction of pluripotency occurring in vivo after implantation. __C__) Different dynamics of naïve, early post-implantation and lineage priming factors during the abandoning of the pluripotent state. Adapted from Kalkan and Smith 2014.

The rapid dissolution of the naïve network upon LIF or 2i withdrawal suggests that it might be intrinsically poised to collapse by factors already present in self-renewal conditions (Fig1.14). For instance, *Fgf4* is a direct target of Oct4 and Sox2 complexes and FGF4/Erk signaling is able to downregulate *Nanog* expression (Ambrosetti et al., 1997; Lanner et al., 2010; Nichols et al., 1998; Silva et al., 2009). *Tcf3* is also directly regulated by Oct4 and it can repress key pluripotency factors such as *Esrrb*, *Klf2*, *Nanog* and *Tfcp2l1* (Guo et al., 2011; Pereira et al., 2006; Yi et al., 2008). Of note, *Tcf3* deletion makes ESCs resistant to differentiation (Guo et al., 2011; Pereira et al., 2006). *Esrrb* and *Tfcp2l1* are targets of multiple putative pluripotency disassembling factors, hinting that they could be pillars keeping the ground state network together. Accordingly, simultaneous knock down of both genes accelerate the exit from pluripotency. It is likely that the irreversible transition out of the ground state requires stable silencing of pluripotency associated loci, and this task can be carried

out by de novo DNA methyltransferases as Dnmt3a and Dnmt3b, which are upregulated during the transition to primed state (Martello & Smith, 2014).

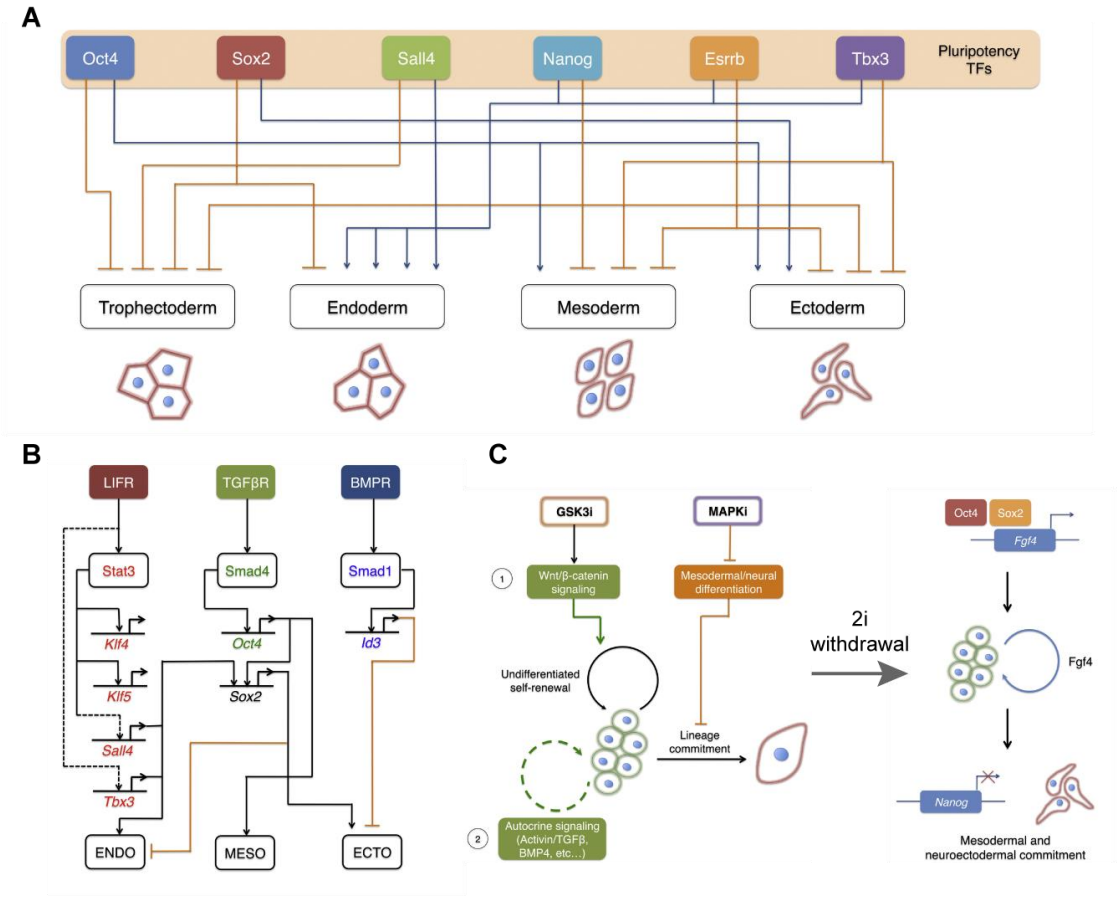


Figure 1.14: Lineage priming by intrinsic poisoning of the pluripotency network.

__A__) Each pluripotency factor can regulate in a positive or negative way the specification towards a certain lineage. However, the cross-regulation of the pluripotency factors balances the individual lineage-specific programs, resulting in no net commitment to any lineage. __B__) The cross-inhibition of the programs primed by each transcription factor is reinforced by extrinsic cues. __C__) Poisoning of the pluripotent state by autocrine signals, such as the production of Fgf4, might lead to the exit from pluripotency upon removal of 2i or LIF and serum. Adapted from Loh and Lim 2011.

Single cell gene expression studies shed more light on the mechanism of the transition from pluripotency to lineage specification. At the population level the termination of naïve pluripotency appears gradual and even reversible in the first stages because of asynchronous single cell dynamics (Kalkan et al., 2017; Trott & Martinez Arias, 2013). Considering individual cells, it is clear that each pluripotency factor follows a characteristic downregulation kinetic, and as soon as the expression

of a set of strictly naïve factors is lost, demarked by the extinction of Rex1 expression, the ground state is no longer accessible. At the same time, "primed" post-implantation epiblast genes are upregulated and, only at the end of this transition, lineage specific markers can be detected, implying the possibility to start new programs (Acampora et al., 2016; Boroviak et al., 2015; Kalkan et al., 2017). However, some reports highlight that individual core pluripotency factors are retained in subsets of cells during the transition and are later co-expressed with lineage specific markers: Oct4 and Nanog are often associated with the primitive streak marker Brachyury, whilst *Sox2* is expressed with *Sox1* (Trott & Martinez Arias, 2013). The finding is in agreement with the notion that Sox2 represses the Primitive Streak fate, whereas Oct4 inhibits neural fate (Thomson et al., 2011). In these studies, the markers of different lineages, such as Brachyury and Sox1, were mutually exclusive, arguing against the possibility of multilineage priming during the loss of pluripotency (Trott & Martinez Arias, 2013).

1.5.3 Germ layers formation in vitro

Many different protocols have been implemented to recapitulate the formation of germ layers *in vitro*, with the bivalent purpose of increasing the understanding of the mechanisms underlying this process and producing multiple tissues for cell therapy applications. The culture regimen adopted for this purpose, often rely on the perturbation of multiple signalling pathways, found to be involved in the process by developmental biology studies or by empirical observations. Equally often, a comprehensive picture of the consequences of triggering these signals on the transcriptional network executing the decision has yet to be achieved. In parallel with the efforts to improve the differentiation conditions, many mESC reporter lines have been generated introducing sequences coding for fluorescent proteins inside loci expressed early during the specification of a certain lineage (Fehling et al., 2003; P. Gadue et al., 2006; Ng et al., 2005; Tada et al., 2005; Ying et al., 2003). Given the inevitable heterogeneity of the response to the differentiation stimuli, the reporter line approach offers the advantage to isolate more homogeneous populations, sharing the expression of the gene controlling the reporter.

1.5.3.1 Primitive streak and endoderm/mesoderm formation

In vivo, distinct mesoderm subpopulations derive from different regions of the PS, whereas definitive endoderm arises from anterior Primitive Streak. Whilst in absence of Wnt or Nodal/TGF-β no PS is generated, it is generally agreed that Wnt favors the development of posterior PS and mesoderm derivatives, and high level of Nodal/Activin signaling are instead necessary for anterior structures (definitive endoderm). Multiple studies using ESCs have confirmed these observations also *in vitro*, adopting the expression of *Brachyury* as proxy for the formation of PS-like structures and later of mesoderm (Fig1.15) (Murry & Keller, 2008). Wnt activation enhances the formation of PS populations and cardiac mesoderm (Turner, Rué, et al., 2014; Ueno et al., 2007), whereas its inhibition blocks the emergence of cells expressing *Brachyury* (Lindsley et al., 2006; Naito et al., 2006; Turner, Trott, et al., 2014). Regarding the role played by Nodal/Activin/TGF-β signaling, it has been reported that elevated concentrations of Nodal induce anterior/definitive endoderm cells, whereas low levels favor the formation of mesoderm/posterior tissues (Gadue et al., 2006; Kubo et al., 2004; Yasunaga et al., 2005). Using reporters driven by the anterior PS markers *Foxa2* or *Goosecoid* in combination with a reporter for *Brachyury*, independent groups observed the emergence of cells expressing the anterior genes in Nodal/Activin-induced cultures, including cells able to generate both endoderm and mesoderm (Gouon-Evans et al., 2006; Tada et al., 2005). Notably, cells fated to become posterior PS, when exposed to high levels of Activin are converted into endoderm cell types, indicating the existence of a certain degree of plasticity of the fates at this stage (Fig1.15). One important conclusion deriving from these findings is the difficulty in clearly defining when a fate is fully acquired relying solely on the expression of one or two markers. If taken individually Wnt and Nodal/TGF-β pathways display the capacity to promote different descendants of the PS, the inhibition of one of the two, *in vitro*, prevents the induction of the PS fate by the other, suggesting the presence of interplay between the two signaling (Gadue et al., 2006;Turner, Trott, et al., 2014).

On the other hand, genes like Flk-1 (*Kdr*) and PDGFR have been widely used to monitor the emergence of mesoderm fate from the PS (Ema et al., 2006; Kataoka et al., 1997). When added to differentiating cultures in the absence of serum, BMP4

contributes to the formation of a Brachyury^{+} PS population, and Flk-1^{+} hematopoietic mesoderm. Nevertheless, BMP is dispensable for the generation of the PS, and its actions in this context are considered mediated by Wnt and Nodal (Ng et al., 2005; Nostro et al., 2008; Park et al., 2004).

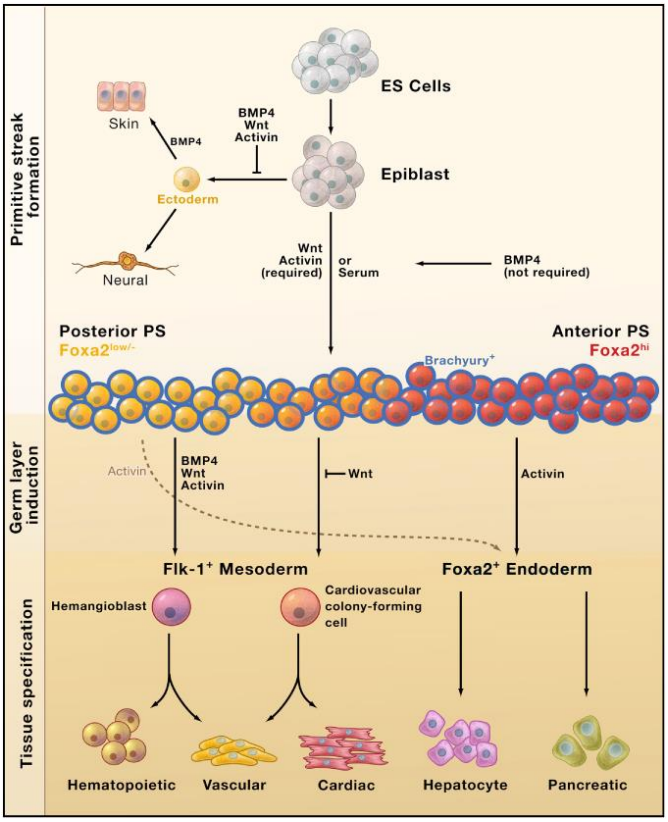


Figure 1.15: ESCs differentiation and formation of Primitive Streak derivatives.

Schematic representation of mESCs differentiation to Primitive Streak and its derivatives via epiblast-like state. The levels of Activin/Nodal signaling are considered crucial for the decision of the Primitive Streak derivative to induce: high dose of Nodal/Activin leads to the formation of anterior PS derivatives/endoderm, whereas low levels favor the formation of posterior fates/mesoderm. As hypothesized by the default model of neural induction, the ectoderm arises in absence of the signals promoting the specification towards the PS. Adapted from Murry and Keller 2008.

Detailed experiments have shown that *in vitro* differentiations not only reproduce the transcriptional changes typical of the PS formation, but also other key features of this structure as the Epithelial-Mesenchymal Transition (EMT) (Turner, Rué, et al., 2014). Starting from the characteristic dome-shaped pluripotent colonies, under the influence of Activin or the Wnt agonist, CHIR, cells form epithelial monolayers reminiscent of the post-implantation epiblast, and later adopt a

mesenchymal-like morphology, starting to migrate in concomitance with the onset of *Brachyury* expression. Downregulation of E-cadherin, typical of the pluripotent cells, rearrangement of the actin cytoskeleton, increase of the expression of fibronectin basally and formation of lamellipodia and filopodia are only examples of the profound structural modification that the cells undergo *in vitro*. As cells become migratory, they also lay down an extracellular matrix, mainly composed of fibronectin, that helps their migration (Turner, Rué, et al., 2014).

Chemically derived cell-permeable small molecules offer many advantages over conventional ligands activating signal transduction, in particular in the perspective of clinical applications, such as the possibility to be synthesized in high amount and purity, and greater control on the timing of action. In the attempt to identify small molecules enabling endoderm differentiation of mouse and human ESC more efficiently than Activin/Nodal, two products of de novo synthesis were discovered, named IDE1 and IDE2 (Borowiak et al., 2009). Both molecules alone induced the differentiation towards endoderm of 70-80% of the total cells, whereas the combination of the two had no synergistic effect. Beside inducing the expression of a reporter controlled by the *Sox17* promoter, marker of definitive and extraembryonic endoderm, the developmental potential of IDEs-induced cells has been tested *in vivo* and *in vitro* (Borowiak et al., 2009). After six days of IDE1 differentiation, the cells injected into the developing gut tube, an endoderm derivative, were able to integrate and expressed appropriate markers. Moreover, when IDE1 was used instead of Activin A to start the endoderm differentiation, it increased by four times the yield of pancreatic progenitors obtained upon further differentiation, confirming the capacity of the small molecules to originate functional cell types. Even though the mechanism of action of these compounds has still to be clarified, the induction of Smad2 phosphorylation, and the possibility to block their effects inhibiting Activin/Nodal-specific TGF-β receptors, strongly support activation of the TGF-β pathway (Borowiak et al., 2009).

1.5.3.2 Neuroectoderm formation: Default mechanism?

The ectoderm germ layer, deriving in large part from the anterior epiblast regions which do not contribute to the PS formation, originates both neuroectoderm and skin tissues. Similarly to *in vivo,* the mechanism of ectoderm induction from ESCs is often

referred to as “default”, because it is observed in conditions devoid of inhibiting factors present in the serum and of exogenous inductive stimuli (Murry & Keller, 2008; Ying et al., 2003). The presence of BMP, Wnt and activin/Nodal in the media, indeed, repress the formation of neuroectoderm (Aubert et al., 2002; Kubo et al., 2004; Ying et al., 2003). However, studies adopting a knock-in line bearing a GFP reporter under the control of the specific neuroectoderm marker *Sox1*, proposed that the formation of neuroectoderm relies on the endogenous production of FGF signals, adding a new perspective to the “default” model (Ying et al., 2003). In apparent agreement with this finding, pharmacological inhibition of Mek by PD184352 (one of the two inhibitors of the 2i cocktail) blocks neuroectoderm specification (Stavridis et al., 2007). Yet, mESCs treated with SU5402, an inhibitor of the Fgfr1 receptor tyrosine kinase activity, or even *fgfr1* mutant cells are still able to differentiate down the neural lineage (Smukler et al., 2006). This leaves open the possibility that neural induction might require MAPK activation through a mechanism independent of FGF receptor; Insulin-like growth factors (IGFs), for instance, can activate the MAPK cascade and are also constant components of the N2B27 neural induction medium (Gaulden & Reiter, 2008). An alternative explanation proposed for the role of FGF is that this signal might be required to reach the primed state of pluripotency, enabling the competence for germ layer differentiation, rather than neural induction *per se* (Kunath et al., 2007; Stavridis et al., 2007). An attempt to reconcile the requirement for FGF activation and BMP inhibition in neural induction, comes from the report that the FGF induced MAPK cascade reduces the activation of BMP signaling promoting SMAD1 degradation (Pera et al., 2003). Nevertheless, it was showed that mutation of the MAPK phosphorylation site of SMAD1 did not prevent neural induction in mouse embryo, arguing against the role of this interaction as unique action of FGF during neural fate acquisition (Aubin et al., 2004).

Adding new emphasis to the “default” mechanism, small molecules inhibitors of BMP/TGF-β were shown to promote neural commitment of mouse EpiSCs (Najm et al., 2011). Similarly, the inhibition of both Activin/Nodal-SMAD2/3 and BMP-SMAD1/5/8, by respectively SB431542 and Noggin, proved a very effective system to induce neural fate from human ESC, counterparts of mouse EpiSCs (Chambers et al., 2009). Cells induced with such protocols acquired an anterior neuroectoderm identity, expressing *Otx2* and *Foxg1*, and could be further converted in multiple central

nervous system cell types (Ozair et al., 2013). Further evidence of the connection between general TGF-β inhibition and neural induction derives from the preferential acquisition of neural fate of mESCs in which SMAD4, the co-SMAD partnering with both SMAD2/3 and SMAD1/5/8, is absent (Sonntag et al., 2005). This and other studies highlight the neural inducing properties of blocking Nodal signaling in addition to BMP (Fig1.16a). For instance, mESCs lacking the Nodal co-receptor *Cripto*, or hESCs overexpressing the Nodal inhibitors *Lefty* or *Cerberus*, and *Nodal* mutant mice, exhibit precocious or increased propensity towards neural specification (Camus et al., 2006; Smith et al., 2008; Sonntag et al., 2005; Vallier et al., 2004). Some of the mechanisms proposed for the neural differentiation by TGF-β inhibition involve preventing the access to other fates or releasing from repression neural priming factors such as Zeb2, Nr2f2 and Zfp521. Regarding the second mechanism, in the pluripotent state these genes are under TGF-β/SMAD-sustained Nanog and Oct4 negative regulation. Therefore, Nodal inhibition results in Nanog downregulation and release from repression of the factors, which in turn block Oct4 while inducing neural-specific genes (Chng et al., 2010; Kamiya et al., 2011; Rosa & Brivanlou, 2011).

The function of Wnt signaling in the neural specification scenario is debated, with reports sustaining a repressive action over neuroectoderm formation and others arguing for the necessity of its activation (Aubert et al., 2002; Kan et al., 2004; Otero et al., 2004). Much of this inconsistency might be explained by considering the many and even opposite context-dependent functions that Wnt exerts in naïve, primed pluripotency and lineage specification. With this in mind, multiple studies agree on a positive effect of Wnt inhibition by Dkk1 or secreted receptors (sFRPs), over the acquisition of neural fate from mESCs (Fig1.16a) (Aubert et al., 2002; Verani et al., 2007). Moreover, other studies shed light on a cross-talk between BMP, FGF and Wnt actions, identifying in SMAD1 a crucial node for the integration of the information encoded by these pathways (Eivers et al., 2008). It has been postulated that the level of BMP controls the intensity of Smad1 activation, whereas MAPK and GSK3 regulate the duration of the signal by phosphorylating the linker region of Smad1, triggering its degradation. The phosphorylation by the BMP receptor of the C terminal domain of Smad1, which determines its activation, is the trigger itself of the double linker phosphorylation, causing the polyubiquitination of Smad1 by the E3 ubiquitin ligase Smurf1 and its degradation (Fig1.16b) (Fuentealba et al., 2007).

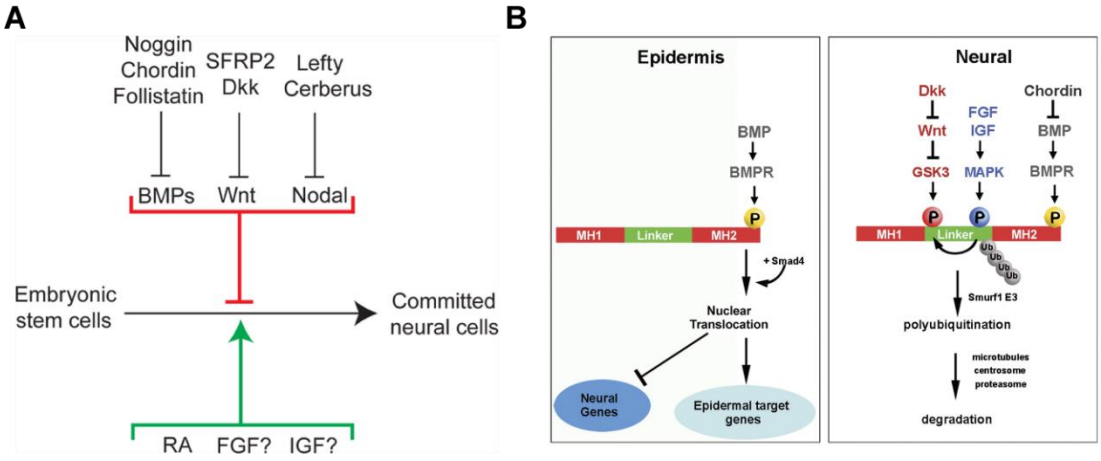


Figure 1.16: Multiple mechanisms of induction of the neural fate.

__A__) Many secreted inhibitors of BMP, Wnt and Nodal, responsible of the induction of neural fate in vivo, are also able to promote the differentiation of ESCs to neuroectoderm. On the contrary, it is still debated whether retinoic acid, a potent inducer of neural fate from ESCs, is also important for neural induction in vivo. Adapted from Gaulden and Reiter 2008. __B__) Proposed model including Wnt inhibition and FGF signaling in the default mechanism of neural induction by BMP pathway inhibition. The phosphorylation of the Smad1 linker region by GSK3 and MAPK, in fact, triggers its degradation and the inactivation of BMP signaling, with the consequently favored acquisition of the neural fate. Adapterd from Eivers, Fuentealba, and De Robertis 2008.

1.5.3.3 Role of RA in neuroectoderm induction

One of the greatest discrepancies between embryogenesis and ESC neural differentiation is the outstanding capacity of retinoic acid to induce neural fate from ESCs, compared to the little evidence for its intervention in neural induction *in vivo* (Gaulden & Reiter, 2008). Since the pioneering studies of Strickland (Strickland & Mahdavi, 1978) RA has been adopted to induce neuroectoderm formation from mouse and human ESCs in suspension aggregates termed embryoid bodies (EBs) (Bain & Gottlieb, 1994; Bibel et al., 2007; Gottlieb & Huettner, 1999), or later in adherent monolayer cultures (Andrews, 1984; Przyborski et al., 2000). The treatment of ESCs with RA recapitulates multiple aspects of neuroectoderm lineage commitment *in vivo*, as the repression of components of the pluripotency network such as Nanog, Oct4 and Klf4 (Ben-Shushan et al., 1995; Mendoza-Parra et al., 2011), the onset of the neuroectoderm transcriptional program, along with blockage of mesoderm and endoderm differentiations (Bain et al., 1996; Okada et al., 2004). A study in hESCs

proposed a difference in the mechanism of RA action in these cells compared to mouse pluripotent cells. In the former, RA does not trigger a cell autonomous program of differentiation and prolonged supply of RA is required to repress other fate options whereas, in mouse ESCs, RA is able to initiate a cell autonomous program in even few hours (Berg & McBurney, 1990; Tonge & Andrews, 2010). Indeed, after only 6 to 8 hours from RA addition, a first transcriptional response can be noticed, characterized by the activation of RA direct targets, followed by a second wave of transcriptional change between 24 and 48 hours, determined by the exit from pluripotency (Mahony et al., 2011; Semrau et al., 2017). Noteworthy, some of the original studies claiming efficient neural conversion in "default" conditions or absence of extrinsic cues, presented caveats like pre-culture of the cells in serum containing medium or induction of differentiation in B27 with retinyl acetate (providing substrates for RA synthesis), leaving room for a RA action upon endogenous production at nanomolar concentrations in these differentiations (Lenka and Ramasamy 2007; Smukler et al. 2006; Tropepe et al. 2001; Kim et al. 2009; Engberg et al. 2010).

1.5.3.4 Profiling transcriptional changes during ESCs lineage acquisition

In a recent work published by our group (Sladitschek & Neveu, 2019), the changes of gene expression in response to external cues provided with the differentiation protocols have been explored, exposing mESCs to separate conditions promoting ectoderm, mesoderm or endoderm differentiation (Borowiak et al., 2009; Torres et al., 2012; Ying et al., 2003), and sequencing the RNA of samples collected every 24 hours for a total of six days. The transcriptional snapshots obtained for each germ layer were connected to delineate trajectories of gene expression changes in each differentiation, represented in the Principal Component Analysis (PCA) in Fig1.17a. The divergence of the three profiles argues for a segregation of gene expression in lineage-specific programs as early as 24 hours from the start of the differentiation, without observing common pluripotent states. Transcriptomes derived from spatially defined regions of mouse gastrula at different stages of development were recently made available (Peng et al., 2019), and they were superimposed on the principal component space created by our *in vitro* germ layers, to test whether they were closely related. Remarkably, the transcriptomes derived from proximal posterior sections of E6.5-E7.5 gastrula, where the Primitive Streak lies, resulted close to the *in vitro*

endoderm trajectory, coherently with the origin of this germ layer from the PS (Lewis & Tam, 2006) (Fig1.17d-e). Moreover, between E7.0 and E7.5, transcriptomes coming from proximal mesoderm sections overlaid the *in vitro* mesoderm profile (Fig1.17e), whereas those originating from anterior epiblast regions projected on the *in vitro* ectoderm trajectory (Fig1.17f). Furthermore, it was noticed that mESCs cells treated with Activin A and FGF2, to induce the conversion into EpiSCs- primed pluripotent state (Guo et al., 2009), presented transcriptional profiles closely resembling intermediate states of the endoderm fate trajectory. In other words, the *in vitro* differentiation towards endoderm derivatives, triggered by the small molecule IDE1, seemed to transit through an epiblast-like primed state prior to the emergence of definitive endoderm (Sladitschek & Neveu, 2019).

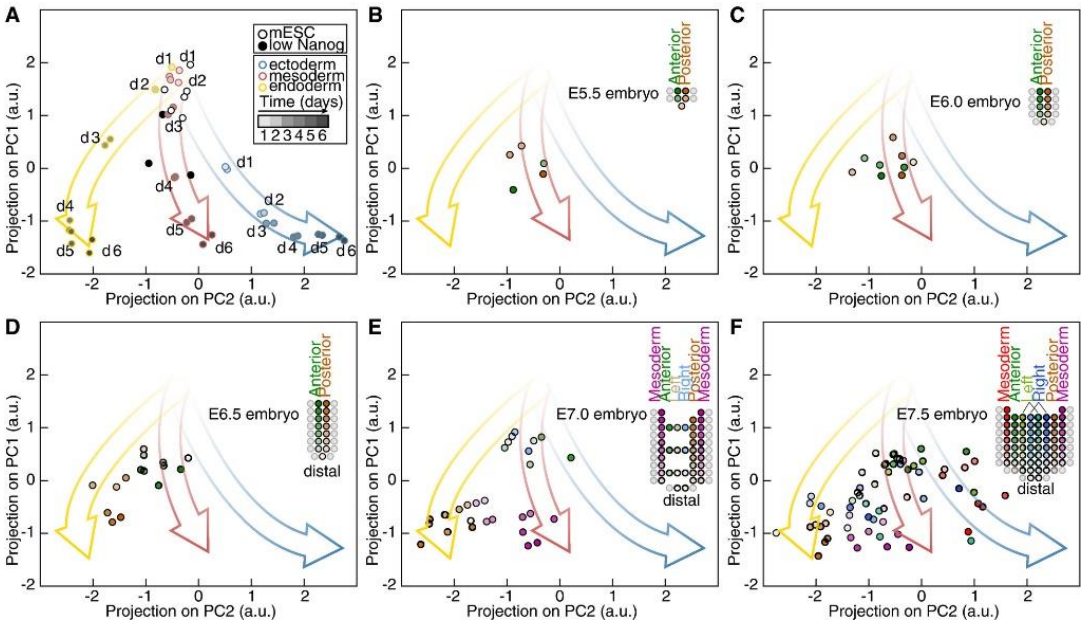


Figure 1.17: Comparison of in vitro generated germ layers with defined regions of gastrulating mouse embryo.

__A__) Principal component analysis of gene expression profiles during mESC differentiation to endoderm, mesoderm, and ectoderm. __B-F__) Projection of transcriptomes derived from spatially defined regions of mouse embryos at E5.5 (__B__), E6.0 (__C__), E6.5 (__D__), E7.0 (__E__), and E7.5 (__F__) stages on the Principal component coordinates generated by the lineage trajectories in __A__. Adapted from Sladitschek and Neveu 2019.

As previously mentioned, the generation of stable lines by the integration of fluorescent reporters into lineage specific *loci*, is a convenient way to follow the specification towards different fates. A widely adopted reporter line to track neuroectoderm formation is the *Sox1*_GFP knock-in line, bearing the insertion of the GFP coding sequence in the *Sox1* locus (Aubert et al., 2003). Starting from this line, the reporter H2B-3xTagBFP was targeted to the T(*Brachyury*) locus, to mark formation of primitive streak and mesoderm cell types. Finally, to follow in parallel the specification to endoderm, H2B-mCherry was integrated into the Eomesodermin (*Eomes*) locus (Sladitschek & Neveu, 2019), a T-box transcription factor expressed in the anterior PS and pivotal for definitive endoderm formation during mouse embryogenesis (Arnold et al., 2008; Russ et al., 2000; Teo et al., 2011).

2 AIM

Many progresses have been made in mimicking the formation of germ layers from mESCs, developing protocols inspired from embryogenesis, and enabling the formation of functional derivatives that can be re-implanted in the embryo and participate to development. This proves a solid knowledge of the basic signaling requirements for the formation of the germ layers. However, much has yet to be examined with respect to how the cells start and eventually lock a fate decision among many available at the same time; the influence of the timing of extrinsic cues over this decision and how multiple signals arrange coherently to determine fates; as well as when cells are only fated and not committed to a lineage program and whether the germ layers in culture show any of the inter-dependence demonstrated during development.

The combination of *in vitro* differentiation of reporter lines, as the above-mentioned triple knock-in line (3KI), together with flow cytometry and time-lapse imaging enable to readily detect, at the single cell level, the effect of applying different extrinsic cues on multiple fates, bearing a potential not yet exhausted. The aim of my project is to mimic the formation of the Primitive Streak and the subsequent induction of neuroectoderm fate in a much easily manipulatable context than mouse embryo, by establishing culture conditions recapitulating these processes. I will exploit such system to investigate the individual contributions and the interplay of the signaling pathways responsible for the formation of neuroectoderm during development and, to revisit the role of the retinoic acid among the neuroectoderm inducing cues.

3 RESULTS

3.1 Using a reporter line to track the specification towards Primitive Streak and neuroectoderm fates

The first part of my project focused on the characterization of conditions recapitulating multiple aspects of the formation of the Primitive Streak (PS), in order to deploy this system to investigate the mechanisms controlling the acquisition of the neural fate. To achieve this purpose, I started by carrying out the differentiation protocols leading towards the generation of germ layers, using the triple knock-in line (3KI) established in our group (Sladitschek & Neveu, 2019). Bearing a BFP reporter under the control of *Brachyury*, GFP integrated in the *Sox1* locus, and mCherry driven by the anterior PS marker *Eomes*, in fact, the line enabled the simultaneous investigation of Primitive Streak/endoderm and neuroectoderm fates.

When the cells are exposed to low serum concentrations and the small molecule IDE1, the mRNA levels of the Primitive Streak (PS) marker *T* (*Brachyury*) increase from 48 hours, reaching a plateau between four and five days of differentiation (Fig3.1a). Likewise, the fluorescence intensity of the BFP reporter, controlled by *T* promoter in the 3KI line, peaks within four and five days and then decreases (Fig3.1b). I isolated the cells expressing BFP by fluorescence-activated cell sorting (FACS) every 24 hours of the differentiation procedure, sequenced the total RNA and compared the expression of multiple PS genes with wild type cells differentiated in the same way. As showed in the heatmap in Fig3.1c, the BFP^+ cells revealed upregulation of almost all the PS markers tested, divided into groups according to the positions along the anteroposterior axis, as well as markers of definitive endoderm and mesoderm (particularly presomitic mesoderm). In other terms, BFP expression in this culture condition marks a population with a transcriptome resembling the PS, endoderm and mesoderm.

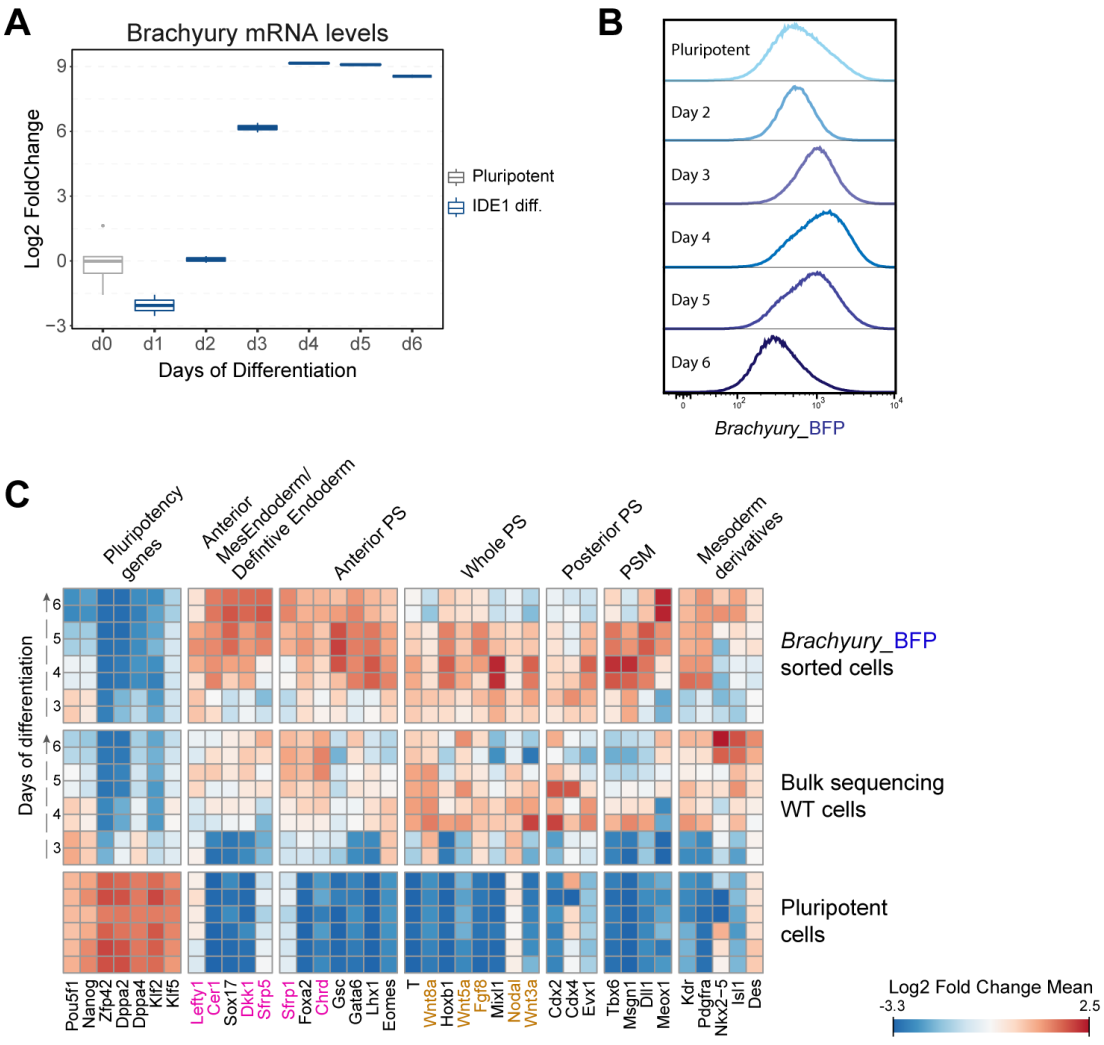


Figure 3.1: Acquisition of Primitive Streak -like fate by the Brachyury_BFP^{+} population during IDE1 differentiation.

A) Box plot representing Brachyury relative expression measured as Log_2 of the fold change of the normalized read counts of wild type mESCs, differentiated according to standard IDE1 protocol, compared to pluripotent mESCs. B) Histograms of the BFP fluorescence intensity changes of 3KI cells undergoing IDE1 differentiation, measured every 24 hours by flow cytometry. C) Heatmap showing the relative expression of Primitive Streak, definitive endoderm and mesoderm markers (following Kojima et al., 2014) of the 3KI line BFP^{+} sorted cells, or of wild type mESCs during IDE1 differentiation, and pluripotent mESCs. The color code refers to minimum and maximum value of the Log_2 of the fold change calculated on the mean of the normalized read count for each given gene.

On the other hand, during the neuroectoderm differentiation driven by high concentration of retinoic acid (RA), the expression of the neuroectoderm specific marker *Sox1* is induced within 24 hours from RA addition, and a total of two days from the cell seeding, and gradually decreases with time (Fig3.2a). Sox1 marks a stage of proliferative neural progenitors, and its downregulation indicates the transition towards later stages of neural differentiation. Accordingly, the fluorescence intensity of the GFP reporter expressed from the *Sox1* locus in the 3KI line, increases during the first days of differentiation and declines after four days (Fig3.2b). As for BFP, I purified the GFP^+ cells and analyzed the expression of neuroectoderm specific genes in comparison to wild type cells exposed to the same treatment. The cells selected for the expression of GFP exhibited a clear neural profile, with upregulation of genes coding for *Zic1*, *Ascl1* and several bHLH factors responsible for the differentiation of the progenitors (Fig3.2c).

Taken together these observations point out that the use of the 3KI line in combination with optimized differentiation protocols enables efficient monitoring of specification towards neuroectoderm or Primitive Streak fates, rendering the system suitable for the investigation of mechanisms underlying their formation.

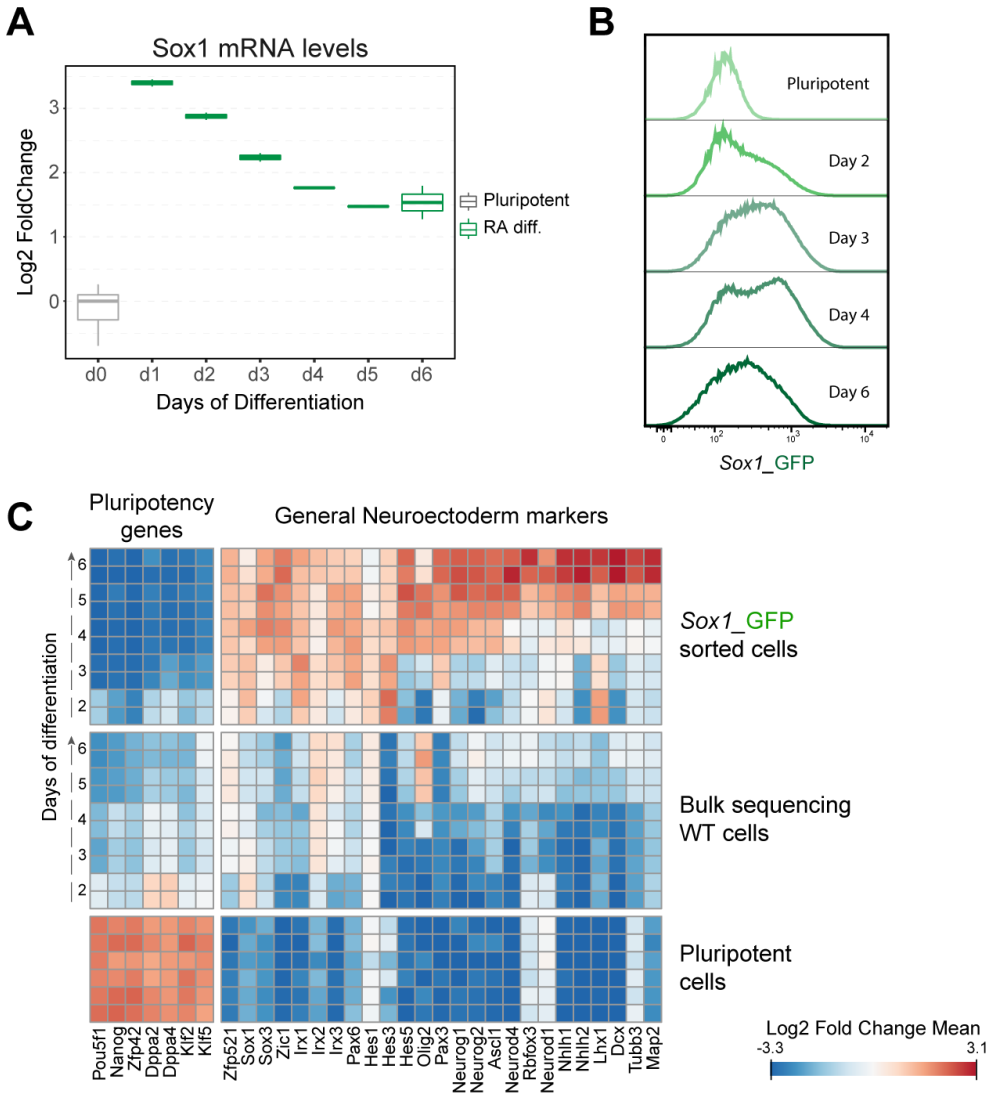


Figure 3.2: Neuroectoderm fate acquisition by the Sox1_GFP$^+$ population during RA differentiation.

__A__) Box plot depicting Sox1 relative expression measured as Log_2 of the fold change of the normalized read counts of wild type mESCs, differentiated according to standard RA protocol, compared to pluripotent mESCs. __B__) Histograms representing GFP fluorescence intensity changes of the 3KI cells undergoing RA differentiation, measured every 24 hours by flow cytometry. __C__) Heatmap showing the relative expression of general neuroectoderm markers of the 3KI line GFP$^+$ sorted cells or wild type mESCs undergoing standard RA differentiation, and pluripotent mESCs. The color code refers to minimum and maximum value of the Log_2 of the fold change calculated on the mean of the normalized read count for each given gene.

3.2 The canonical Wnt pathway elicits the formation of the Primitive Streak-like cells *in vitro*

To further test the resemblance of the cells expressing BFP during the IDE1 differentiation to the PS, I sought to verify whether they rely on the same signaling requirements as their *in vivo* counterpart. Despite being both indispensable for the correct formation of PS in mouse embryo, the activation of Activin/Nodal/TGF-β is usually preferred to the one of Wnt signaling by protocols that aim at the formation of PS and endoderm in culture. I investigated the requirement of canonical Wnt activation for *T* expression, by either exposing the 3KI line cells to CHIR99021 (CHIR), to promote the β-catenin transduction cascade, or to the small molecule inhibitor XAV939 (XAV). BFP expression normally starts within two days of IDE1 treatment; however, concomitant inhibition of Wnt signaling by XAV almost completely abrogated the reporter expression. On the contrary, the use of CHIR alone led to BFP expression in most of the cells and with fluorescence intensity detected much higher than in IDE1, Activin A or Wnt3a conditions (Fig3.3a). Considering that *T* has been reported as a canonical Wnt responsive gene (Arnold et al., 2000; Vonica & Gumbiner, 2002; Yamaguchi et al., 1999), there was the possibility that its expression could be directly triggered by Wnt without additionally promoting the PS formation. Wnt activation via CHIR is also used in combination with the Mek inhibition in the well-known 2i condition, to maintain naïve pluripotency. To test the early transcriptional changes driven by Wnt, I purified the BFP^+ cells arising 48 hours after the addition of CHIR in PS promoting condition and, in parallel, I added CHIR to cells in ES complete medium (pluripotency maintaining regime) for the same amount of time, before RNA sequencing. The comparison between the CHIR induced BFP^+ population and cells cultured 48 hours in IDE1, showed upregulation of general and posterior PS genes, along with presomitic mesoderm markers (Fig3.3b). Therefore, the expression of multiple PS markers, together with reduction of pluripotency ones argues in favor of a wider action of Wnt pathway in the PS formation than activation of *T* alone.

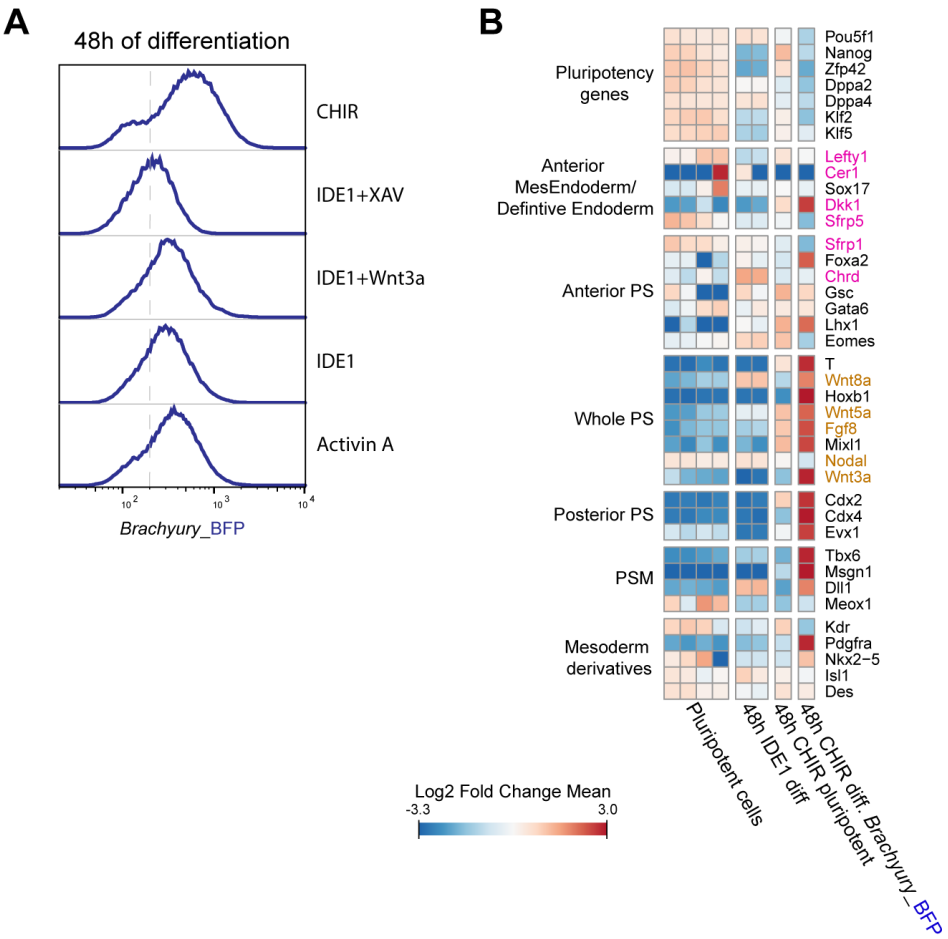


Figure 3.3: Canonical Wnt control of the Primitive Streak-like fate.

__A__) Histograms depicting BFP fluorescence intensity analyzed by flow cytometry after 48 hours of treatment of the 3KI line cells with the indicated molecules. __B__) Heatmap analyzing early transcriptional effects of Wnt activation, evaluating the relative expression of Primitive Streak, definitive endoderm and mesoderm markers after 48 hours of differentiation. The conditions compared are 3KI line BFP^{+} cells sorted after 48 hours of CHIR induction; 3KI line cells in pluripotency maintaining medium with addition of CHIR for 48 hours, wild type mESCs after 48 hours of IDE1 differentiation, and pluripotent mESCs.

3.3 Detection of canonical Wnt signaling during Primitive Streak-like differentiation with a novel reporter line

Albeit pointing to a clear involvement of Wnt pathway in the decision of undergoing PS fate, evidences of its activation and variations of its activity during this process were still elusive. To verify the presence of Wnt signaling during standard low serum-IDE1 protocol, I decided to create a new triple reporter line starting from the *Sox1*_GFP, *T*_BFP double knock-in line (2KI), adding a reporter of canonical Wnt activation (2KI-Wnt). Inspired by the logic of the TOP Flash luciferase reporter assays, I generated a construct in which TCF/LEF responsive elements drive the expression of a destabilized fluorescent protein NLS_SCARLET_PEST2D reporter. In the newly derived 2KI-Wnt line, the detection of Scarlet demarks cells in which the β-catenin transduction cascade is promoting the expression of Wnt target genes. The differentiation in Fig3.4a, performed with the 2KI-Wnt line in PS-like regime, displayed the appearance of cells with active Wnt signaling between day 2 and 3, with the Scarlet^{+} fraction and fluorescence peaking around day 4 and then decreasing. Remarkably, Scarlet expression was almost always associated with the one of BFP indicating that Wnt activation and *T* expression follow similar dynamics over time. Confocal imaging of the 2KI-Wnt cells differentiated from two to four days with IDE1 confirmed the presence of cells expressing Scarlet among the BFP^{+} cells (Fig3.4b). To determine the fate of the cells expressing both BFP and Scarlet, I purified the double positive population by FACS after five days of differentiation, along with cells expressing only BFP or neither of the reporters, and I compared their transcription profiles with the bulk sequencing of wild type cells differentiated for the same amount of time. The comparison highlighted that the Wnt^{+}-Brachyury^{+} population expressed multiple PS markers, but with a propensity towards the posterior and presomitic mesoderm genes, whereas the anterior markers were expressed to a lower extent, in agreement with the fates induced by Wnt signaling *in vivo* (Fig3.4c). The above findings confirm that during our PS-like differentiation canonical Wnt signaling is active and underlies the formation of cells belonging to posterior PS and mesoderm.

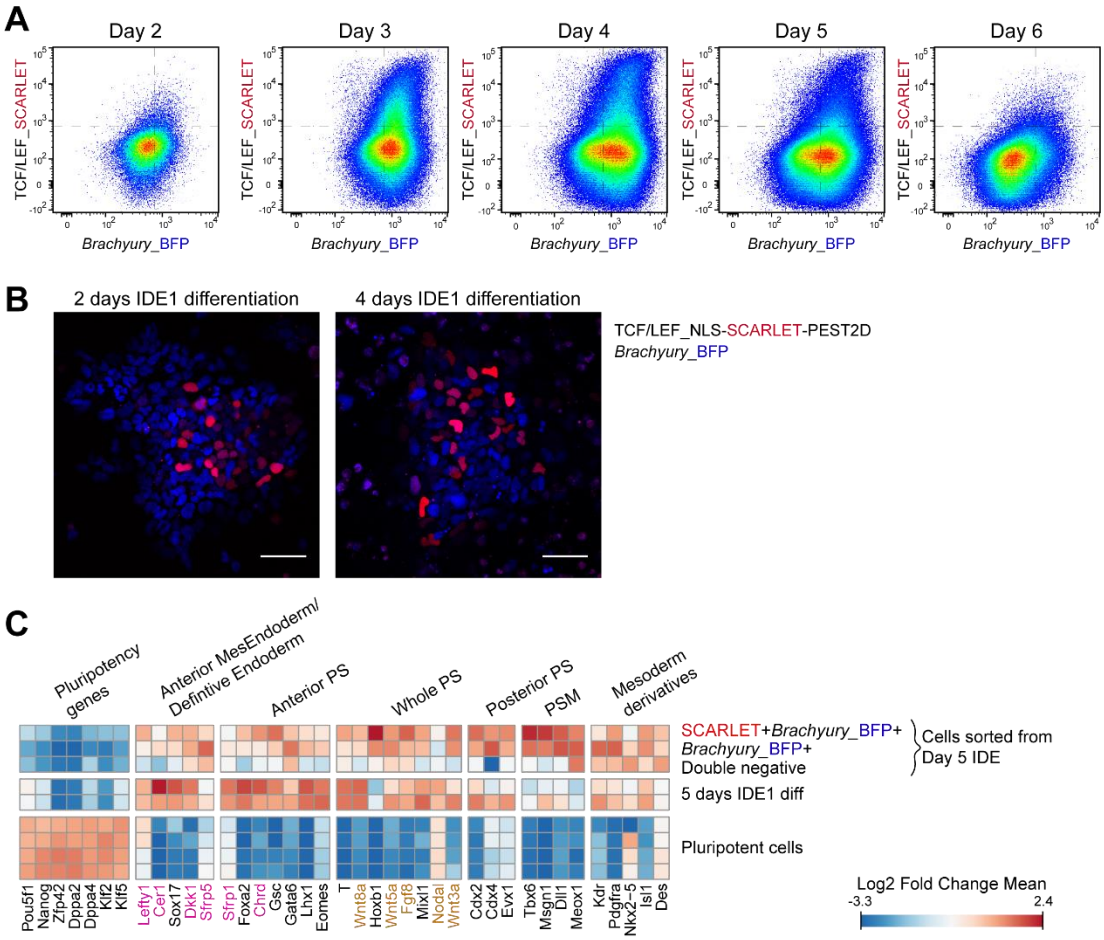


Figure 3.4: Canonical Wnt activation during Primitive Streak-like differentiation detected by the 2KI-Wnt line.

A) Wnt reporter-triple knock-in line (2KI-Wnt) differentiation to Primitive Streak with standard IDE1 protocol, measured every 24 hours by flow cytometry. Pseudocolor density plot displaying the relative density of cell populations along BFP and Scarlet fluorescence intensity axes. Every dot in the plot represent a single cell. The color code reflects the relative cell density within the graph window: blue (low) to red (high). The dotted line separates into four quadrants the space according to expression of BFP and Scarlet. The axes scale used to measure the fluorescence intensity is biexponential. ***B****) Confocal imaging of 2KI-Wnt cells after two and four days of IDE1 differentiation. Scarlet fluorescence marks a sub-population of the BFP^{+} Primitive Streak-like cells with active Wnt signaling. Scale Bar=50µm.* ***C****) Heatmap analyzing the fate of cells with active Wnt signaling after five days of standard IDE1 differentiation. The conditions compared are Scarlet^{+} BFP^{+}, Scarlet $^{-}$ BFP^{+}, and double negative populations sorted from five days of differentiation, wild type mESCs differentiated for the same time, and wild type mESCs in pluripotency maintaining medium.*

3.4 TGF-β is dispensable to start the formation of Primitive Streak-like cells but is necessary to induce the definitive endoderm

Given the extensively reported crosstalk between Activin/Nodal/TGF-β and Wnt pathways, the above-mentioned functions of the Wnt pathway could depend, at least in part, on an indirect activation of TGF-β. To evaluate the role of Wnt in absence of TGF-β, I blocked this pathway with the inhibitor SB431542 (SB43) while activating Wnt, and I used the *Eomes*_mCherry 3KI line to assess the outcome on PS and definitive endoderm fates. Comparing cells differentiated for four days in IDE1 medium, and cells exposed to both SB43 and CHIR, highlighted a reduced but clearly present BFP fluorescence in the latter case, implying that TGF-β is not strictly required for *T* expression and the induction of the PS (Fig3.5a). At the same time, the mCherry reporter signal controlled by the late PS/ anterior definitive endoderm marker *Eomes*, is completely extinguished when TGF-β is inhibited but Wnt is active (Fig3.5a). The above findings, together with a raised proportion of cells expressing the GFP reporter in this condition, suggest a requirement for TGF-β activation to form anterior primitive streak cell types and block neuroectoderm formation. Isolating BFP^{+} and BFP^{-} cells from either standard IDE1 or SB43 and CHIR differentiation, I gained more insights into the identity of the cells deriving from Wnt activation and TGF-β inhibition. As anticipated by the flow cytometry data, this population displayed multiple general PS markers, albeit to lower levels compared to the IDE1-induced BFP cells (Fig3.5b). Moreover, it presented a clear modified propensity from anterior, definitive endoderm fates typical of IDE1 induced cells, to posterior PS and presomitic mesoderm derivatives, similarly to what observed for the cells presenting active Wnt in the 2KI-Wnt reporter line. Overall, the data presented entail a quite close resemblance between the PS-like population and its *in vivo* counterpart, from the transcriptome level to the underlying signaling network. Such close resemblance paved the ground for investigating functions of the PS in our system, such as its neuroectoderm inducing properties.

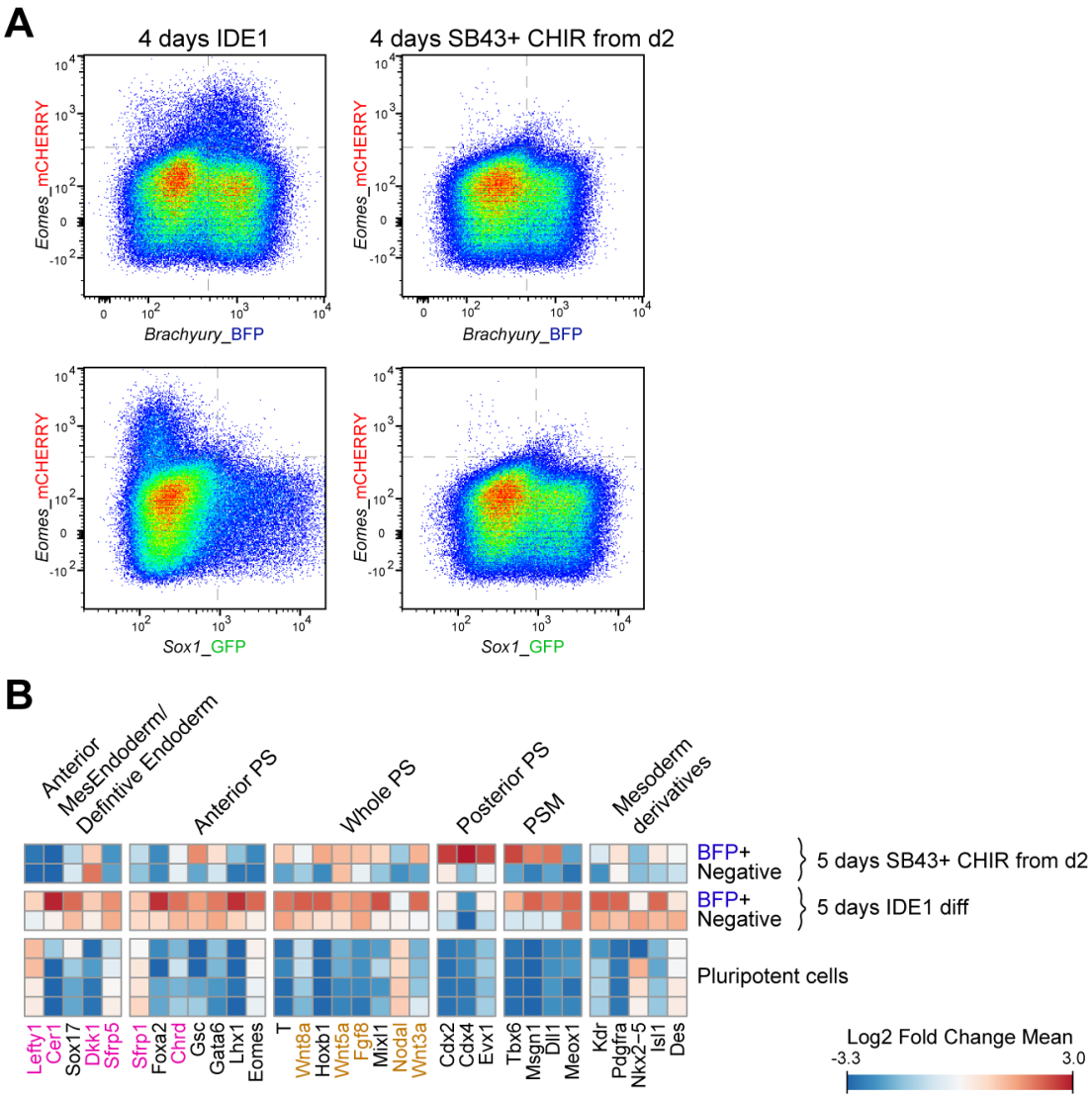


Figure 3.5: Changing the fate propensity of the Primitive Streak-like population upon Nodal/TGF-β inhibition.

A) Pseudocolor density plots displaying four days of differentiation of the Eomes_mCherry triple knock-in line (3KI) towards Primitive Streak fate, via standard IDE1 protocol or four days of SB43 with addition of CHIR the last 48 hours before the analysis. ***B)*** *Heatmap depicting the change of fate of cells undergoing Primitive Streak-like differentiation in conditions inhibiting Nodal/TGF-β signaling and promoting canonical Wnt. The samples analyzed are BFP^+ and BFP^- cells sorted from five days of differentiation with SB43 and CHIR added 48 hours before measuring, BFP^+ and BFP^- cells sorted from five days of standard IDE1 differentiation, and wild type mESCs in pluripotency maintaining medium.*

3.5 Induction of neuroectoderm derivatives from the PS-like population

Under low serum-IDE1 culture regime, the majority of cells adopts a PS-like destiny and further differentiates in definitive endoderm or mesoderm cell types. Considering that during mouse development the PS and its derivatives are paramount organizers of the initial stages of neuroectoderm formation, I wondered whether it could be possible to investigate the onset of neuroectoderm from factors secreted by the PS, exploiting the readily manipulatable system I had characterized. Notably, starting between three and four days of the PS-like differentiation a group of cells showing robust GFP expression gradually and reproducibly emerges (a part of which also exhibits expression of the BFP reporter) (Fig3.6a). To be able to capture the GFP progression in parallel with the PS evolution over time, I introduced an H2B-iRFP reporter in the background of the *Sox1*_GFP, *T*_BFP 2KI line to mark constitutively all the cell nuclei. Using this line, I performed a time-lapse imaging experiment starting from three days of the low-serum IDE1 regime, to capture the onset of the GFP population. A group of GFP^+ cells gradually arose from a group of cells undergoing PS-like differentiation, partially sharing also BFP fluorescence, as depicted in the time frames in Fig3.6b. To test whether the expression of GFP was stable and demarking cells fated to neuroectoderm, I isolated them by FACS and sequenced their mRNAs. In addition, I re-seeded the GFP^+ population in N2B27 medium, maintaining it in culture with regular passaging to test whether the cells preserve GFP expression. Notably, seven days from the re-seeding after sorting, GFP fluorescence was present in most of the cells and at considerably high levels (Fig3.7a). The finding implies that the GFP^+ cells originating in the PS-like condition are in fact stably expressing the GFP reporter even when cultured in isolation. Therefore, I re-sorted the cells maintained in culture, separating high GFP and low GFP fractions, as well as the cells not expressing GFP, and I tested the expression of neuroectoderm markers in these populations and in the GFP^+ cells originally isolated from the IDE1 differentiation. They all displayed upregulation of many general neural lineage genes in comparison to BFP^+ PS-like cells or pluripotent cells (Fig3.7b). Of note, the cells with intermediate levels of GFP fluorescence after extensive culture period showed a more mature neural progenitor transcriptome with respect to the population with

higher GFP fluorescence, confirming that the reduction of GFP and thus Sox1 is indicative of later stages of the differentiation.

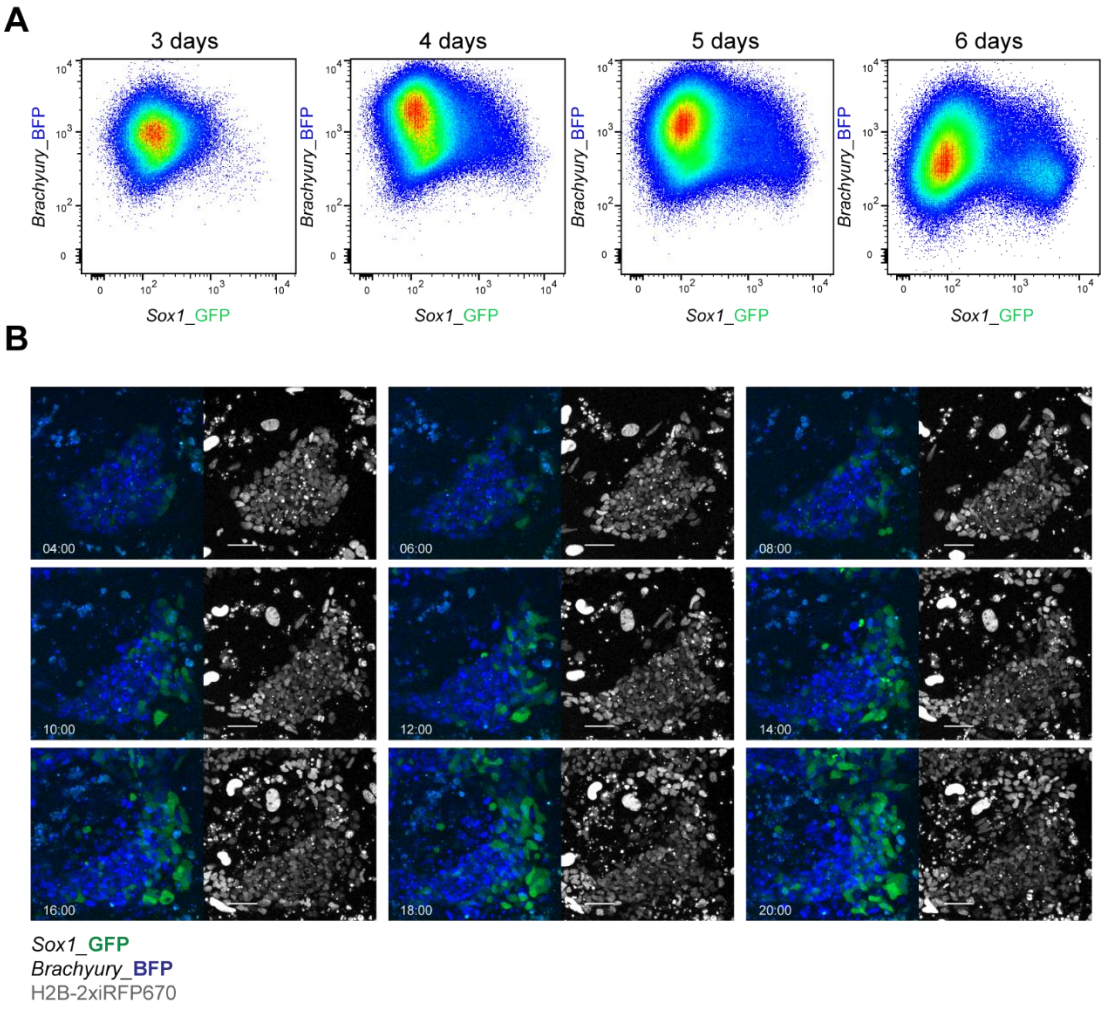


Figure 3.6: Onset of a GFP$^+$ population in Primitive Streak-like differentiation.

***A**) Pseudocolor density plots displaying the emergence of a GFP$^+$ population during six days of standard Primitive Streak-like differentiation. **B**) Time-lapse imaging experiment of cells during the standard IDE1 differentiation. The imaging started 72 hours after the beginning of the protocol and the time indicated refers to hours from the beginning of the imaging experiment. The constitutively expressed H2B-iRFP reporter was used to visualize all the cell nuclei and automatically set the focus plane. Scale bar=50µm.*

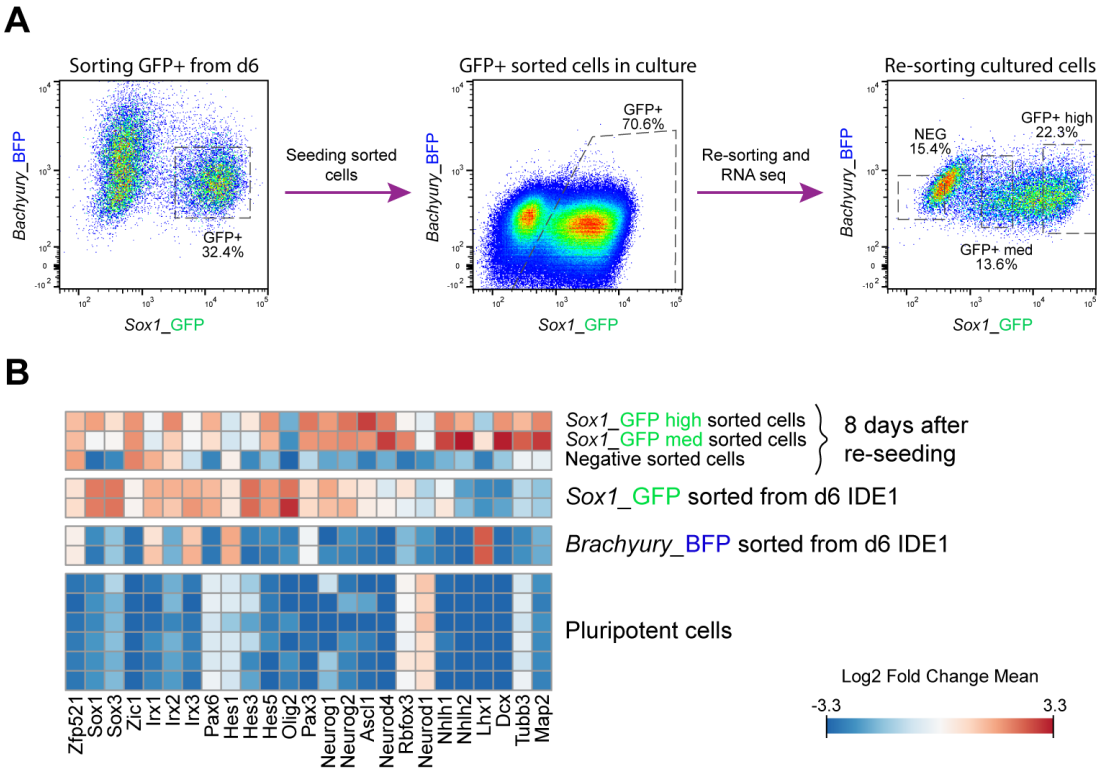


Figure 3.7: Stable expression of GFP reporter in IDE1 regime marks neuroectoderm cell types.

A) GFP$^+$ cells sorted with DB Melody from six days of standard IDE1 differentiation were reseeded in N2B27 medium in laminin coated dish (left plot); after seven days and two passaging rounds, GFP expression was analyzed by DB Fortessa flow cytometer (central plot), and after a total of eight days high GFP$^+$, low GFP$^+$ and GFP$^-$ populations were sorted by DB Melody before RNA sequencing (right plot). ***B****) Heatmap showing the relative expression of neuroectoderm specific genes in the cells isolated as represented in the left and right plots in* ***A****), in addition to GFP$^+$ and BFP$^+$ populations sorted from standard IDE1 differentiation and pluripotent mESCs.*

The first neuroectoderm type, originating in the anterior epiblast, is responsible for the formation of the rostral part of neural axis, induced by the action of the PS and its derivatives, and is later followed by more posterior neuroectoderm fates. I sought to understand which kind of neuroectoderm was represented in the GFP^+ cells arising in the PS-like condition. For this reason, I tested the expression of anterior, hindbrain or posterior/spinal cord neural markers from the cells isolated for GFP fluorescence in comparison to the BFP sorted ones. Strikingly, the GFP population shows upregulation of all the three categories of markers, opening the question about the mechanisms of induction of the various neuroectoderm types in this system (Fig3.8). Therefore, the conditions sustaining the formation of a structure resembling the PS also lead to the generation of various neural fates with the mechanisms underlying these processes suitable for exploration.

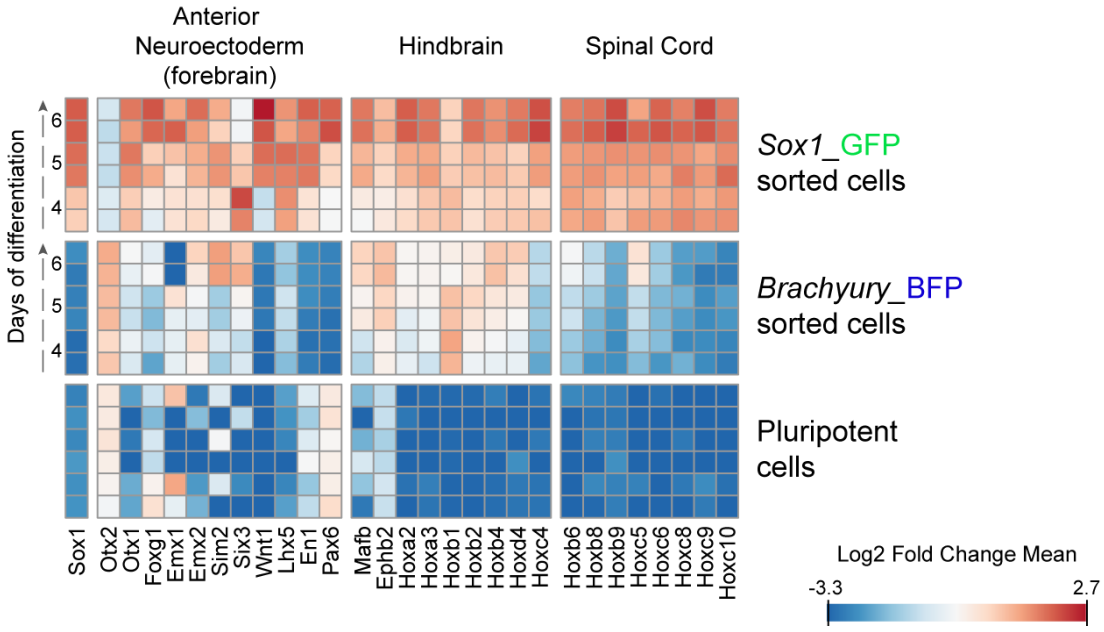


Figure 3.8: Neural fates subtypes included in the GFP+ population emerging during Primitive Streak-like differentiation.

Heatmap showing the relative expression of forebrain, hindbrain and spinal cord markers (following Gouti et al., 2014) during standard IDE1 differentiation. The conditions compared are the GFP^+ BFP^- and the GFP^- BFP^+ cells, sorted after four, five and six days of differentiation, and wild type mESCs in pluripotency maintaining medium. The color code refers to minimum and maximum value of the Log_2 of the Fold Change calculated on the mean of the normalized read count for each given gene.

3.6 Signaling factors secreted by the Primitive Streak-like cells might control the induction of neural fates

The emergence of multiple neuroectoderm types in the presence of a population mimicking the PS under many aspects, entailed that the induction of neural fate might be orchestrated by factors produced by this population, as it occurs *in vivo*. Besides expressing ligands of Wnt, Nodal/ TGF-β and BMP/ TGF-β pathways, in fact, the cells marked by the BFP fluorescence revealed upregulation of genes coding for secreted inhibitors such as Dkk1, Cerberus, Chordin and Noggin, responsible for the induction of the neuroectoderm in the anterior epiblast *in vivo* (Fig3.9a). Although the effect of these pathways is well characterized during development, their individual contribution to the acquisition of neural fate is yet unclear owing to the difficulties to study this process in the embryo. Therefore, I sought to understand the effect of activating or inhibiting these pathways individually on the specification towards neuroectoderm. The addition of Activin A or BMP4 respectively from the beginning of the differentiation or after two days, almost completely abrogated the detection of the GFP reporter and the formation of neuroectoderm (Fig3.9b). Substantially different is the effect of adding CHIR to the PS-like regime, not resulting in repression of GFP expression and highlighting a fundamental distinction between Wnt and TGF-β actions over the neural fate (Fig3.9b). The difference is confirmed when these signaling cues are applied on cells already undergoing neuroectoderm specification induced by retinoic acid. In these conditions, in fact, Activin A impaired the GFP^+ population, whereas Wnt activation did not affect negatively the GFP fluorescence, which was rather higher than standard level (Fig3.9c). In the same experiment the combination of Activin A, BMP4 and Wnt3a completely blocked the differentiation towards neuroectoderm (Fig3.9c). From these experiments I concluded that the factors produced by the PS-like population affect differently the generation of neural derivatives, with the two TGF-β pathways being unequivocally detrimental for this fate, whereas the Wnt pathway shows a much more context dependent effect.

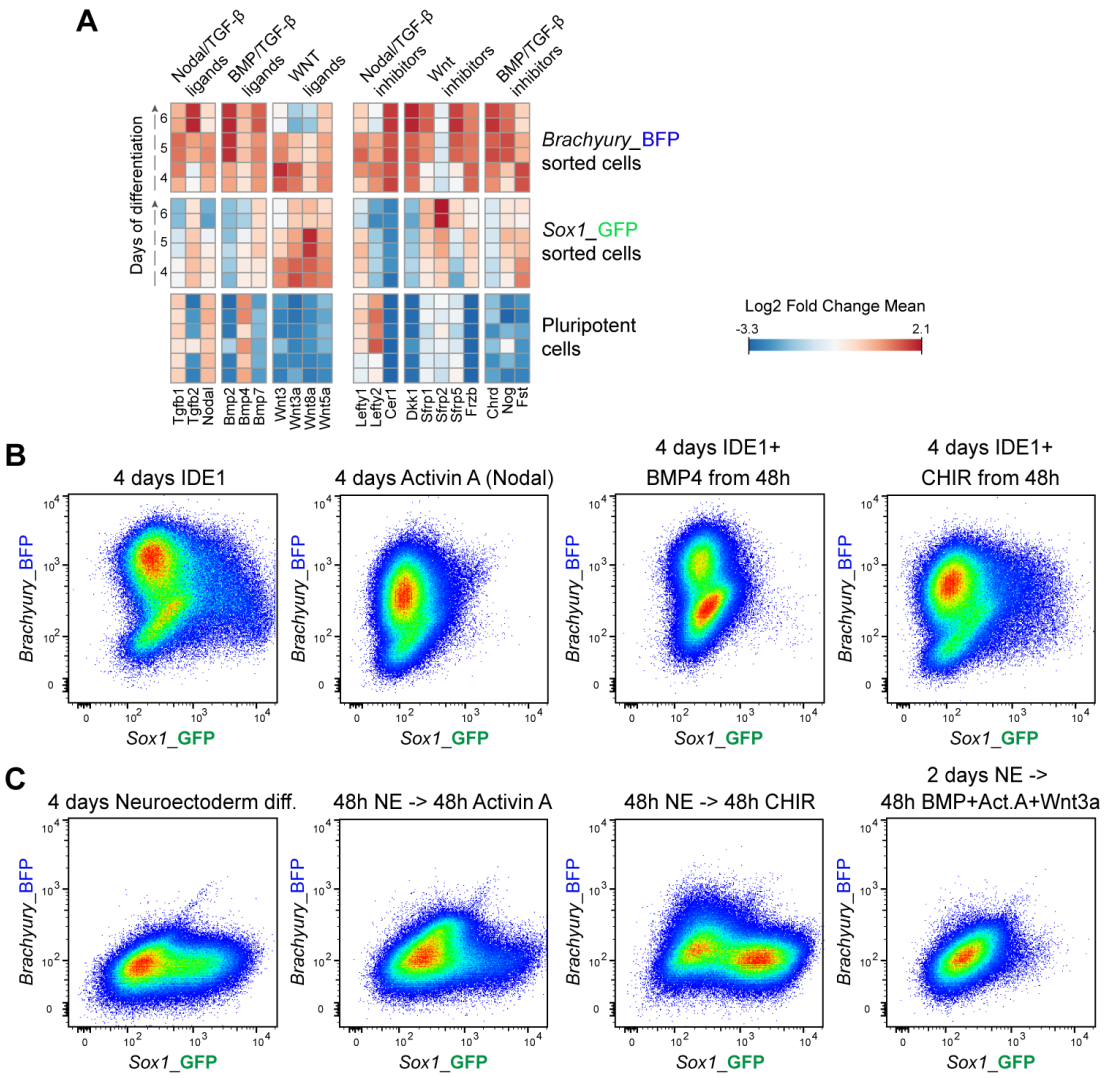


Figure 3.9: Effects of the signaling molecules secreted from the Primitive Streak-like population on neural fate acquisition.

A) Heatmap depicting the relative expression of genes coding for ligands and secreted inhibitors of Wnt and TGF-β pathways during IDE1 differentiation. Compared are the GFP$^-$ BFP$^+$ and the GFP$^+$ BFP$^-$ cells sorted after four, five and six days of differentiation, and wild type mESCs in pluripotency maintaining media. ***B****) Pseudocolor density plots highlighting different expressions of the GFP reporter after four days of differentiation in standard IDE1 regime, four days of Activin A without IDE1, four days of IDE1 with BMP4 in the last 48 hours, or four days of IDE1 with CHIR in the last 48 hours before measurement.* ***C****) Pseudocolor density plots showing the effect of Wnt or TGF-β activation over cells pre-directed towards the neuroectoderm lineage by retinoic acid. Compared are cells differentiated for 24 hours with N2B27 alone and, according to the conditions tested, 3 days of 1μM RA, 24 hours of RA and 48 hours of Activin A, 24 hours of RA and 48 hours of CHIR, 24 hours of RA and 48 hours of BMP4, Activin A and Wnt3a together.*

3.7 Secreted inhibitors of Wnt and TGF-β pathways independently induce the formation of neuroectoderm

During development, the inhibition of Wnt or TGF-β signaling are considered crucial for the origin of neuroectoderm, and mice bearing null mutations in either the Wnt inhibitor, *Dkk1*, or the two BMP inhibitors, *Chordin* and *Noggin*, present severe truncation of forebrain structures. After investigating the individual effect of Wnt and TGF-β activation with respect to neural fate decision, I focused on the neuroectoderm inducing properties of their single inhibition. Exposing mESCs to secreted inhibitors of Wnt (Dkk1 or SFRP1), or the small molecule XAV, equally led to the expression of GFP in most of the cells, with variable fluorescence intensity (Fig3.10a). Likewise, inhibition of BMP/ TGF-β (Noggin) or Nodal/ TGF-β (SB43) favored the specification towards neuroectoderm over the acquisition of the PS fate (Fig3.10b). Thus, single inhibition of Wnt or TGF-β can induce neural cell types, independently, even from pluripotent cells. To test with a different approach the importance of this inhibitory activity for the formation of neuroectoderm derivatives in the low serum-IDE1 regime, I introduced null mutations in the two BMP inhibitors, *Chordin* and *Noggin*, by CRISPR-Cas9 in the *Sox1*_GFP, *T*_BFP knock-in line. In agreement with the effect of BMP inhibition and the role of the two inhibitors *in vivo*, simultaneous mutation of *Chordin* and *Noggin* led to a remarkable reduction of the GFP^+ fraction during the PS-like differentiation (Fig3.10c). Taken together the above observations reveal the existence, in the *in vitro* scenario, of a delicate balance between the agonists of Wnt or TGF-β, promoting the acquisition of a PS-like fate, and their inhibitors, which instead promote specification towards neural lineage.

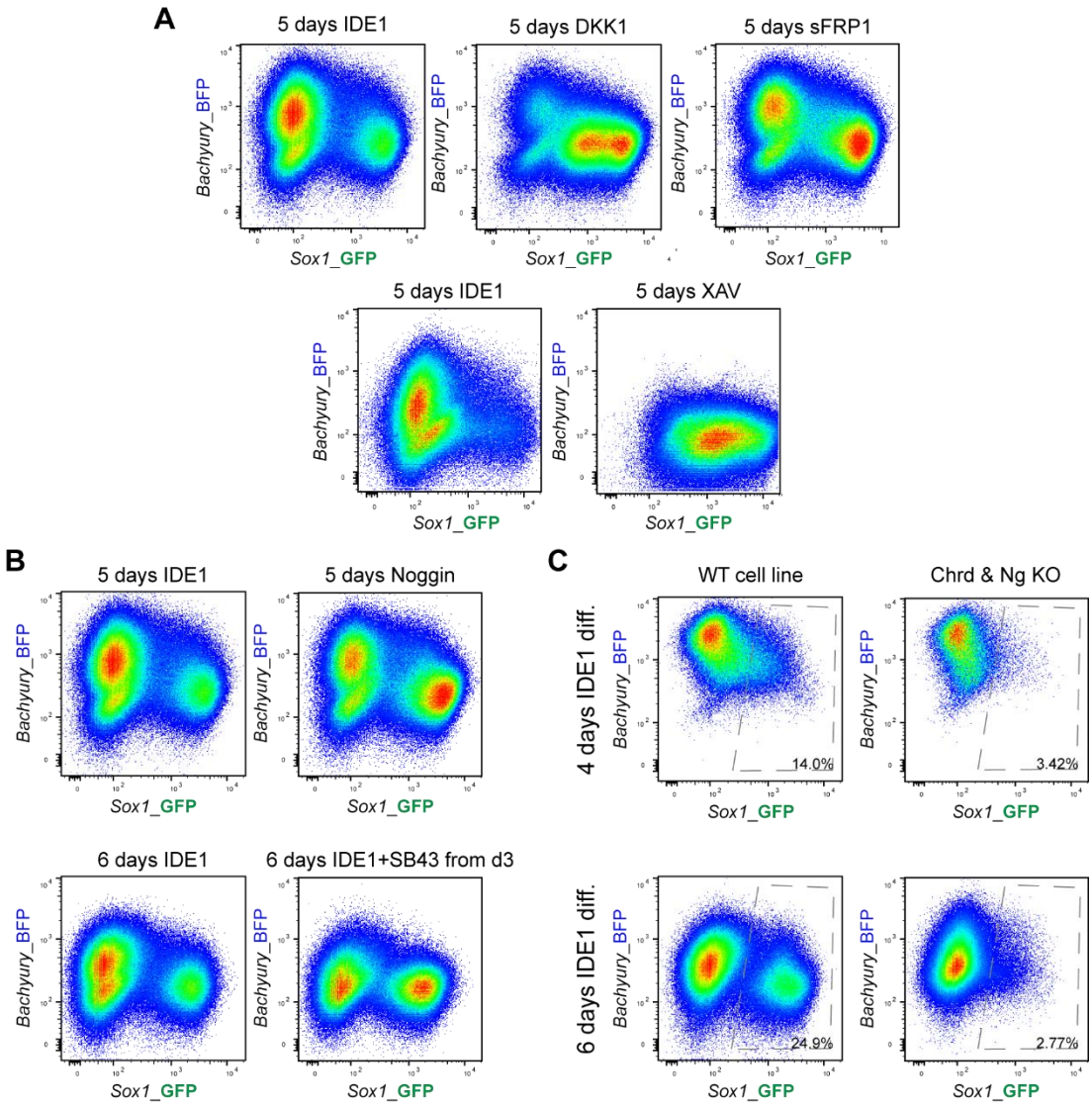


Figure 3.10: Induction of neural fate via Wnt or TGF-β inhibition.

__A__) Pseudocolor density plots showing the effects of Wnt inhibition over induction of the neural fate. The conditions analyzed consist of five days of differentiation with standard IDE1 protocol; five days of differentiation with the secreted Wnt inhibitors Dkk1 or sFRP1, or with the small inhibiting molecule XAV. __B__) Pseudocolor density plots showing the effects of TGF-β inhibition over induction of the neural fate. Compared are five or six days of standard IDE1 differentiation and respectively five days of differentiation with the BMP inhibitor, Noggin, or six days of IDE1 with SB43 added for the last 72 hours before the analysis. __C__) Effects on the neuroectoderm induction by the Primitive Streak-like population of the combined mutations of the BMP inhibitors Chordin and Noggin after four and six days of standard IDE1 differentiation.

3.8 Signaling cues underlying the formation of different neuroectoderm subtypes

Next, I sought to reconstruct the signals, in the intricated network generated by the PS, responsible for the formation of the neuroectoderm types emerging in IDE1 culture. Therefore, I compared the transcription profile of series of conditions promoting the formation of GFP cells and analyzed the expression of forebrain, hindbrain, or spinal cord markers. Notably, distinct inductive stimuli resulted in the upregulation of specific neural markers, as showed in the heatmap in Fig3.11. The inhibition of Wnt signaling appeared to upregulate anterior neuroectoderm transcripts, whereas Wnt activation in combination with Nodal/ TGF-β inhibition determined expression of genes demarking the most posterior part of the neural axis, the spinal cord (Fig3.11). Notably, the profile of the intermediate neuroectoderm category along the anteroposterior axis, the prospective hindbrain, derived from cells differentiated with high concentration of retinoic acid. This result opened to the possibility that retinoic acid, together with Wnt and TGF-β pathways, might be one of the mechanism adopted by the PS-like population to control the specification towards the neuroectoderm fate, and not only represent a stratagem to induce derivatives of this lineage *in vitro*.

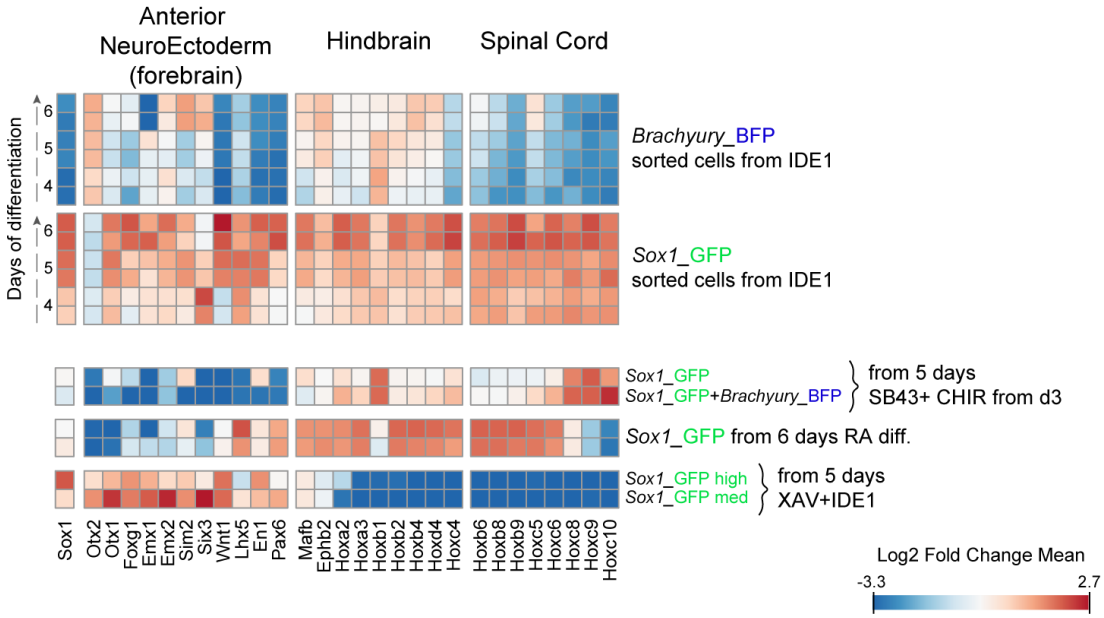


Figure 3.11: Reconstruction of the stimuli contributing to the formation of the neural cell types emerging in IDE1 differentiation.

Heatmap showing a connection between different neuroectoderm types and the signals potentially contributing to their formation. The conditions compared are GFP^+ BFP^- and GFP^- BFP^+ cells sorted after four, five and six days of differentiation in the first two groups of samples and: GFP^+ and GFP^+BFP^+ cells from five days of differentiation with SB43 and CHIR added the last 48 hours; GFP^+ from six days of standard RA differentiation; cells showing high or intermediate levels of GFP after five days of IDE1 and XAV treatment.

3.9 Exogenous perturbation of the retinoic acid signaling affects the formation of neuroectoderm from the PS-like cells

To verify the influence of retinoic acid (RA) on the neuroectoderm formation from the PS-like population, I added high concentrations of RA (1uM) to the low serum-IDE1 regime starting from three days of differentiation. When compared to the course of the standard differentiation protocol, RA gradually but completely switched the fate of the differentiating cells during 72 hours of treatment, from BFP^+ /PS-like to GFP^+ population (Fig3.12). On the contrary, when the RA signaling was completely blocked by the potent inverse agonist/inhibitor, AGN193109 (AGN), the formation of neuroectoderm from the PS population seemed to be reduced although not completely extinguished, and the fraction of cells in which GFP fluorescence was detected by the end of the differentiation appeared diminished (Fig3.12). These observations prove that RA can affect the fate decision in the PS context, when added exogenously.

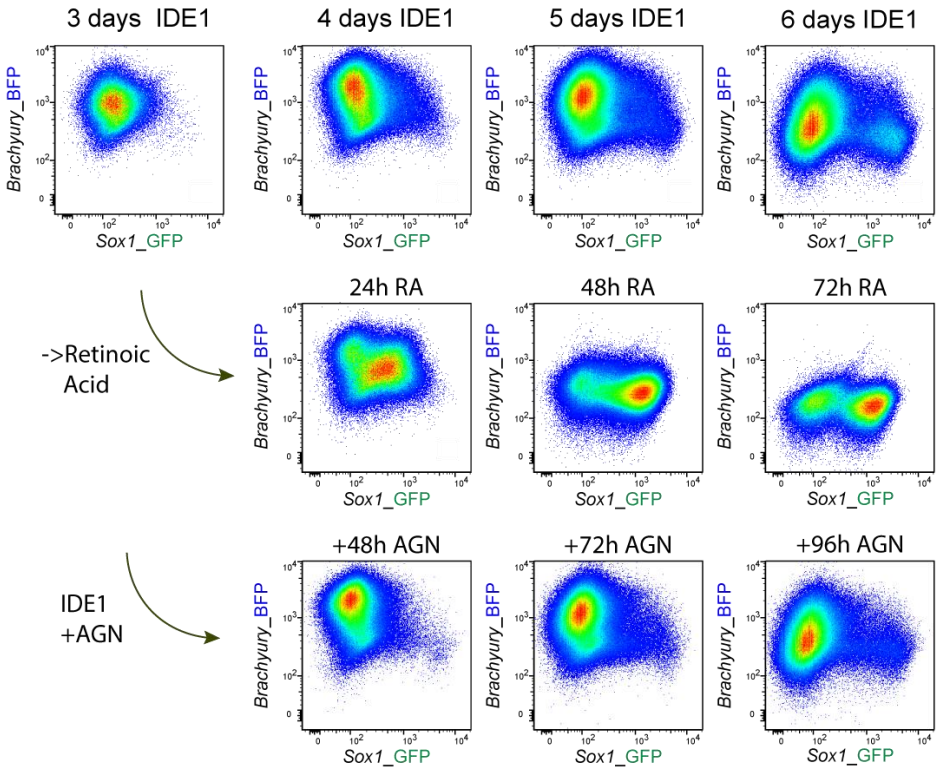


Figure 3.12: Activation or inhibition of RA pathway during IDE1 differentiation.

Pseudocolor density plots displaying the differences of BFP and GFP expression after six days of differentiation in standard IDE1, or IDE1 and 1μM RA added from three days of differentiation, or IDE1 and AGN added after from two days of differentiation.

3.10 Retinoic acid controls multiple aspects of neural fate acquisition from mESCs

The use of retinoic acid (RA) at high concentrations is a well-established and effective way to drive neural fate acquisition from mESCs. However, the presence of RA is considered dispensable to differentiate pluripotent ESCs to neuroectoderm and the general need for the molecule in this context is debated. Before exploring the effect of RA in the complex PS scenario, I sought to understand better its control over the neural fate in the N2B27 chemically defined condition, used in the neural differentiation form pluripotent cells. As anticipated in the section 3.1, in this regime the expression of the neural progenitor marker *Sox1* peaks early during the differentiation and gradually decreases (Fig3.2a). Differentiating pluripotent cells for the same amount of time with either N2B27 devoid of RA components (N2B27), a formulation of the same medium with substrates for RA synthesis (N2B27 vit.A), or with high concentration of RA, highlighted the RA capability of sustaining the entire

course of the differentiation. In the RA condition, by the end of the differentiation most of the cells displayed GFP fluorescence at reduced levels as expected for mature neural cell types (Fig3.13a). On the other hand, in the absence of RA signaling components, the same amount of time led to the acquisition of neural fate by only a minor fraction of cells, in the form of high GFP fluorescence, whereas in N2B27 vit.A an higher proportion of cells expressed high level of GFP reporter (Fig3.13a). Next, I wondered whether time and concentration of RA were both responsible for the control of RA on the decision to undergo neuroectoderm specification demarked by the increase of GFP fluorescence, and the differentiation of the neural progenitors revealed by the reporter downregulation. For this reason, I tested the effect of exposing the cells for only 24 hours to RA and continuing the differentiation in N2B27. Unexpectedly, the pulse of RA was sufficient to start GFP expression in most of the cells, in a self-sustained way that appeared stable also after 24 or 96 additional hours from the removal of RA (Fig3.13b). I further tested whether the expression of the reporter was stable, by passaging and re-seeding the cells in N2B27, and I additionally purified the GFP^+ cells to analyze their transcriptome (Fig3.14). Even after re-plating and culturing the cells for 48 additional hours after the 96 hours already passed from the RA removal, the population preserved seemingly intact high GFP fluorescence (Fig3.14a). Compared to GFP^+ cells purified from six days of continuous RA differentiation, the cells derived from the RA pulse demonstrated higher expression of early neuroectoderm genes, such as *Sox1* and Zic1, and reduced levels of mature neural progenitors markers (Fig3.14b). Thus, exposing the cells for as little as 24 hours to high concentrations of RA is sufficient to start a self-sustainable neural program. I additionally tried to understand the role of RA concentration on this phenomenon by differentiating the cells for six days either in N2B27 and N2B27 vit.A, or with a pulse of RA at concentrations ranging from 0.1 nM to the usual 1µM, followed by N2B27 till the end. The lowest concentration tested of 0.1 nM seemed not to have affected the course of the differentiation, resembling the N2B27 condition (Fig3.14c). However, any higher concentration adopted, from 1nM to 1µM (10nM and 100nM data not shown), led to the acquisition of the high GFP state by most of the cells, highlighting how little RA is necessary for the occurrence of this effect (Fig3.14c). The results here reported point out a broader action of the RA pathway regarding neural fate decision, rather than being only an inductive signal for *in vitro* protocols.

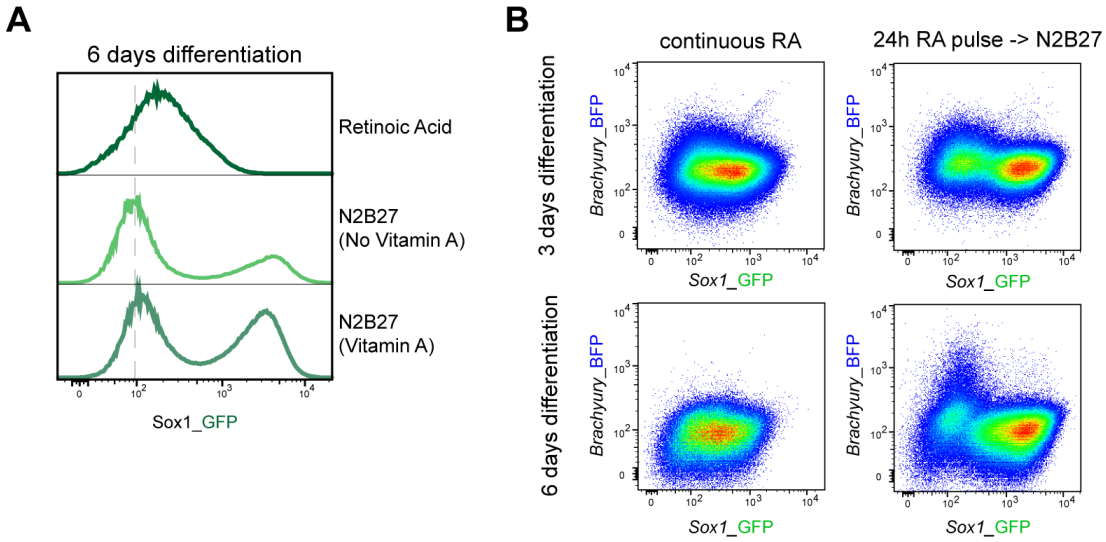


Figure 3.13: Retinoic acid signaling control of neural lineage specification.

Neuroectoderm differentiation of pluripotent cells in the absence of serum in different formulations of the N2B27 medium. ***A****) Histograms depicting the GFP fluorescence intensity after 6 days of differentiation in standard 1µM RA regime, N2B27 not containing vitamin A or N2B27 formulation with vitamin A (N2B27 vit.A).* ***B****) Pseudocolor density plots comparing three and six days of differentiation with 1µM RA protocol or pulsing RA activation for 24 hours and continuing the differentiation in N2B27.*

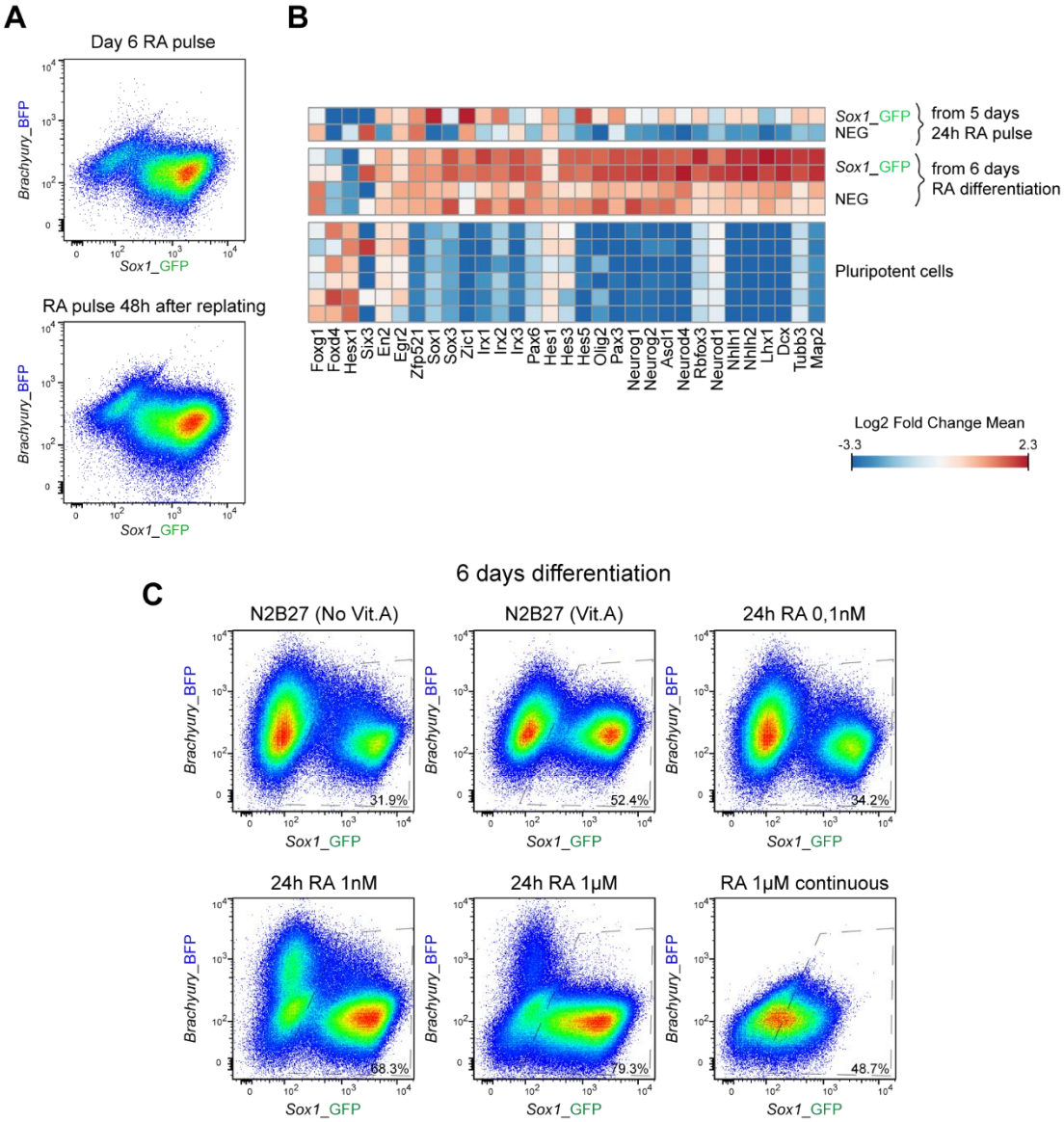


Figure 3.14: Stable GFP expression induced by the RA pulse marks neural cell types.

__A__) Cells showing stable GFP fluorescence at high levels after a RA pulse and 96 hours in N2B27 (upper plot), and following additional passaging and re-seeding in N2B27 onto laminin coated dish for 48 hours (lower plot). __B__) Heatmap showing the relative expression of neuroectoderm specific genes in cells isolated after 24 hours of RA pulse and additional 72 hours of culture in N2B27 compared to cells continuously exposed to RA, and pluripotent mESCs. __C__) Pseudocolor density plots comparing the results of six days of differentiation after 24 hours of RA pulse using different starting concentrations.

3.11 Generation of a retinoic acid sensor line to detect RA signaling during acquisition of neural fate

As indicated by the RA pulse experiments, the presence of continuous RA stimulation in the standard differentiation protocol is necessary to render the $Sox1^+$ state as transitory as possible. Nevertheless, the temporal relationship between activation of the RA signaling and acquisition of the neuroectoderm fate was still elusive. To determine the activation status of the RA pathway at any moment during the differentiation, I decided to generate a new triple reporter line, which in addition to *Sox1*_GFP and *T*_BFP, would bear a reporter of RA signaling. I took as model the system used by Rossant and colleagues to detect RA signaling during mouse embryogenesis, based on an hspLacz reporter controlled by the retinoic acid Responsive Elements (RAREs) contained in the *RARβ* promoter (Rossant et al., 1991). Therefore, I generated a triple knock-in line (2KI-RA) in which the RAREs drive the expression of an NLS_SCARLET_PEST2D reporter. To test both the sensitivity of the line and the effect of different amounts of RA on the GFP expression, I treated the 2KI-RA cells with various concentrations of RA for 24 hours. Detection of $Scarlet^+$ cells was possible from as little as 1nM RA, whereas the intensity of the fluorescence raised constantly with the increase of the concentration up to 1µM, condition in which the highest fraction of cells was found to have active RA signaling along with the highest fluorescence intensity (Fig3.15a). The progressive increase of Scarlet fluorescence was also associated with gradually higher GFP fluorescence peaking at 100nM RA (Fig3.15a). I verified whether the $Scarlet^+$ cells were indeed characterized by active RA signaling, isolating the cells according to Scarlet fluorescence, GFP fluorescence, or both after 36 hours of 100nM RA stimulation, and analyzing the expression of RA target genes. Comparing their transcription profiles, I observed that the two $Scarlet^+$ categories showed upregulation of direct RA targets as *Cyp26a1*, *Stra6*, *Stra8* and *Cdx1* among the others, whereas the neuroectoderm markers *Sox1*, *Zic1* and *Ascl1* were more expressed in the Scarlet $^-$ GFP^+ population (Fig3.15b).

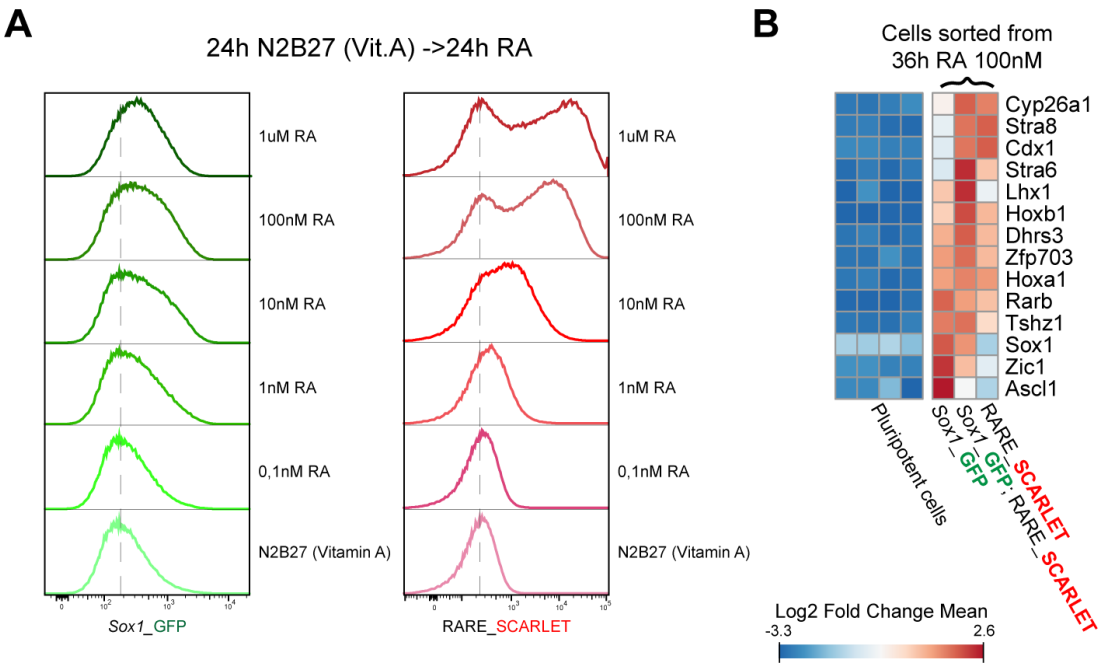


Figure 3.15: Retinoic acid signaling detection in combination with neural specification through a RA sensor line.

A) Histograms representing GFP and Scarlet fluorescence intensity changes after 24 hours of differentiation of the RA sensor line with the indicated concentration of RA. ***B)*** *Heatmap showing the relative expression of various RA target gens and neural markers upon 36 hours of differentiation with 100nM RA. The conditions analyzed are GFP$^+$ Scarlet$^-$, GFP$^+$ Scarlet$^+$ and GFP$^-$ Scarlet$^+$ RA sensor line cells sorted, and pluripotent mESCs.*

3.12 Retinoic acid induced specification towards neuroectoderm lineage monitored via the sensor line

To gain insights into the temporal regulation of the acquisition of neural fate in relation to RA response, I used flow cytometry to monitor the changes in Scarlet and GFP expression every 24 hours after stimulation with RA (100nM) for a total of four days. Within 24 hours, most of the cells exhibited activation of Scarlet reporter and the GFP fluorescence began to rise, reaching the maximum of GFP$^+$ cells between 48 and 72 hours (Fig3.16a). On the contrary, the fraction of cells expressing Scarlet, decreased after the first 48 hours. Unexpectedly, the detection of Scarlet and GFP fluorescence in the same cells was quite rare, the two appearing rather anti-correlated for most of the time points examined, with cells expressing Scarlet to high level displaying lower levels of GFP and vice versa. The co-expression was observed mainly at early time points and among cells with intermediate levels of both reporters. I

noticed a similar behavior also by performing confocal imaging every 24 hours of cells undergoing neural differentiation induced by 100nM RA (Fig3.16b). When I adopted 1µM of RA, instead of using 100nM, the intensity of Scarlet fluoresce after 48 hours was higher, whereas GFP fluorescence was remarkably lower, rendering the double positive population not as distinct as the 100nM RA condition (Fig3.16c).

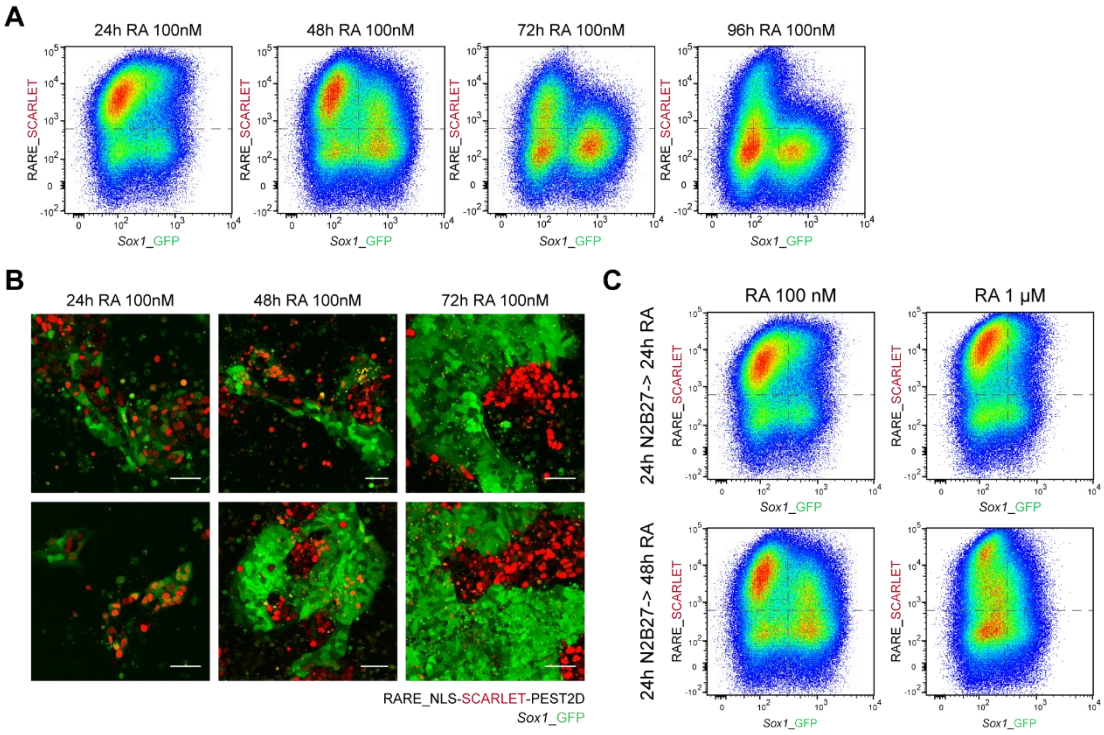


Figure 3.16: Dynamics of RA signaling activation during neurectoderm specification.

All the differentiations were carried out using the 2KI-RA sensor line in N2B27 with vitamin A regime, adding RA after 24 hours from the seeding. ***A****) Pseudocolor density plots of cells differentiated for five days towards neuroectoderm and analyzed by flow cytometry every 24 hours after RA addition.* ***B****) Confocal imaging of cells differentiated for 72 hours with RA, acquiring images every other 24 hours. Scale bar=50µm.* ***C****) Pseudocolor density plot depicting the change of GFP and Scarlet expression, after 24 and 48 hours from the addition of either 100nM or 1uM RA. The dotted line separates into four quadrants the space according to expression of Scarlet and GFP.*

To reconstruct the order of events following RA addition to the media, I performed a time-lapse imaging experiment using a derivative of RA sensor line bearing a constitutively expressed H2B-iRFP reporter to visualize all the nuclei. Within 4 hours from the addition of RA, Scarlet fluorescence was already detectable in some cells, although the fluorescence levels kept increasing in the same cells at least

for the first 12 hours (Fig3.17). Between 20 and 24 hours GFP^+ cells also started to appear, originating prevalently from cells expressing Scarlet. As already observed from flow cytometry, cells with an intermediate level of Scarlet fluorescence were more likely to co-express the GFP reporter as compared to high Scarlet ones (Fig3.17). Overall, these experiments indicate that the RA sensor line enable rapid and sensitive detection of RA signaling during neuroectoderm specification and it is suitable to study the mechanisms underlying both processes at the same time.

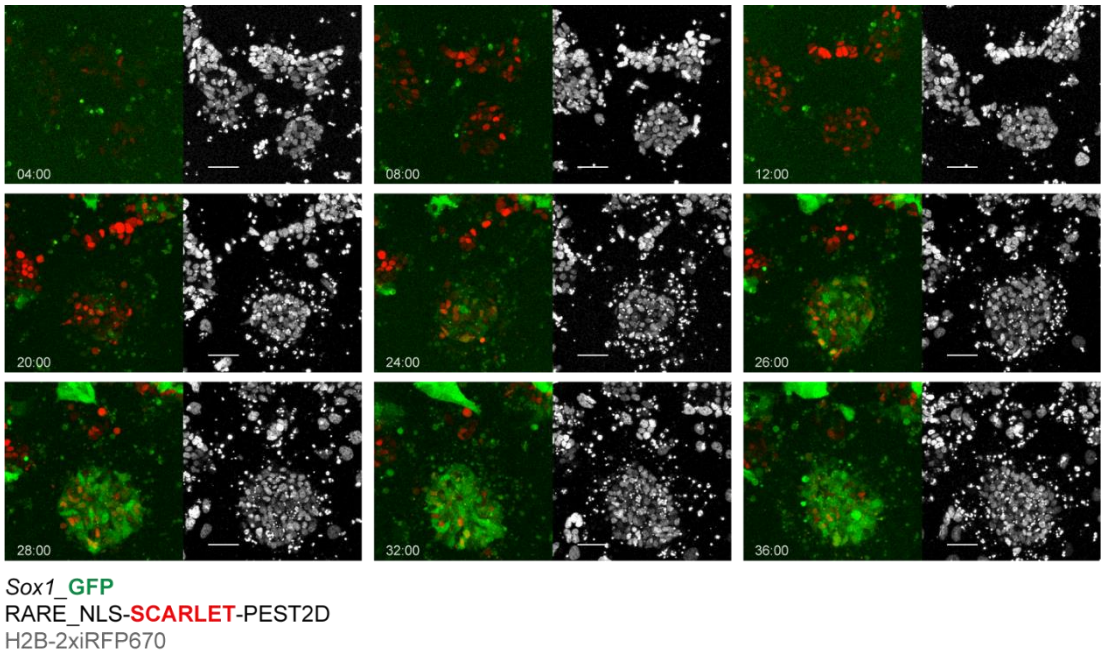


Figure 3.17: Retinoic acid activity and acquisition of neural fate monitored by the RA sensor line.

Time-lapse imaging experiment showing the order of events during neural fate acquisition, from RA response to expression of the GFP reporter. The experiment started upon the addition of 100nM RA to the medium after 48 hours of culture in N2B27. The time indication refers to hours from the beginning of the experiment. The H2B-iRFP reporter was used to visualize all the cell nuclei and automatically set the focus plane. Scale bar=50µm.

3.13 Detection of retinoic acid signaling in the Primitive Streak-like population, and its effects on the acquisition of neural fate

The fact that exogenous RA addition can elicit formation of neuroectoderm from both the PS-like population and pluripotent mESCs, clearly highlight the existence of a mechanism controlled by this pathway enabling the formation of neural cell types.

However, identifying an endogenous RA production from the PS-like population would allow to study the function of this pathway among the other neural induction activities, a context resembling more closely early phases of mouse development. Almost all RA production of the early mouse embryo is generally attributed to the ALDH1A2 enzyme, implicated in the last step of RA synthesis (Fig3.18a). Thus, the expression of *Aldh1a2* was an important condition to be met for the presence of endogenous activation of the RA signaling. Analyzing the mRNA levels of *Aldh1a2* during the PS-like differentiation, it is possible to notice that the gene is quite specifically upregulated in the BFP/PS-like population as compared to pluripotent cells and the GFP^+ cells from the same differentiation (Fig3.18b). To further test the presence of endogenous RA activation, I adopted the RA sensor line in the PS promoting condition. Scarlet positive cells were detected starting from three to four days of the differentiation, but even more remarkably these cells were also positive for the BFP reporter, meaning that they are a subgroup of the PS-like population and its derivatives (Fig3.18c). The relevance of these findings derives from the fact that *Aldh1a2* expression and RA signaling activity are detected in the PS region in the early mouse embryo as well (Ribes et al., 2009). Encouraged by the presence of endogenous RA signaling and the described capacity of RA to control the neuroectoderm fate, I decided to explore the interdependence between the pathway and the fate determination in this system. For this reason, I opted for a pharmacological approach to test the effect of perturbing the RA signaling, using AGN or vitamin A instead of RA to stimulate the activation of the pathway by increased RA production. As anticipated in Fig3.12, the AGN inhibitor, blocking RA signaling, caused the abrogation of Scarlet fluorescence and resulted in a partial but reproducible reduction of the GFP^+ fraction by the end of the differentiation (Fig3.18d). Adding vitamin A, instead, led to a rapid boost of activation of the pathway, as proved by the increased fraction of $Scarlet^+$ cells, followed by an augmentation of the fraction of cells expressing GFP (Fig3.18d). The last observations clearly denote the existence of an RA-dependent mechanism of neuroectoderm induction from the PS, and entail that the amount of RA precursors is a limiting factor for the acquisition of neural fate controlled by RA.

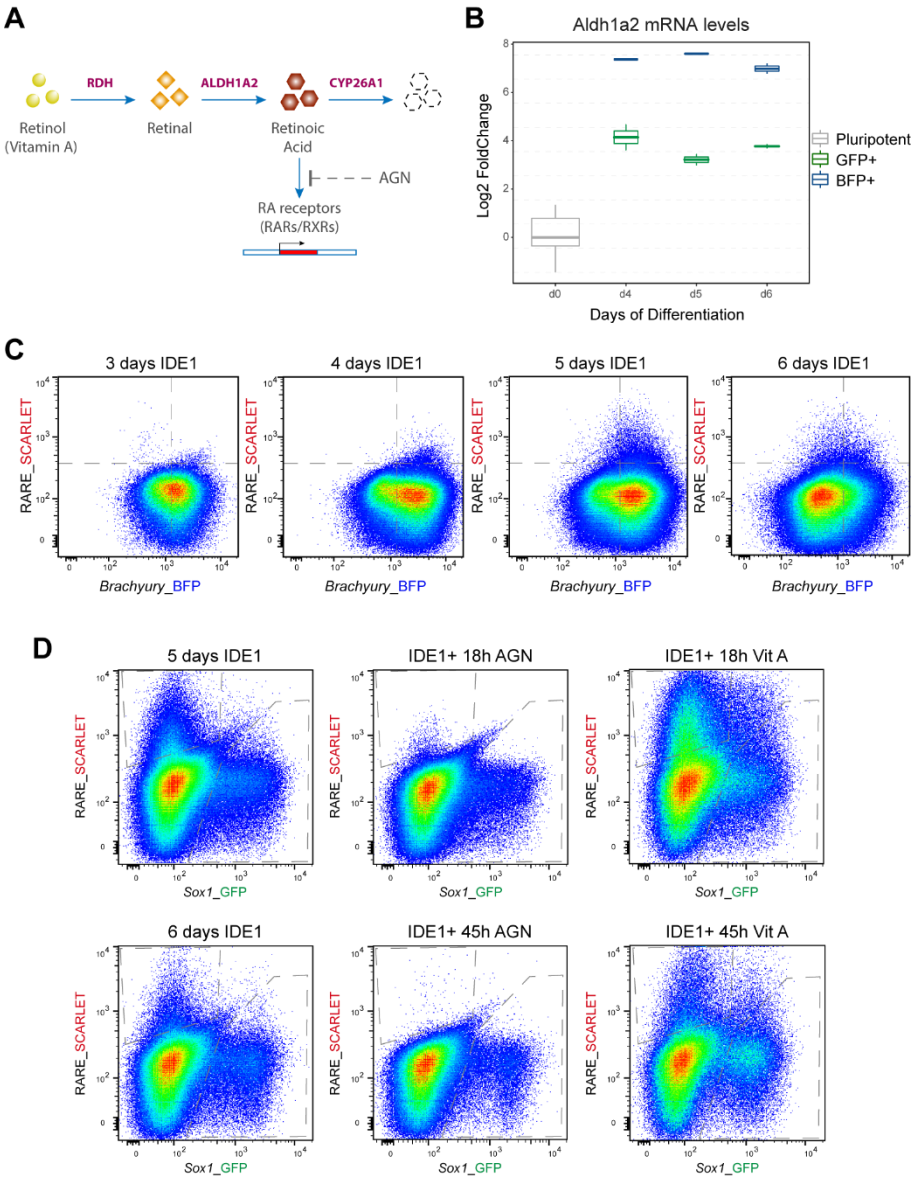


Figure 3.18: Detection of endogenous RA signaling during the Primitive Streak formation and consequences of its perturbation on neural fate acquisition.

__A__) Schematic depicting the main steps of RA metabolism and signaling. The main enzymes catalyzing each reaction are reported. __B__) Box plot of Aldh1a2 relative expression measured as Log_2 of the fold change of the normalized read counts in the BFP^+ and GFP^+ populations sorted from IDE1 differentiation compared to pluripotent mESCs. __C__) Pseudocolor density plots showing the onset of $Scarlet^+$ cells within the BFP^+ population during six days of IDE1 differentiation of the 2KI-RA sensor line. __D__) Pseudocolor plots comparing Scarlet and GFP expression from the 2KI-RA line after five and six days of the standard IDE differentiation, RA inhibition with AGN in addition to IDE1 for the last 18 and 45 hours before the analysis, or stimulation of RA production adding vitamin A to IDE1 from the last 18 and 45 hours before measurement.

3.14 The degrading enzyme Cyp26a1 limits the response to retinoic acid

The fine tuning of RA signaling during development is achieved by controlling not only its sites of production but also its degradation. As for *Aldh1a2*, the transcripts coding for the RA degrading enzyme Cyp26a1 were particularly enriched in the PS-like BFP^+ cells, hinting that this crucial mechanism for the control of RA response could be in place also in our scenario (Fig3.19a). As an alternative approach to explore the effects of RA signaling perturbation, I decided to introduce null mutations in both *Cyp26a1* alleles (Fig3.19b). Before RA signaling begins, which according to the data in Fig3.18c occurs between three and four days of IDE1 differentiation, the absence of Cyp26a1 causes no alteration of the differentiation propensity (Fig3.19c). However, the measurement of five days of differentiation of the knock-out line exhibited clearly more cells fated to neural lineage. The difference was even more evident when RA signaling was enhanced by providing more substrate for its production adding vitamin A. The supply of vitamin A shifts the balance of the fates in favor of neuroectoderm even in wild type cells, but in absence of Cyp26a1 had even more dramatic effect, with most of the cells eventually expressing GFP within 48 hours (Fig3.19c). Looking at the RA response both in the Cyp26a1 KO condition alone and in the presence of vitamin A, revealed that the increase in the GFP^+ cells was combined with enhanced RA activation both in terms of the fraction of cells expressing Scarlet and its fluorescence intensity (Fig3.19d). From these results, I concluded that Cyp26a1 limits the response to RA both in terms of its intensity and of the fraction of cells responding.

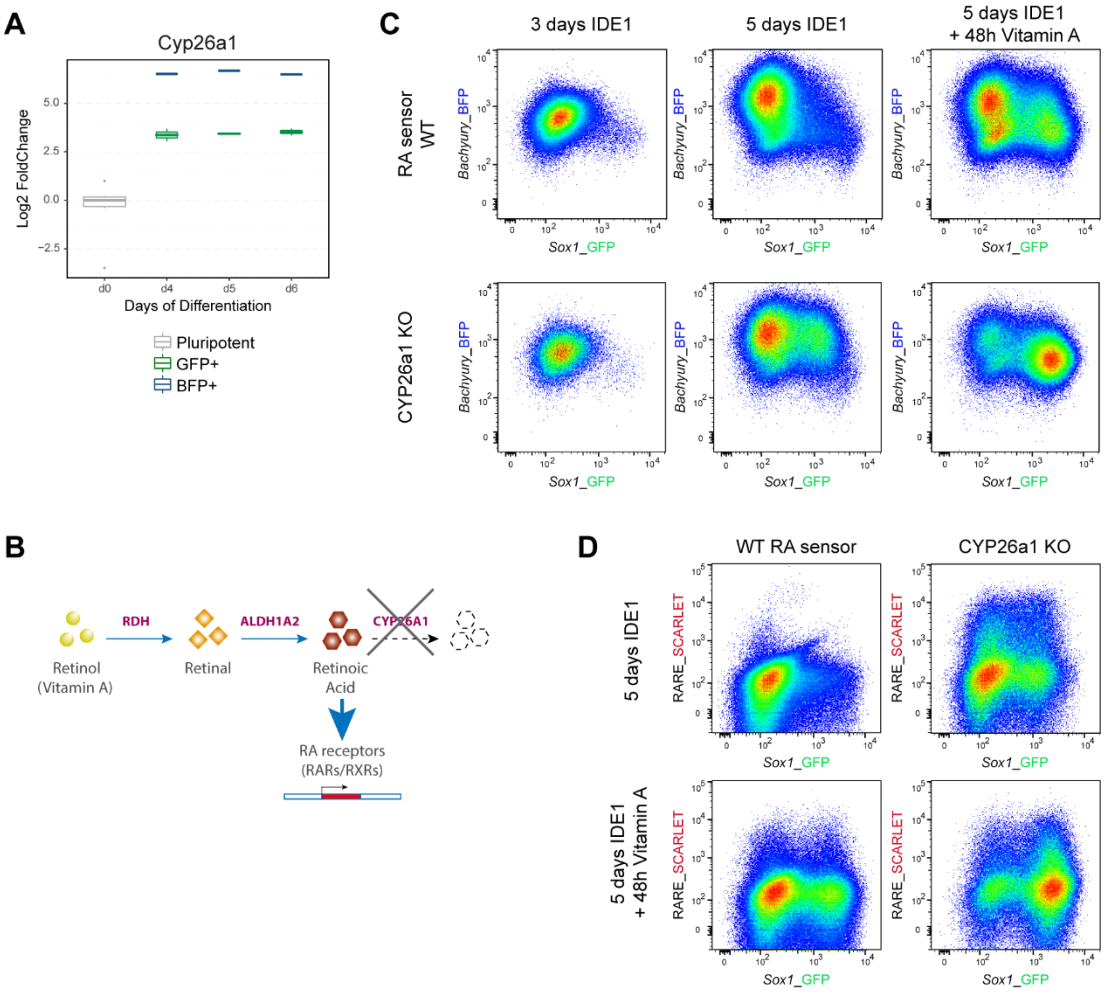


Figure 3.19: Cyp26a1 KO line response to endogenous RA and acquisition of Neural fate.

__A__) Box plot of Cyp26a1 relative expression measured as Log_2 of the fold change of the normalized read counts in the BFP^+ and GFP^+ populations sorted during IDE1 differentiation, compared to pluripotent mESCs. __B__) Schematic depicting the RA metabolism and main expected consequences of Cyp26a1 KO on RA signaling. __C__) Pseudocolor density plots displaying the effects of Cyp26a1 KO on neuroectoderm fate specification measured after three or five days of standard IDE1 differentiation, or adding vitamin A in the last 48 hours before measurement. __D__) Alternative representation of the conditions in panel C, showing the changes in RA sensor activation upon Cyp26a1 KO in comparison to the WT RA sensor line.

To gain more insights about the mechanism of action of Cyp26a1 and of RA degradation, I investigated further the behavior of the knock-out cells when different concentrations of RA were applied exogenously or when RA signaling was stimulated only transiently. Differentiating the wild type sensor line and the Cyp26a1 KO line for 96 hours with either 1 nM or 100 nM RA, highlighted some interesting differences

(Fig3.20a). At low concentrations (1nM RA), the differences in Scarlet expression between WT and Cyp26a1 KO line were minimal, although the latter presented, after 96 hours, cells in which the response reached higher levels compare to WT (Fig3.20b). At 100nM RA the dynamics of the RA activation are slightly different according to the presence or absence of the degrading enzyme. In the KO line, in fact, the peak of response in terms of Scarlet fluorescence intensity was reached already within 24 hours, with most of the cells displaying remarkably high fluorescence before its rapid decline continuing the differentiation. In the line with wild type Cyp26a1 alleles, the highest Scarlet fluorescence was reached only between 24 to 48 hours and with a peak lower than in the Cyp26a1 KO. In addition, it was possible to observe the persistence of a group of cells showing high Scarlet levels by 96 hours of treatment, meaning that they are likely delayed responders, much reduced in the Cyp26a1 KO (Fig3.20a). Analyzing the outcome of these differences on the neural fate decision, it was clear that at 1 nM RA the Cyp26a1 KO exhibited considerably more cells with high GFP fluorescence than the WT counterpart. At high RA concentration, the differences in terms of GFP expression trends were reduced, with the exception of the 48 hours time point, when in the Cyp26a1 KO condition more cells responded expressing the GFP reporter (Fig3.20a). To test whether Cyp26a1 absence would impair the possibility of the RA pulse to induce the stable high GFP population, I treated both Cyp26a1 KO and WT cells with various RA concentrations for 24 hours and continued the differentiation with N2B27 alone for additional four days (Fig3.20c). Continuous RA treatment, or pulse ranging from 1 nM to 1μM of RA, showed no difference both in presence and absence of Cyp26a1, entailing that the degradation of RA is not a primary mechanism for the pulse phenomenon. Nevertheless, in the Cyp26a1 KO as little as 0.1nM RA became sufficient to drive GFP expression and high fluorescent levels in most of the cells (Fig3.20c). Taken together these results seem to point to a role of Cyp26a1 as crucial threshold barrier against the RA response; when the degrading enzyme is removed, the threshold to overcome in order to start the RA signaling is much lower and this is particularly evident when using low concentrations of RA rather than high ones.

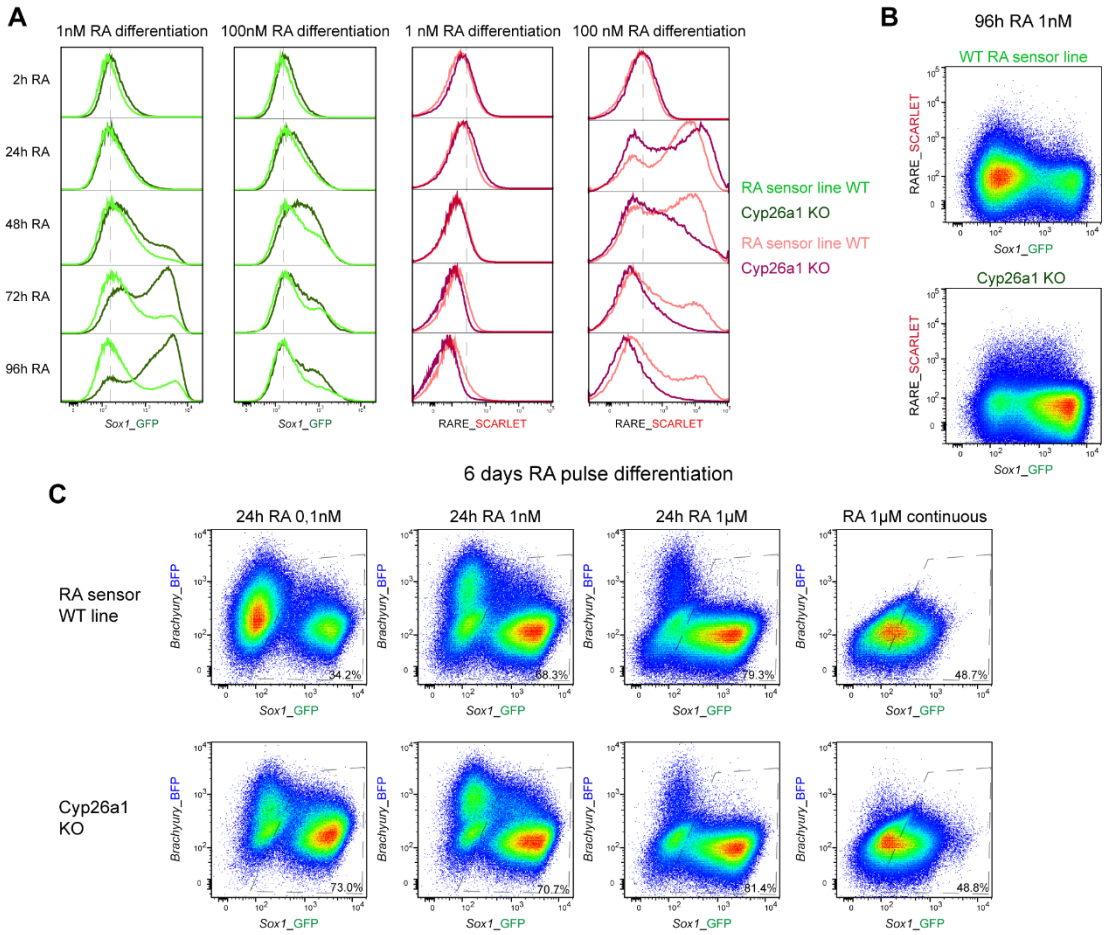


Figure 3.20: Enhanced response to exogenous RA in the Cyp26a1 KO line.

__A__) Histograms depicting the differences in GFP and Scarlet levels of 2KI-RA sensor line and Cyp26a1 KO line differentiated for 96 hours in N2B27 with vitamin A, supplying either 1nM or 100nM RA after the first 24 hours. __B__) Pseudocolor density plots presenting GFP and Scarlet levels 96 hours after the differentiation of the WT RA sensor line and Cyp26a1 KO line with 1nM RA. __C__) Pseudocolor plots comparing the effects of 24 hours of RA pulse at different concentrations, in the Cyp26a1 KO line and WT RA sensor line.

3.15 RA signaling perturbation affects neural induction by both Wnt or Nodal/TGF-β inhibition

The various neuroectoderm types induced from the PS-like population seemingly derived from different inductive stimuli, among which RA, that is both present and plays a role in this context. Therefore, I sought to investigate the presence of interplay between the different mechanisms of neural induction and whether perturbing RA signaling would affect other means of promoting the neural fate. In order to do that, I carried out low serum-IDE1 differentiation with both the WT RA sensor line and the Cyp26a1 KO line and applied XAV, SB43 or AGN from the last three days of differentiation to block respectively Wnt, Nodal/TGF-β or RA pathway (Fig3.21). RA inhibition rescued the increase of the GFP^+ fraction consequent to the absence of Cyp26a1. Nodal/TGF-β inhibition, as previously discussed, resulted in higher proportion of cells fated to neural lineage and, its effects were even more dramatic when combined to the Cyp26a1 KO, causing expression of GFP from majority of the cells. This effect was instead reversed when TGF-β and RA inhibition were combined, with a clear reduction of the final fraction of GFP^+ cells (Fig3.21a). Similarly, the combination of XAV and AGN had even more striking effect because blocking RA signaling completely prevented any increase of neural cell types by Wnt inhibition (Fig3.21b). In sum, the perturbation of the RA pathway by either AGN or Cyp26a1 KO highlights the involvement of this signal in the acquisition of neural fate by Wnt or Nodal/TGF-β inhibition.

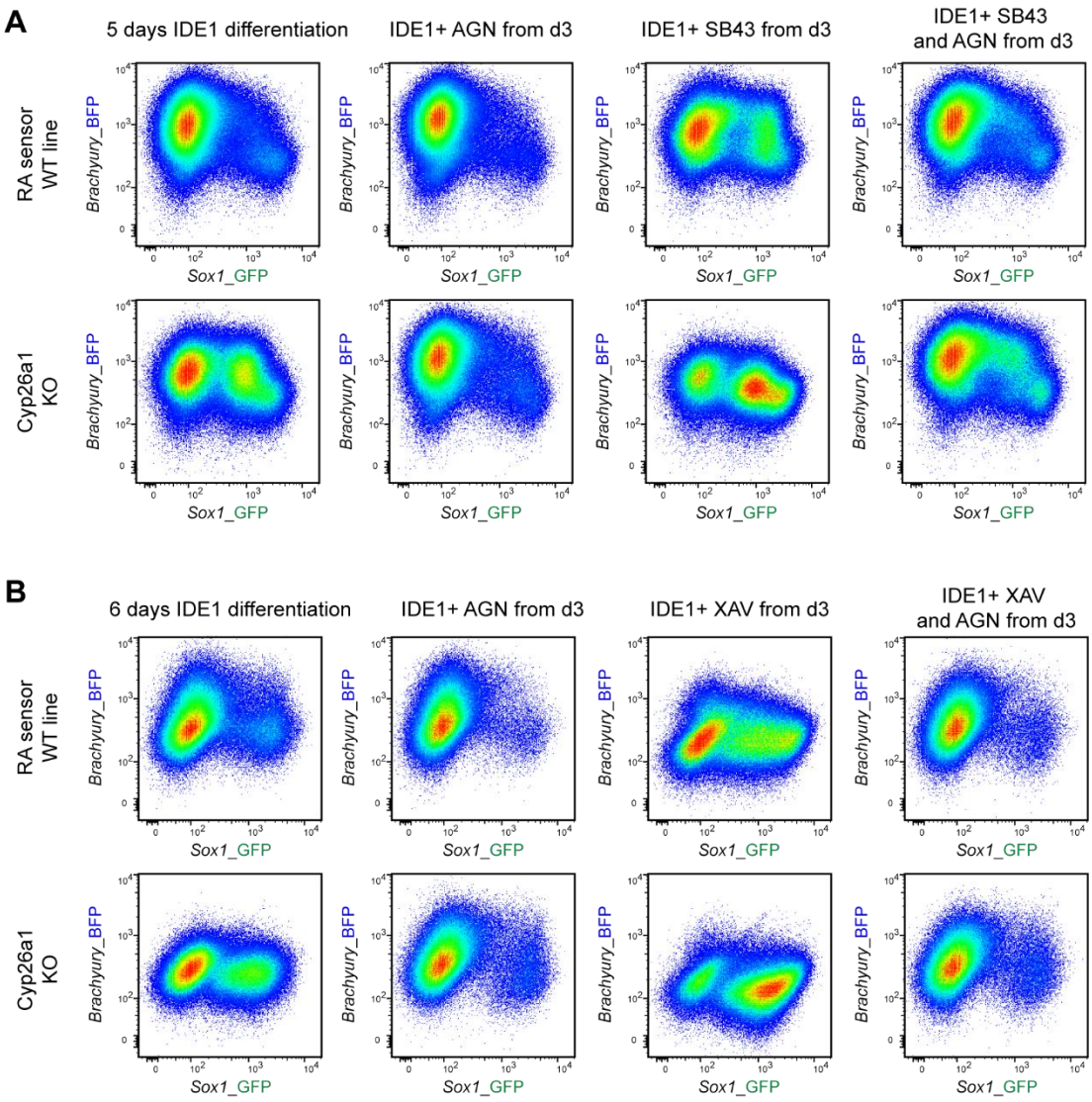


Figure 3.21: Interplay between RA signaling and Wnt or Nodal/TGF-β inhibition in the induction of neuroectoderm during the Primitive Streak-like differentiation.

__A__) Interplay between RA and Nodal/TGF-β pathway explored by comparing GFP and BFP level changes in the 2KI-RA sensor line and Cyp26a1 KO upon five days of standard IDE1 differentiation, or five days of differentiation carried out in IDE1 with addition of AGN, SB43 or both in the last 48 hours before the analysis. __B__) Interplay between RA and Wnt pathway investigated as in the panel A, by comparing GFP and BFP level changes in the 2KI-RA sensor line and Cyp26a1 KO upon six days of standard IDE1 differentiation, or six days of differentiation carried out in IDE1 with addition of AGN, XAV or both in the last 72 hours before the analysis.

3.16 Residual RA activation in the *Aldh1a2*$^{-/-}$ RA sensor line

The phenotype of mice bearing mutations in the *Aldh1a2* gene, exhibiting lack of RA signaling but apparently normal induction of forebrain structures, led to consider RA as a factor transforming the emerging neuroectoderm, rather being involved at the induction stage. Nevertheless, the combination of the findings above presented points out the role played by the RA in the induction of neural lineage. Thus, I decided to attempt the *in vitro* recapitulation of this phenotype, mutating the *Aldh1a2* gene in the RA sensor line and observing the outcomes of its manipulation (Fig3.22a). When differentiated in IDE1 regime, the Aldh1a2 KO line showed no expression of Scarlet reporter and RA signaling activity (Fig3.22b). Differently from the effect of the inhibition of the pathway by AGN, the absence of RA activation in this case was associated with seemingly normal fraction of GFP^+ cells (Fig3.22b). Additional differences which could be appreciated were a slightly higher BFP level and more cells showing both BFP and GFP fluorescence at a given time point (Fig3.22b). To test whether the Aldh1a2 KO cells were truly incapable of producing RA and RA signaling was absent, I used AGN or vitamin A supply. Surprisingly, even though the disappearance of Scarlet expression in the standard differentiation was hinting to an absence of residual RA, the addition of AGN could further decrease the fraction of GFP^+ cells (Fig3.22c). Moreover, in the presence of vitamin A cells exhibiting intermediate levels of Scarlet fluorescence re-emerged, while almost all expressed GFP (Fig3.22c). The immediate implication of this observation is that the RA production is not entirely extinguished when Aldh1a2 is removed, at least in this *in vitro* scenario. The second implication is that extremely low concentration of RA might induce the acquisition of neural lineage without eliciting Scarlet expression. To test this possibility, I added sub-nanomolar concentrations of RA for 48 hours to the Aldh1a2 KO line in the IDE1 differentiation. As hypothesized, no Scarlet expression was detected, but still such low amount of RA slightly increased the fraction of GFP^+ cells compared to standard differentiation (Fig3.22d). Therefore, contrary to what was indicated by the *Aldh1a2* null mouse phenotype, the KO results in the RA sensor line argue for a residual activation of RA signaling and for the possibility to still metabolize vitamin A.

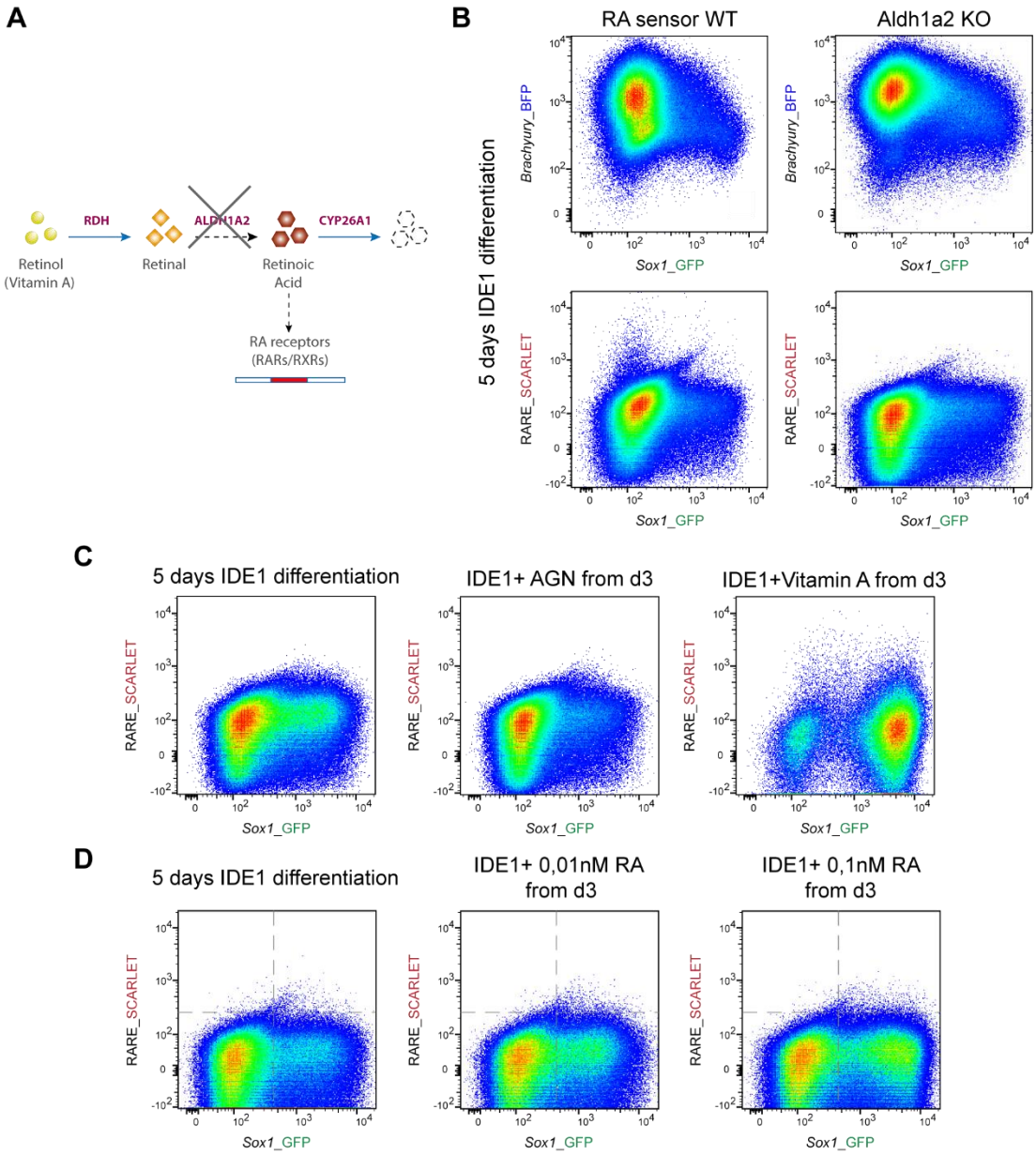


Figure 3.22: In vitro recapitulation of the Aldh1a2 null genotype in the RA sensor line.

__A__) Schematic depicting the RA metabolism and main expected consequences of Aldh1a2 KO on RA signaling. __B__) Pseudocolor density plots showing the effects of Alh1a2 mutation in the RA sensor line on the Primitive Streak population (BFP), neuroectoderm induction (GFP) and RA activation (Scarlet). __C__) Pseudocolor plots presenting the differences in Scarlet and GFP expression of the Aldh1a2 KO line after five days of IDE1 differentiation, upon inhibition of RA signaling with AGN, or providing vitamin A in the last 48 hours before measurement. __D__) Pseudocolor density plot of Aldh1a2 KO cells differentiated for five days either in standard IDE1 condition or adding for 48 hours 0.01 and 0.1 nM RA, three days after the beginning of the differentiation. The dotted line separates into four quadrants the space according to expression of Scarlet and GFP.

3.17 The WT and the KOs RA sensor lines respond differently to increasing doses of vitamin A

To better understand the balance between RA synthesis and degradation and its consequences on the neural fate decision, I decided to compare the behavior of the 2KI-RA sensor line, the Cyp26a1 KO line, and the Aldh1a2 KO line in response to increasing concentrations of vitamin A. The Cyp26a1 KO condition showed higher sensitivity to the presence of vitamin A, with the response reaching the plateau at lower vitamin A concentrations and greater fraction of cells presenting Scarlet fluorescence compared to the WT control (Fig3.23a). The enhanced RA response was associated to consistently higher fraction of GFP^+ cells. In the case of *Aldh1a2* mutation, the heterozygous condition showed decreased response to any vitamin A concentration from the point of view of both $Scarlet^+$ and GFP^+ fraction. On the contrary, *Aldh1a2* null mutant revealed dose dependent and more penetrant response to vitamin A, and the percentage of cells expressing Scarlet raised along with the vitamin A concentrations, eventually involving a higher fraction of the entire population than in WT control. This behavior was associated with detection of the GFP reporter in most of the cells even at the lowest concentrations of vitamin A tested (Fig3.23a). Another interesting aspect of the Aldh1a2 KO was about the change in fluorescence intensity of the BFP and GFP reporter (Fig3.23b). In the standard differentiation, in fact, the Aldh1a2 KO line reported higher BFP fluorescence than both WT sensor line and the Cyp26a1 KO; but when vitamin A was added at any of the concentrations tested, the BFP fluorescence decreased to the same levels of the other two lines (Fig3.23b). On the other hand, GFP displayed a basal wider range of fluorescence in Aldh1a2 KO than in the other two conditions, and from the lowest concentration of vitamin A tested its levels became considerably higher (Fig3.23b). The results indicate that the Aldh1a2 KO cells are still able to process vitamin A, although less efficiently, and even more cells than WT counterpart respond to RA signaling.

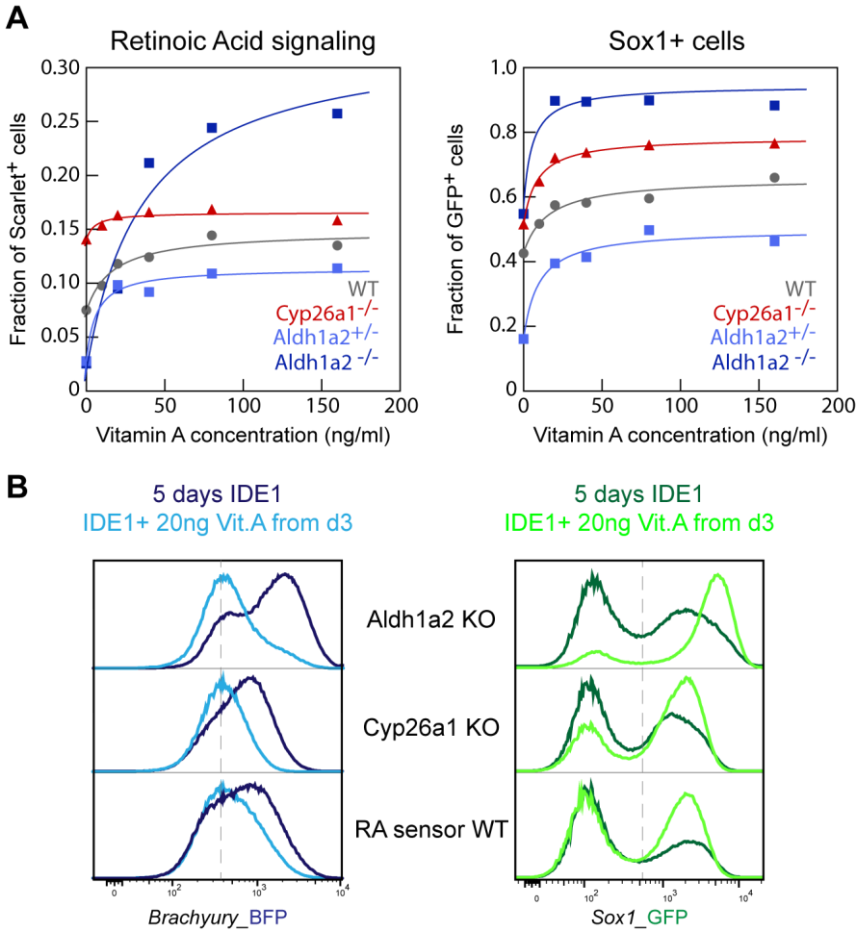


Figure 3.23: Different responses of WT and KOs RA sensor lines to increasing concentrations of vitamin A in IDE1 differentiation.

A) Trends of Scarlet$^+$ and GFP$^+$ cell fractions of the WT RA sensor line, Cyp26a1 KO, and Aldh1a2 heterozygous and null mutant lines, differentiated for five days in IDE1, providing various concentrations of vitamin A for the last 48 hours before the analysis. ***B****) Histograms showing the variations in BFP and GFP fluorescence levels after five days of differentiation in the indicated conditions.*

Intrigued by the fundamentally opposite defects of the Cyp26a1 and Aldh1a2 KO lines with respect to the RA metabolism, I sought to explore their behavior when they were differentiated in combination. For this reason, I carried out a co-culture experiment using the Cyp26a1 KO line and a derivative of the Aldh1a2 KO line, bearing the H_2B-iRFP reporter, to be able to distinguish the two KO clones by detecting iRFP fluorescence. I seeded the two lines together in different proportions, and then analyzed after five days of differentiation the expression of reporters in the whole population and in the individual KOs (Fig3.24). Looking at the mixed population, the higher the fraction of the Aldh1a2 KO the lower the Scarlet$^+$ cells

proportion and the percentage of GFP^{+} cells compared to all the cells measured (Fig3.24). Taking in consideration the single contribution of Cyp26a1 KO or Aldh1a2 KO, in the latter case I noticed the reactivation of the Scarlet reporter and that most of the Aldh1a2 KO cells presented GFP fluorescence when added at lower ratio (Fig3.24). These results demonstrate the paracrine action of RA in this system enabling Scarlet expression in the Aldh1a2 KO line, but also that the Cyp26a1 KO increases the availability of RA in the medium, having on the GFP expression of Aldh1a2 KO the same effect as providing vitamin A.

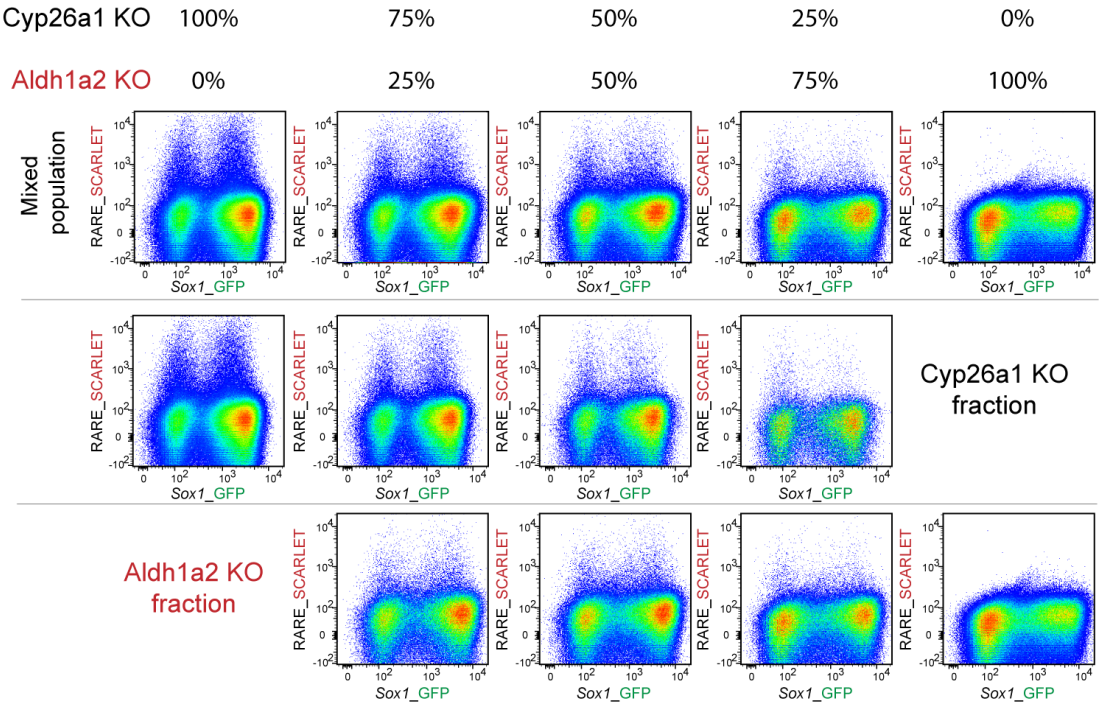


Figure 3.24: Co-culture of Cyp26a1 KO and Aldh1a2 KO lines during Primitive Streak-like differentiation.

The Cyp26a1 and Aldh1a2 KO lines were mixed in different proportions at the moment of the cell seeding and differentiated for five days in the standard IDE1 condition. An Aldh1a2 KO line bearing the constitutively expressed H2B-iRFP reporter was used in this experiment to discriminate BFP, GFP and Scarlet fluorescence deriving from the Aldh1a2 KO and the Cyp26a1 KO. The first row represents to the entire samples analyzed, whereas the second and third rows derive from the separation of Cyp26a1 KO and Aldh1a2 KO contributions to each fluorescent channel.

3.18 The RA receptors knock-out renders the cells insensitive to RA signaling

An additional layer of regulation of the RA response is the availability of its receptors, the main executors of RA actions during development. Unfortunately, studying the functions of individual RA receptors during development has proven particularly challenging, owing to their ability to compensate each other functions and overlapping domain of expression. *In vitro* reconstitutions of the signaling machinery helped in understanding the basic principles of how the signal transduction works, but also generated artifacts. As a possible way to achieve a condition completely devoid of RA, I attempted the combined mutation in the 2KI-RA line of the three RARs, claimed to carry out almost all the developmental functions of RA, as compared to the other family of RA receptors, the RXRs. To evaluate indirectly the abrogation of RAR transduction, I looked for the absence of the typical consequences of RA activation: RA induced modification of GFP expression, parallel repression of *Brachyury* (BFP levels), and possibility to prevent these changes by AGN. Strikingly, the putative triple RAR KO presented no difference in the BFP and GFP fluorescence both with 100nM of RA or with RA and inhibiting concentration of AGN (Fig3.25a). Moreover, I tested the effect of a RXR specific agonist (LG268) and a RAR specific agonist (Ch55) and, intriguingly, I observed that in the triple RAR KO Ch55 did not cause any visible change of reporter expressions, whereas the RXR ligand reduced the BFP levels and slightly increased the GFP^+ fraction. The last observation confirmed the lack of the RAR functions, while showing that in absence of RARs, RXRs activation might still influence the acquisition of neural fate, although this effect is not observed using all-*trans*-RA as ligand (Fig3.25a). Analyzing the RA activation by measuring Scarlet expression, I found no reporter expression in all the conditions tested in the triple RAR KO, proving the absence of RA signaling (Fig3.25b). Of note, the RXR agonist induced expression of only few Scarlet cells even in the WT sensor line, whereas the RAR specific ligand is able to elicit strong Scarlet fluorescence in these cells. Overall, the knock-out of the three RARs seems to provide an unprecedented condition of absence of RA signaling hinted by the lack of the most common effects of this pathway in our system.

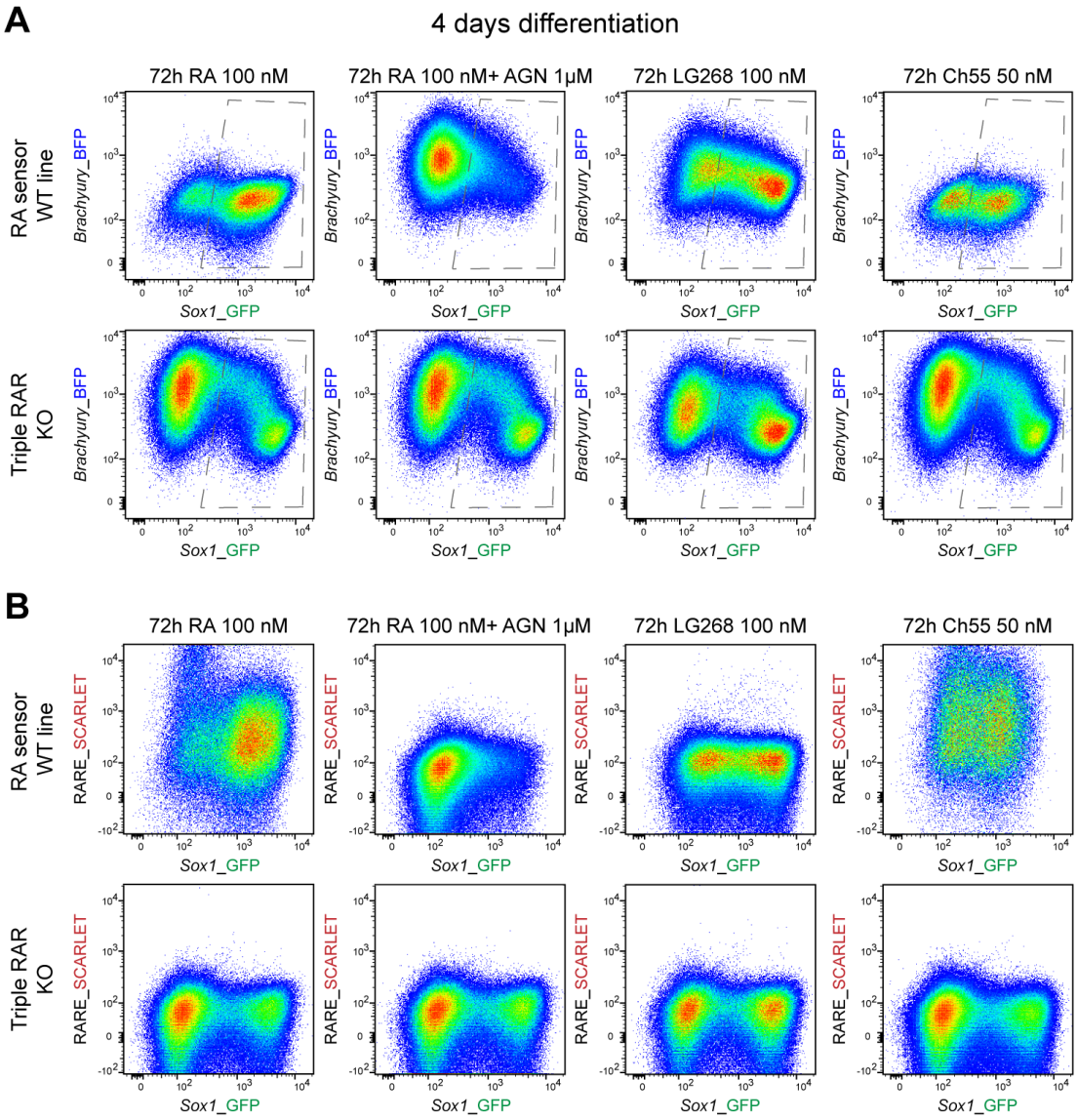


Figure 3.25: Triple RAR mutant line response to perturbation of the RA signaling.

The WT RA sensor line and the triple RAR KO line were differentiated for four days in N2B27 vit.A, adding 72 hours before the analysis 100nM RA; 100nM RA and inhibiting concentration of AGN (1µM); 100nM of the specific RXRs ligand, LG268; 50nM of the RAR agonist, Ch55. ***A-B****) Pseudocolor density plots showing the changes in GFP, BFP and Scarlet fluorescence in the reported conditions.*

After verifying the functional absence of RA transduction in the triple RAR KO, I sought to test its behavior in the PS-like culture regime, also in relation to other pathways. In a preliminary experiment I induced the differentiation of the triple RAR KO for five days in low serum/IDE1 condition and observed no expression of Scarlet, and insensitivity of the line to AGN also in this context (Fig3.26a and b). Unexpectedly, a considerable fraction of GFP^+ cells was found in standard differentiation of the triple KO, together with globally lower BFP levels (Fig3.26a). The inhibition of Nodal/TGF-β or Wnt pathways only slightly modified the fraction of GFP^+ cells while, only in the case of XAV, still decreasing BFP fluorescence (Fig3.26a). The reduced efficacy of Wnt and TGF-β inhibition in the triple RAR KO sustains the previously anticipated effect of blocking RA signaling by AGN being able to hinder both ways of neuroectoderm induction. Nevertheless, the presence of robust GFP expression in the basal condition of the triple RAR KO highlights an important aspect of the RA function possibly missing when RARs are removed, that is ligand-independent repression of target genes. This possibility complicates further use of RA receptors KOs to accomplish a state devoid of RA signaling.

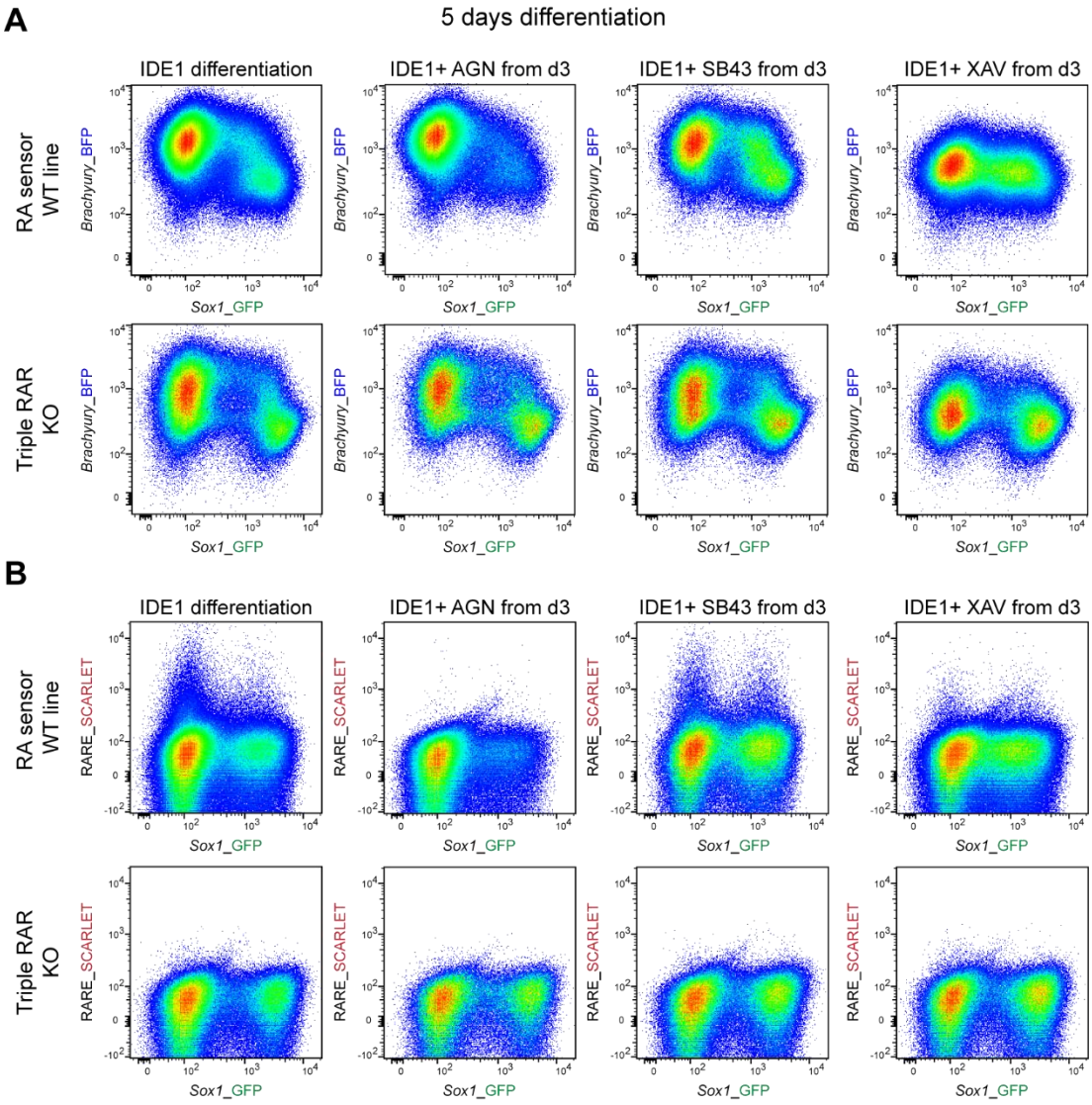


Figure 3.26: Differentiation of the triple RAR KO line in Primitive Streak-like condition, and inhibiting Wnt or Nodal/Tgf-β signaling.

The WT RA sensor line and the triple RAR KO line were differentiated for five days in standard IDE1 regime, or inhibiting for the last 48 hours of differentiation respectively RA (AGN), Nodal/Tgf-β (SB43) or Wnt (XAV) pathways before the flow cytometry analysis. ***A-B)*** *Pseudocolor density plots showing the changes in GFP, BFP and Scarlet fluorescence in the explained conditions.*

3.19 RA reporters based on different DNA binding topology reveal more widespread RA activation

The classical retinoic acid responsive elements (RAREs) consist of two direct repeats of the 5'-RGKTCA-3' consensus sequence separated by a variable number of base pairs from each other, according to which they are abbreviated $DR_{0\text{-}10}$ (direct repeat). One of the most common sequence in terms of both widespread presence in RA target genes and extensive characterization, is the DR5, which is also the one used in the original mouse reporter lines and in the RA sensor line presented in this project. However, the residual RA signaling in the Aldh1a2 KO, undetected by the quite sensitive DR5 reporter, and the possibility to influence the specification to the neural lineage without activating Scarlet expression (low RA concentrations; RXR agonist in the triple RAR KO), implies that the RA response might be more widespread than what is estimated by using this reporter. Looking at previous reports for alternative DR binding sites to test as novel RA responsive elements, there is a peculiar sequence found in the *Mafa* locus, which consist of the combination of three DRs: DR2, DR0 and DR8 (Moutier et al., 2012). The intriguing aspect of this regulatory region was the co-existence of the DR2, known to be responsive to RARs, with the novel DR8 and DR0 sequences, which could potentially account for a broader RA response. A new triple reporter line was generated, introducing in the *Sox1*_GFP and *Brachyury*_BFP 2KI line, the NLS-SCARLET-PEST2D reporter downstream of the *Mafa* RAREs. The use of the novel triple reporter line during the PS-like differentiation led to the detection of roughly ten times more cells expressing Scarlet (Fig3.27). The fact that Scarlet expression was almost entirely abrogated when AGN was added to the IDE1 differentiation confirmed that the reporter was responsive to RA (Fig3.27). In the attempt to investigate the effect of Wnt or TGF-β inhibition on this broader RA activation with the new line, I found persistence of $Scarlet^{+}$ cells in both conditions, particularly noticeable in the case of XAV, as it is barely detected with the classical RA reporter (Fig3.27). Despite being preliminary, these last results anticipate a possible change in perspective about the contribution of RA signaling to processes where it was not considered to be involved using the classical mouse reporter line.

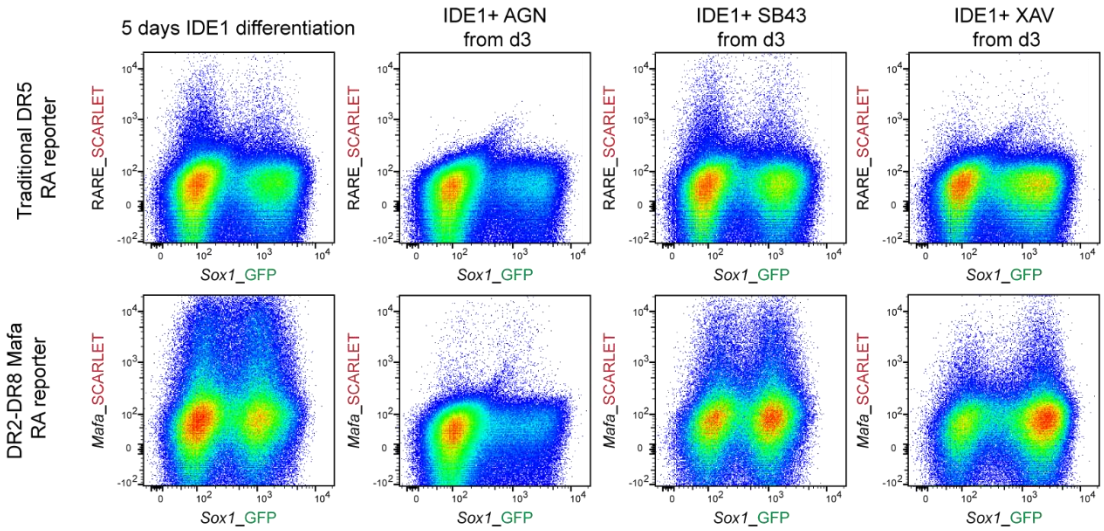


Figure 3.27: Primitive Streak-like differentiation of the novel Mafa RA sensor line and inhibition of Wnt or Nodal/Tgf-β signaling.

Pseudocolor density plots displaying the differences in the RA signaling activity detected by Scarlet using the classical DR5-based RA sensor line and the novel Mafa-RARE RA sensor line. The two reporter lines were differentiated for five days according to standard IDE1 protocol, or in IDE1 and additional inhibition respectively of RA (AGN), Nodal/Tgf-β (SB43) or Wnt (XAV) pathways for the last 48 hours of differentiation before the flow cytometry analysis.

4 DISCUSSION

Owing to the complexity of the processes occurring during gastrulation, the study of the formation of the primary germ layers has proven particularly challenging *in vivo*. Only in recent years, technological improvements allowed progresses in disparate directions, ranging from the acquisition of the complete transcriptomes of germ layers from mouse gastrula, to the development of culture systems with micropattern devices enabling the bidimensional distribution of geometrically ordered germ layers, till the possibility to differentiate clusters of cells in tridimensional structure resembling the mouse gastrula, known as gastruloids (Martyn et al., 2018; Peng et al., 2019; Turner et al., 2017; Van Den Brink et al., 2014). To investigate the inter-relationship between the Primitive Streak program and the acquisition of neural fate, the interplay of the many neural inductive cues and the role played by retinoic acid in this context, I instead adopted as strategy the differentiation of mESC lines, maximizing the use of fluorescent reporters, with the advantage of studying such complex processes with fairly simple assays such as flow cytometry and confocal imaging. For this reason, I used throughout the entire project derivatives of the 2KI line previously generated in our group, harboring the GFP reporter under the control of the neural progenitor marker *Sox1*, and BFP in the locus of the Primitive Streak and mesoderm marker *Brachyury* (*T*).

4.1 Characterization of conditions mimicking the formation of the Primitive Streak at a transcriptional and signaling level

As much of the efforts of the last decades of stem cell research focused on defining conditions enabling the derivation of mature cell types of all the three germ layers, less explored is the intermediate stage of the Primitive Streak (PS) formation. Perhaps for the same reason, in the stem cell field, the T-box transcription factor Brachyury is more often regarded as a mesoderm marker, regardless of its role in the PS *in vivo*. To achieve conditions resembling the generation of the PS, as first part of my project, I opted for the analysis of a protocol requiring low serum and the small molecule IDE1. The use of IDE1 was originally proposed as a strategy even more effective than Activin treatment to generate definitive endoderm cell types (Borowiak et al., 2009). Studies published by our group indicate that the transcriptional profile of

the cells undergoing endoderm differentiation induced by IDE1 seems to transit through a stage resembling Epiblast stem cells (EpiSCs) (Sladitschek & Neveu, 2019). Moreover, the cell line harboring the BFP reporter downstream of the *Brachyury* promoter highlights the expression of the Primitive Streak marker during the differentiation towards endoderm, which is unusual (Fig3.1b). The findings pointed to the possibility of mimicking the steps going from naïve pluripotency to the induction of the Primitive Streak and the formation of endoderm cell types *in vitro*. However, the transcriptomic analysis of the cells undergoing IDE1 differentiation and expressing *Brachyury*_BFP, reveals a more complex scenario than what was previously reported. In addition to the expected definitive endoderm profile, it is possible to observe the signature of a Primitive Streak-like population, and expression of mesoderm genes, in particular of presomitic mesoderm (Fig3.1c). Noteworthy, this is an exclusive effect of the IDE1 treatment, as cells differentiated to endoderm with Activin A alone, or towards the EpiSC state, with Activin and FGF2, lack the expression of general and posterior Primitive Streak genes or mesoderm markers (Fig4.1). Therefore, the cells in IDE1 regime transit through an epiblast/PS-like state, but then proceed further with the formation of endoderm and mesoderm cell types.

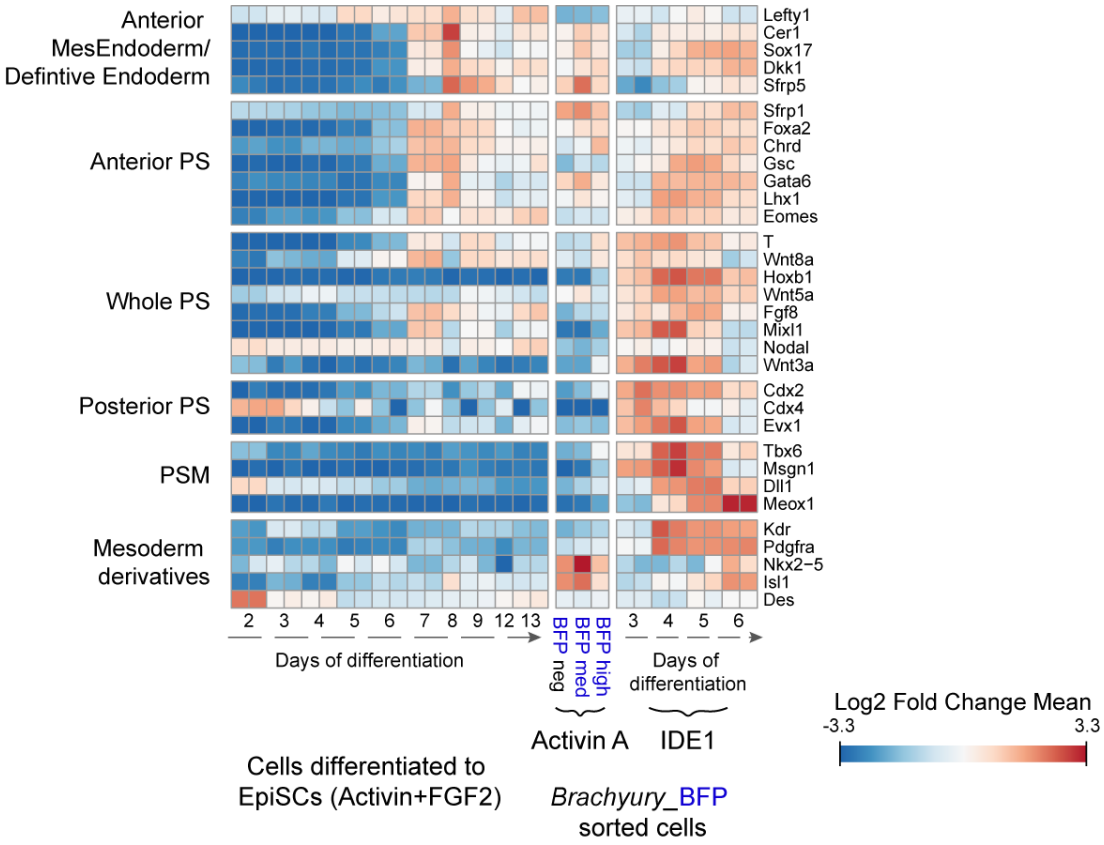


Figure 4.1: The unique properties of the IDE1 induced differentiation.

Heatmap comparing the Primitive Streak inducing properties of IDE1 and Activin A treatments. The conditions compared are from left to right: bulk sequencing of cells differentiated for 13 days with Activin A and FGF2 to reach an Epiblast Stem Cell state; high and low BFP$^+$ cells sorted after five days of Activin A differentiation and additional 24 hours in IDE1; BFP$^+$ cells sorted from three to six days of differentiation in standard IDE1 regime.

From studies conducted in the mouse embryo, it was concluded that both Wnt and Nodal signaling are necessary for the formation of the Primitive Streak. I tested whether the PS-like population originating in culture was subjected to the same signaling requirement, attempting as well to dissociate the individual contributions of the two pathways in this process. Although also in culture both signals can induce *Brachyury* expression, detected via BFP fluorescence, and start the Primitive Streak program when cross-talk is allowed, there is no Primitive Streak induction when Wnt is inhibited by XAV in the presence of IDE1 (Fig3.3a). On the other hand, when the Nodal/TGF-β pathway is inhibited with SB43, the Wnt agonist CHIR is still able to start *Brachyury* expression and a PS-like program, but characterized by the expression of posterior PS and mesoderm genes (Fig3.5b). Given also the loss of expression of

the anterior marker *Eomes* when Nodal/TGF-β signaling is inhibited (Fig3.5a), the results demonstrate that Nodal might be dispensable to begin the formation of the PS, but is essential for the generation of its anterior derivatives, as the definitive endoderm. This is in complete agreement with the *in vivo* data, showing that mild deficiency of the Nodal pathway impairs the formation of anterior PS cell types (Dunn et al., 2004; Lowe et al., 2001; Vincent et al., 2003). The inhibition of Nodal signaling has however a second unexpected consequence. In fact, the observed reduction of the anterior fate and disappearance of mCherry controlled by *Eomes* is associated with an increased specification to the neural lineage, evidenced by the greater fraction of *Sox1*_GFP$^+$ cells (Fig3.5a). It appears, consequently, that Nodal has the ability to promote anterior PS and endoderm fates and repress the neural one, whereas Wnt favors mesoderm and posterior PS fates and has no negative impact on the neural lineage *per se*. Considering that *Nodal* itself is reported as a positive target of Wnt signaling, I propose a mechanism whereby Wnt starts *Brachyury* expression, the Primitive Streak program and *Nodal* expression, and this in turn is responsible for directing a fraction of the PS cells to anterior destiny, while at the same time repressing the emergence of neuroectoderm cell types. A similar mechanism of interaction of Wnt and Nodal signaling to direct fate decisions at the streak stage has been demonstrated also for hESCs (Funa et al., 2015).

Regarding Wnt signaling and its propensity to induce posterior PS fates, I observed a similar behavior in an independent way, through the 2KI-Wnt cell line, reporting for Wnt/β-catenin signaling activation. Not only this line confirmed the activation of the canonical Wnt pathway in the IDE1 differentiation, before only hypothesized, but also granted the possibility to assess in an unbiased way the fate of Wnt$^+$ cells emerging from the PS-like population. These cells exhibited the transcription profile of posterior PS and mesoderm cell types, matching the role of Wnt signaling *in vivo* (Fig3.4c).

4.2 The neural inducing factors of the Primitive Streak-like population and the onset of neuroectoderm cell types

As shown in Fig3.6, in the conditions promoting the generation of the PS-like population, cells expressing the neural progenitor marker *Sox1* arise, both from Brachyury$^+$ and Brachyury$^-$ cells. I found that the cells expressing *Sox1* in the PS-like

condition belonged, indeed, to the neuroectoderm lineage by their transcriptional profile, and were able to maintain this state for prolonged time even in absence of serum and IDE1 (Fig3.7). The finding highlights an important aspect about the cells belonging to the PS, whose fate *in vivo* is restricted to endoderm and mesoderm lineages by the spatial organization of signaling cues. In fact, the Primitive Streak-like population should be rather considered pluripotent, given its possibility to generate cell types of all the three germ layers, endoderm, mesoderm and even ectoderm. These cells also exhibit an incredible plasticity, demonstrated by the experiment in Fig3.12, where high concentrations of RA can completely convert cells expressing *Brachyury*, in cells fated to the neural lineage and expressing *Sox1*. About the possible origin of the neuroectoderm cell types, it is intriguing to consider that, in the mouse gastrula, the Primitive Streak and its distal tip, the node, are organizers of the neural fate in the anterior epiblast, and they might possess the same organizing functions also in our culture system. As their *in vivo* counterpart, the cells of the PS-like population express genes coding for ligands and inhibitors of Wnt, Nodal and BMP pathways (Fig3.9a). The main difference is probably that, *in vivo*, a marked separation of the fates is achieved early on, with the streak forming endoderm and mesoderm in the posterior epiblast and inducing the formation of neuroectoderm anteriorly, by spatial segregation of the signaling domains, establishing gradients of signals along the main axes. The lack of an ordered spatial distribution in the bidimensional culture clearly reduces the possibility to investigate these processes in the exact way they occur during development. At the same time, it gives the chance to study the behavior of the streak, dissociating the events depending from the inner program and the ones dictated by the localized distribution of the signaling domains, enabling to test how these cells react to their own signals when senders and receivers are not spatially separated, and when the transition from pluripotency to lineage restriction is really achieved.

In these conditions of coexistence in the dish of uncoherent signals leading towards opposite directions, there seems to be a balance between the stimuli favoring and stabilizing the PS and others which instead promote the specification to the neural lineage (Fig4.2a). A similar competition between PS-like and neural-like populations, existing in a dynamic equilibrium, was hypothesized by Tsakiridis and colleagues in EpiSC cultures (Tsakiridis et al., 2014). The balance between the fates is precarious because any perturbation of one of the two affects the other, with the fractions of cells

types adjusting accordingly. This does not only apply to the exogenous addition of retinoic acid, which can induce the neural fate program from Brachyury^{+} cells, but also to the removal of signals sustaining the PS (Wnt and Nodal/TGF-β), or even the absence of inhibition of some of the PS activities, as in the case of Chrd & Nog KO line (Fig4.2b and c). In fact, inhibiting Wnt or TGF-β, either from pluripotent mESCs or after the PS program started, leads the cells towards the neuroectoderm lineage (Fig3.10a and Fig3.10b), whereas removing the BMP inhibitors reduces the fraction of cells directed to the neural fate and might stabilize the PS (Fig3.10c).

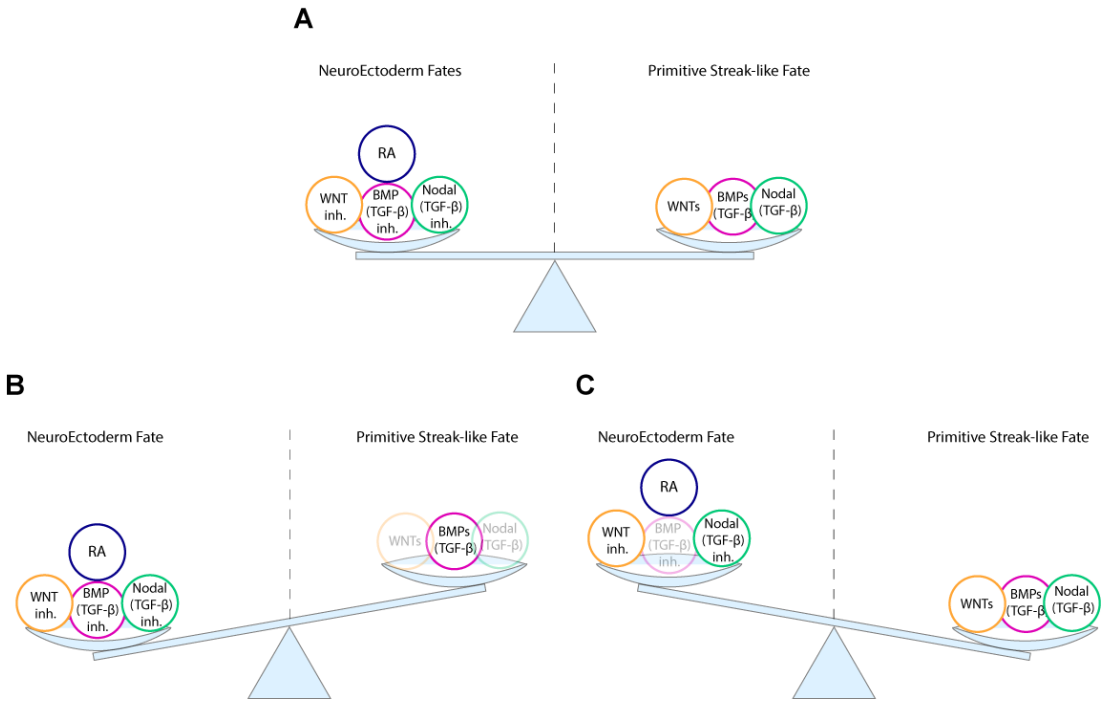


Figure 4.2: The balance between neuroectoderm and Primitive Streak inducing signals.

*Schematic depicting the proposed balance between signals inducing either Primitive Streak differentiation or specification to the neural fate: **A**) in the condition of equilibrium between the fates in standard IDE1 regime; **B**) when Wnt or Nodal/TGF-β are exogenously inhibited; **C**) when BMP inhibition is impaired as in the Chordin and Noggin double mutant line.*

The interesting outcome of this uneven signaling environment is the generation of multiple neuroectoderm types in culture. As shown in Fig3.8, the seemingly homogenous population of cells expressing *Sox1*, isolated by selecting the GFP^{+} cells from the PS-like differentiation, represents in reality a spectrum of neuroectoderm cell types with different developmental origins along the anteroposterior axis. I managed to untangle some of the signals contributing to the formation of these various neural

fates, identifying the inhibition of the Wnt pathway as a mechanism inducing anterior neural fates, the retinoic acid signaling as inducer of hindbrain cell types, whereas cells with spinal cord character emerged as a result of combined inhibition of Nodal and activation of Wnt. An intriguing question still to solve is whether these multiple neuroectoderm types arise all simultaneously, each induced by different cues, or as proposed by the activation-transformation hypothesis, the anterior ectoderm is the first to arise and all the other stimuli act on it driving the formation of more posterior cell types. At least from the transcriptional profile obtained from the $Sox1^+$ cells over time during the PS-like differentiation (Fig3.8) there seems to be no strong indication for an initial anterior induction compared to the others. Regardless of the mechanism of induction, both flow cytometry and time lapse imaging (Fig3.6), evidence that at least a part of the neuroectoderm cell types originate from cells expressing *Brachyury*. This observation is of extreme relevance, because during mouse embryogenesis the co-expression of the PS/prospective mesoderm marker *Brachyury* and neuroectoderm genes, such as *Sox2*, has been reported at E8.5 in the Node-Streak Border (NSB), and the Caudal Lateral Epiblast (CLE) adjacent to the PS. The group of cells derived from these two regions, able to give rise to both mesoderm and neuroectoderm cells of trunk and tail, is collectively known as neuromesodermal progenitors (NMPs), and is responsible for the posterior expansion of the body axis from somitogenesis till the tail bud stage (Henrique et al., 2015). The co-expression of *Brachyury* and *Sox1*, and the presence of both presomitic mesoderm (PSM) and neural progenitor (NPs) cells in the PS-like differentiation, together with the natural evolution during gastrulation of PS regions in neuromesodermal progenitors, suggest the presence of these bipotent progenitors in our culture system. To test this, I isolated the cells expressing both BFP and GFP, only one of the two reporters, or neither of the two after five days of differentiation in the IDE1 regime and analyzed their transcriptional profiles. Strikingly, there is a clear distinction between the $Sox1^+Brachyury^+$ cells, expressing genes of both mesoderm and neuroectoderm cell types and NMP markers as *Nkx1-2*, and the cells expressing only *Brachyury* or *Sox1*, which belonged respectively to the presomitic mesoderm and neural progenitor fates (Fig4.3).

Although preliminary, this result anticipates the possibility that due to the temporally and spatially disordered formation of cell types in the differentiation, a group of PS-like cells might convert in NMPs. Nonetheless, a distinction between

NMPs and the rest of the PS-like cells exists, based on the expression in our culture of node/organizer genes as *Chordin* and *Noggin*, and anterior PS markers as *FoxA2* which are absent in the NMPs. Moreover, the presence of neuroectoderm types spanning the entire anteroposterior axis rather than only posterior ones, and the clear reduction of $Sox1^{+}$ cells in the Chordin and Noggin double mutant line, highlights that the conversion of the NMPs into neural progenitors might contribute to the pool of cells directed to the neural fate, but remains one of the other discussed mechanisms active in the IDE1 regime.

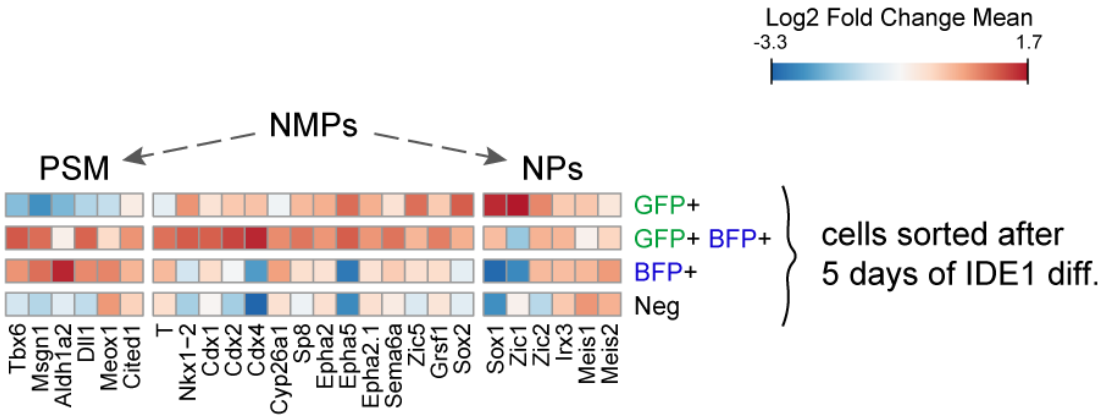


Figure 4.3: Formation of NMPs in the Primitive Streak-like differentiation.

Heatmap showing differential expression of neuromesodermal progenitor (NMPs), presomitic mesoderm (PSM) and neural progenitor (NPs) markers in cells differentiated for five days in standard IDE1 culture and sorted according to the expression of the indicated reporters.

4.3 Retinoic acid and neural induction in the PS-like differentiation

The functions of retinoic acid (RA) signaling from the primitive streak to the headfold stage of gastrulation have rarely been investigated, and its production in the nascent paraxial mesoderm, at the late streak stage, is classically considered the incipit of its activity. However, Ribes and colleagues showed that the expression of the synthesizing enzyme Aldh1a2, as well as RA activity detected by the RARE-*lacZ* mouse line, are already present in the Primitive Streak and in the node at E7.5, with RA activity spreading from the mesoderm also to the adjacent epiblast (Fig4.4). At E7.75, the headfold stage, the expression of the synthesizing enzyme ceases in the Primitive Streak, while the RA signaling progressively decreases in this structure (Ribes et al., 2009). The RA degrading enzyme, Cyp26a1, is instead expressed from around E6.5–E7.0 in the extra-embryonic and embryonic endoderm, and then in

primitive streak and posterior mesoderm. At E7.5, *Cyp26a1* transcripts are transiently detected in the anterior region, where the enzyme is considered to protect the forming neuroectoderm, confining the zone of active RA signaling to the posterior region (Fig4.4) (Fujii et al., 1997; Ribes et al., 2007). In the context described in this report, the quite exclusive upregulation of *Aldh1a2* and *Cyp26a1* in the Brachyury^{+} cells (Fig3.18b and Fig3.19a), as well as the presence of RA signaling activity detected in the PS-like population through the RA reporter (Fig3.18c), strongly reflect the proposed presence of RA in the Primitive Streak during gastrulation. It has been suggested that the RA signaling, at this stage, might limit the ingression of the cells through the streak, restricting the expansion of the mesodermal domain.

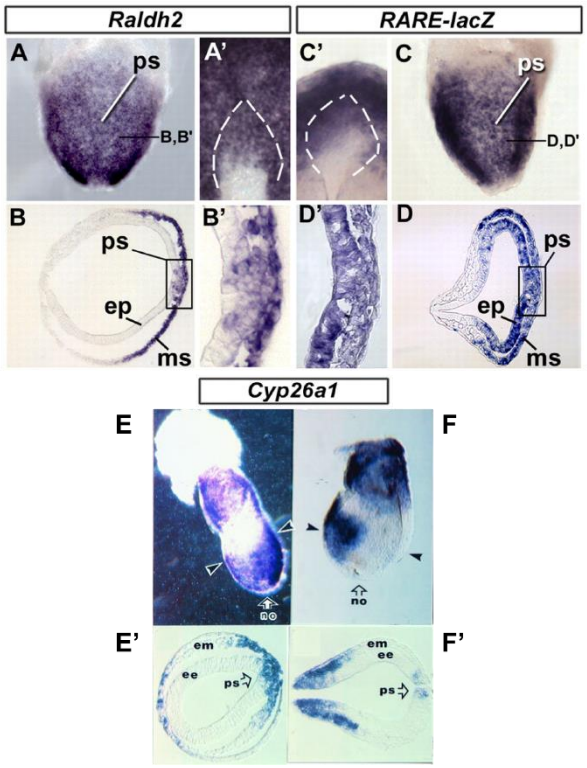


Figure 4.4: Aldh1a2 and Cyp26a1 expression and RA signaling in the Primitive Streak of mouse embryo.

First row: whole-mount detection at E7.5 of Aldh1a2 (Raldh2) transcripts in the primitive streak (A,B,B') and node (A') on the left panel, and lacZ mRNA (RA signaling) in the primitive streak (C,D,D'), node(C') and epiblast layer (D) on the right panel. Adapted from Ribes et al. 2009. Bottom row: whole mount in situ hybridization of Cyp26a1 transcripts at E7.25 in the primitive streak region (E, E'), and at E7.5 in the anterior ectoderm (F, F'). Abbreviations: ep, epiblast; ms, mesoderm; ps, primitive streak; ee, embryonic ectoderm; em, embryonic mesoderm; no, node. Adapted from Fujii et al., 1997.

Regarding the RA functions associated to neuroectoderm specification, the most important observations came from the analysis of *Aldh1a2* mutant mice. These mice display in fact ostensible absence of RA signaling as indicated by the RARE_*lacZ* reporter and, despite presenting several defects in the hindbrain, somites and limbs, they have normal forebrain structures and globally normal expression of the neuroectoderm markers *Sox1* and *Sox2*. This finding led to the conclusion that RA is not required for neural induction but only for posterior neural differentiation as suggested by the absence of Pax6 and Olig2 transcripts in the posterior neural tube of *Aldh1a2*$^{-/-}$ mice (Molotkova et al., 2005). The discussed presence of RA signaling in the condition promoting the formation of the PS-like population, combined with its well-accepted ability to induce neural fate *in vitro*, and the possibility to greatly affect the fraction of Sox1^{+} cells originating from the PS by exogenously activating or inhibiting this pathway (Fig3.12), encouraged to question this paradigm of the RA function in the neural specification. Moreover, the finding that inhibition of RARs or stimulation of RA production by vitamin A perturbs the equilibrium between PS and neuroectoderm fates without completely altering the course of the differentiation (Fig3.18d), proposes RA signaling as one of the factors orchestrating the balance of cell types anticipated before (Fig4.5). Nevertheless, the demonstration of an influence of RA on the normal course of the PS-like differentiation comes from the observation of the Cyp26a1 knock-out line. When cells devoid of this RA degrading enzyme are differentiated according to the IDE1 protocol, a considerably higher fraction of cells acquires the neural fate, meaning that RA degradation is a mechanism limiting the formation of neural cell types by the PS-like population (Fig3.19c). The possibility to increase even more the formation of neural progenitors, by providing vitamin A, highlights that together with RA degradation, another limiting factor in the fate decision is the RA synthesis. Thus, producing more RA or removing Cyp26a1 clearly set the balance between the fates in culture on a new equilibrium point favoring the specification towards neuroectoderm (Fig4.5).

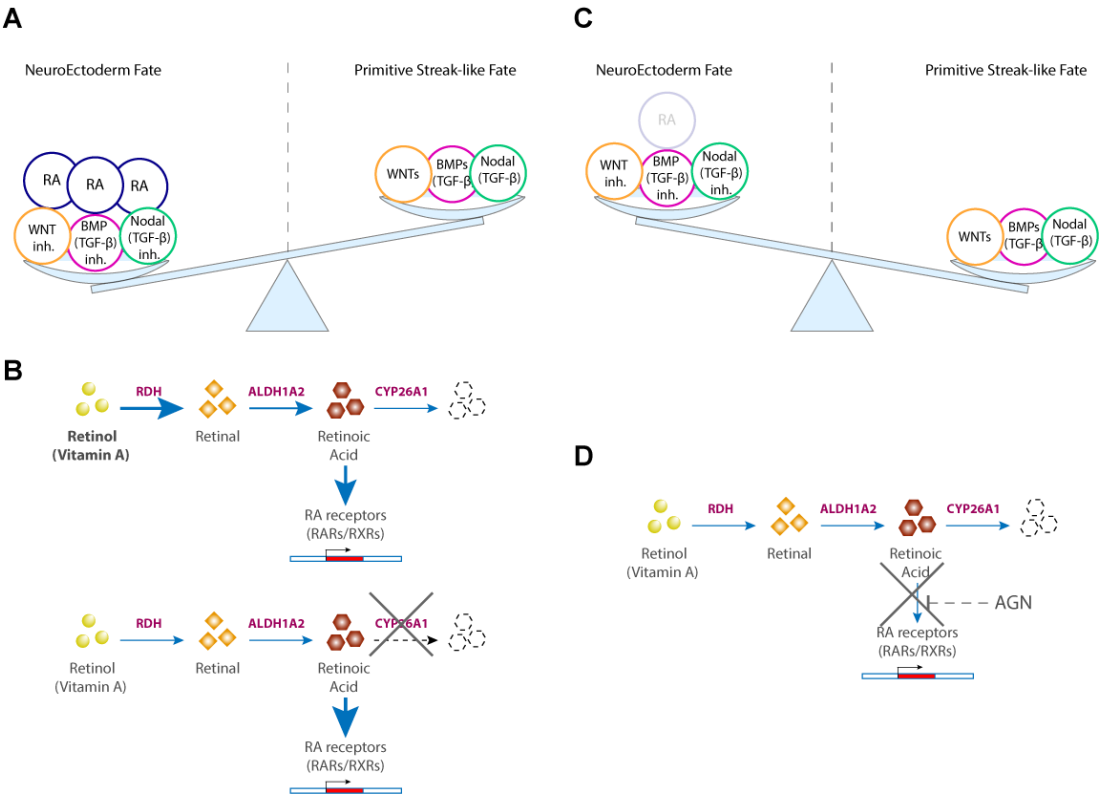


Figure 4.5: The balance between neuroectoderm and Primitive Streak-like fates is altered by RA signaling.

__A__) and __B__) Representation of the shift in the balance between neuroectoderm and primitive streak-like cell types, in favor of the formers when RA signaling is boosted by vitamin A supply or removing Cyp26a1 in the Cyp26a1 KO line. __C__) and __D__) Perturbation of the balance in the direction of reduced neuroectoderm fraction when RA signaling is blocked by AGN.

It is important to mention that RA is not alone in this context, and other signals contribute to the induction of a neural fate from the PS-like population. The simultaneous perturbation of RA signaling, by Cyp26a1 KO or AGN treatment, and inhibition of Wnt or Nodal/TGF-β pathways provided valuable insights into the relationship between these various means of neural induction, highlighting that they are not independent from each other (Fig4.6). More in detail, the combination of Cyp26a1 KO and inhibition of Wnt or TGF-β, results into an increase of Sox1+ cells greater than the individual conditions, although this does not exclude independent mechanisms of action (Fig3.21). On the contrary, combining AGN with XAV or SB43 clearly shows that the inhibition of RA signaling prevents the induction of neural fate by the other mechanisms (Fig4.6). However, given the reported action of AGN as an inverse agonist, it is not possible to discriminate whether its effect on the other neural

induction derive from an enhanced repression of RARs or absence of RA signaling. The former case would mean that some of the genes responsible for the acquisition of neural fate by either XAV or SB43 treatment are also controlled by the RARs, and increasing the repressive action of the receptors by AGN, hinders these mechanisms. On the other hand, the second possibility implies that the effectiveness of XAV or SB43 requires active RA signaling. While the inhibition of Wnt or TGF-β pathways might destabilize the PS-like population (as indicated by *Brachyury*_BFP downregulation in the presence of XAV), RA signaling could work independently, inducing the differentiation of the destabilized cells. To make this distinction it is necessary to achieve a condition devoid of RA signaling, for instance, by removing RA synthesizing enzymes, and preliminary data in this direction suggest that active RA signaling is indeed necessary to induce neuroectoderm at least by Wnt inhibition.

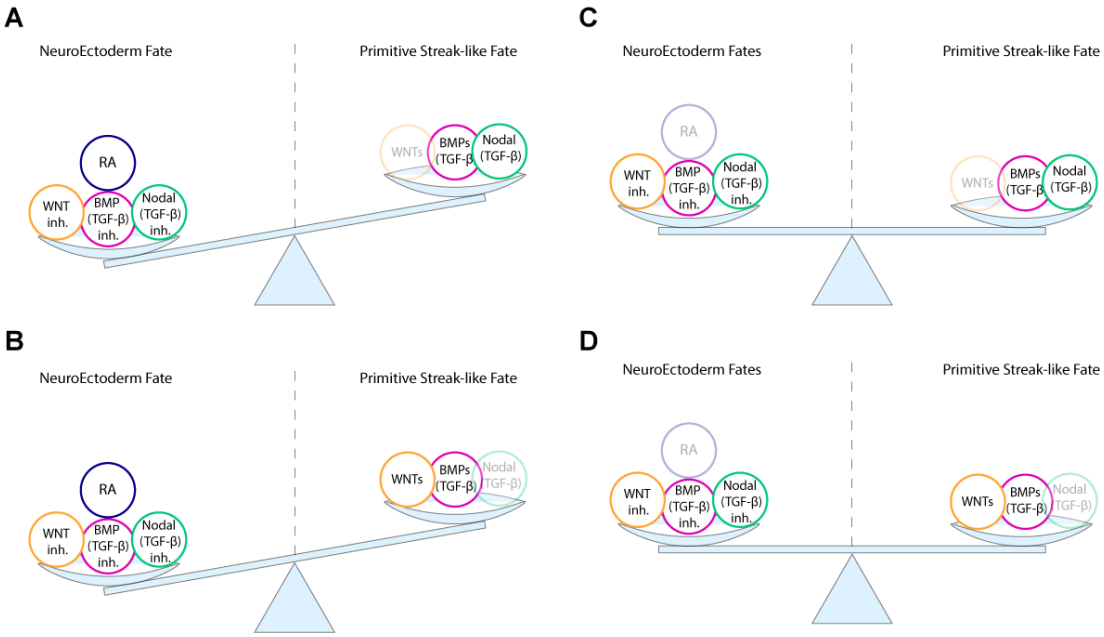


Figure 4.6: Blocking RA signaling re-establishes the balance between neuroectoderm and Primitive Streak-like fates altered by Wnt or TGF-β inhibition.

__A__) and __B__) representation of the altered balance between neuroectoderm and primitive streak-like fates, favoring the former when Wnt (A) or Nodal/TGF-β pathway (B) are inhibited by respectively XAV and SB43. __C__) and __D__) the repression of RA signaling can revert the effect of Wnt (C) or Nodal/TGF-β (D) inhibition, re-establishing a new equilibrium between neural lineage and primitive streak-like cell types.

In the first attempts to accomplish a condition devoid of RA during the PS-like differentiation, I introduced null mutations in the *Aldh1a2* gene in my RA sensor line,

mindful of the *Aldh1a2*$^{-/-}$ mouse phenotype characterized by absence of RA signaling. Accordingly, removing Aldh1a2 leads to the abrogation of the RA reporter expression in the IDE1 regime, without clear reduction in the fraction of cells expressing the *Sox1*. Associated to increased expression of *Brachyury*_BFP reporter (Fig3.22b and Fig3.23b), the results recapitulate the main observations made in the *Aldh1a2*$^{-/-}$ mice (Molotkova et al., 2005; Ribes et al., 2009), providing a solid ground to study this phenotype more in detail. The seemingly unaffected *Sox1* expression in the KO is in contrast with the results here presented, showing that inhibition of RA signaling leads to reduced specification to the neural lineage, whereas in this case the balance between the fates seem not to be perturbed. One possible explanation is that the ostensible unchanged level of GFP reporter is misleading, as a reduction of neural cell types deriving from RA signaling, might be concealed by a compensatory increase of other neuroectoderm subtypes also marker by Sox1. To explore this possibility, the acquisition of the transcriptional profile of the Sox1^{+} cells in the Aldh1a2 KO will be essential. A hint derives from the finding that the Aldh1a2 KO cells present, together with globally higher levels of *Brachyury*_BFP expression, also an increased proportion of cells co-expressing *Brachyury* and *Sox1*, thus putative NMPs, which could indicate a shift between the induced neuroectoderm populations (Fig4.7). A role for RA signaling has indeed been demonstrated in the context of the decision of the neuromesoderm progenitors to differentiate to cells committed to the neural lineage and, thus, the absence of Aldh1a2 synthesized RA might lead to defective specification of the NMPs in culture (Gouti et al., 2017).

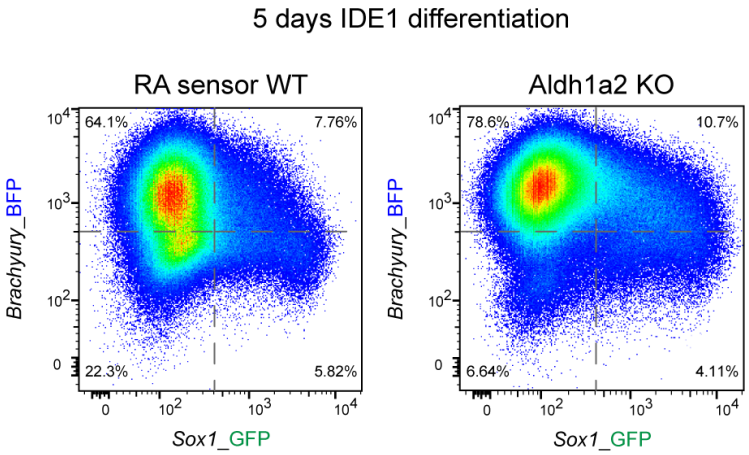


Figure 4.7: Comparison of the proportion between populations in the Aldh1a2 KO and in the WT RA sensor line.

Pseudocolor density plots showing the different proportions of Primitive Streak-like (BFP) and neuroectoderm (GFP) populations in the Alh1a2 KO line and in the WT RA sensor line. After five days of standard IDE1 differentiation the Aldh1a2 KO line displays an increase of BFP^+GFP^+ cells, putative neuromesoderm progenitors. The percentages of the populations derive from the separation into quadrants represented by the dotted lines.

Another hypothesis which I am exploring is that RA production in the Aldh1a2 KO is not completely ceased, and some signaling might be still active, despite the absence of the RA reporter. The results indicating further reduction of the $Sox1^+$ cells upon addition of AGN to the Aldh1a2 KO line IDE differentiation, or the reactivation of the RA reporter when vitamin A is provided, demonstrate that the pathway is still active (Fig3.22c). Therefore, other ALDHs enzymes might take over the production of RA, even though it should be clarified whether the other putative enzymes are operating also in normal conditions, or they are activated by a compensatory mechanism in the Aldh1a2 KO cells. Further implications are that RA can exert its actions on the neuroectoderm fate at such low concentrations to not even trigger the expression of its own reporter, or alternatively that the DR5_RARE RA reporter, used also in the classical mouse experiments, does not account for all RA signaling. The plausibility of the first option is demonstrated by the fact that adding sub-nanomolar (0.01 and 0.1 nM) concentrations of RA to the PS-like differentiation of the Aldh1a2 KO line, allows to slightly shift the balance of the fates towards neural lineage, without inducing Scarlet expression (Fig3.22d). As for the second hypothesis, the existence of a more widespread RA response seems to be supported by reporters controlled by non-

conventional RAREs, as the ones found in the *Mafa* gene. The novel RA sensor generated with the *Mafa cis*-regulatory sequences is responsive to RA signaling and inhibition of RARs by AGN, and reveals RA activation in a greater fraction of cells than the classical DR5 reporter during the PS-like differentiation, even in conditions such as the XAV treatment, where the pathway seems normally almost not active (Fig3.27). This preliminary result underpins the possibility that the RA reporter used in the classical studies, marks only a fraction of the cells in which RA signaling is active. Whether the RA responsive cells detected by the *Mafa* RAREs constitute a completely different population, or a more sensitive version of the DR5 Scarlet^{+} cells, will be clearer only by analyzing their transcriptomes.

4.4 Retinoic acid control over the neuroectoderm specification

The default model of neural induction, applied to ESCs differentiation, argues that the neuroectoderm fate arises in the absence of inhibiting substances contained in the serum, as BMPs, Wnts and Nodal/TGF-β ligands. The first reports showing efficient neural conversion of ESCs using only the N2B27 synthetic medium seemed to confirm this model. However, as published already by Engberg and colleagues, the conclusion made by these first studies are undermined by the use of N2B27 formulations containing precursors of RA or by maintaining the cells in the presence of serum which contains traces of RA precursors (Engberg et al., 2010). The results discussed so far confirm the role of retinoic acid among the neural induction signals during the formation of the PS-like population. Even in conditions devoid of serum and IDE1, comparing the differentiation efficiency of mESCs in N2B27 without RA precursors, N2B27 containing substrates for RA synthesis (N2B27 vit.A), or N2B27 with addition of high concentrations of RA, evidence the higher efficiency of RA in driving the neural fate. However, by simply looking at the effects of a continuous addition of RA, the actual functions played by this pathway during the specification of the neuroectoderm lineage were still hard to grasp. It is important to mention that the acquisition of a neural fate is a stepwise process, in which different groups of transcription factors follow each other determining waves of genes regulation. As a general mechanism, a transcription factor at a given stage negatively controls both the previous set of factors and the following ones, ensuring a certain stability to that stage till new conditions cause its downregulation and trigger downstream events. This is

the case of the transcription factor Sox1, which marks the beginning of the commitment towards the neural lineage. During the RA-induced differentiation, *Sox1* is upregulated and rapidly downregulated within four or five days of culture, and given this transient trend of expression, the GFP reporter controlled by its promoter never reaches high fluorescence levels. Intriguingly, I found that when RA is provided for only 24 hours during the N2B27 differentiation and then removed, the analysis of the GFP fluorescence at multiple time points indicates activation of *Sox1* transcription in most of the cells, at unprecedented high levels and in a stable fashion. The implications of these observations are that RA can commence a stable neuroectoderm program within 24 hours, but also that its continuous supply is mainly required to render the Sox1 state as transient as possible. In other terms, RA signaling must have the dual capability of inducing or accelerating the acquisition of the Sox1 state and of controlling its conclusion. On a practical level, the confirmation that the cells expressing high levels of GFP possess the transcription profile of neuroectoderm, and the stability of this expression, imply that if this mechanism were in place in hESCs as well, it could be a very effective and simple system to obtain a homogenous population of proliferating neural progenitors, allowing their expansion and controlled differentiation. Interestingly, the global response to the RA pulse is not altered in the Cyp26a1 KO line, where RA degradation is defective. This excludes the possibility that the degradation of RA stabilizes the Sox1 state. Nevertheless, the deletion of Cyp26a1 reduces the minimum concentration of RA able to commence stable *Sox1* expression (Fig3.20c).

Intrigued by the fact that RA seems to have more functions than simply inducing the neuroectoderm fate, I generated the RA sensor line to follow the evolution of the RA response during differentiation in combination with the specification to the neural lineage. From the time-lapse imaging in Fig3.17, it is clear that at concentration larger than 100nM, RA acts according to a cell autonomous mechanism, with the detection of the RA reporter only few hours after RA addition, preceding the expression of *Sox1* in most of the cells. However, the analysis of the progression of the RA response by flow cytometry brings to the attention two surprising aspects: the (anti-)correlation of RA activation and *Sox1* expression over time, and the dynamic of the RA signaling. Regarding the first, the co-expression of the RA-responsive Scarlet reporter and *Sox1*-driven GFP in the same cells is much rarer than what was expected and is observed

mainly within 48 hours from RA addition, in cells exhibiting intermediate levels of both reporters (Fig3.16). Concerning the RA signaling activation, it appears that at high enough concentrations most of the cells respond in less than 24 hours, expressing high levels of Scarlet, but later the fluorescence decreases without additional waves of activation, even though RA is continuously present in the media. The more the GFP^+ cells induced over time, the fewer $Scarlet^+$ cells can be detected, suggesting, in my opinion, that the cells showing RA activation at late time points are cells delayed in their response to the differentiation stimuli. Taking into account also that the extent of Scarlet and GFP co-expression seems to be dependent on the doses of RA provided, with higher concentrations allowing for the least co-expression, the dynamics of *Sox1* expression seems to strongly influence the window of RA responsiveness. Therefore, I hypothesize a mechanism whereby Sox1 activation can downregulate the response to RA, or vice versa the decrease of the RA response allows *Sox1* expression. This idea is corroborated by the fact that cells which present higher Scarlet levels seem to express *Sox1* later than the ones downregulating the reporter more rapidly. Overall, the order of events following RA addition are first high intensity of Scarlet expression in most cells, followed by a variable range of co-expression with GFP while Scarlet fluorescence diminishes, and eventually terminating in only GFP being detected, without new expression of Scarlet in the same cells. The nature of this mechanism is elusive and intriguing at the same time, because the behavior of the cells exposed to the RA pulse strongly suggests that at least two separated responses to RA are necessary for a normal differentiation course, implying that the cells must be able to respond to RA even after activating Sox1. This opens multiple scenarios, one being that if the initial activation of RA signaling is high/prolonged enough the Sox1 state is intrinsically destabilized, perhaps by the expression of additional transcription factors which are not induced by the RA pulse. Another possibility is that the "second" response downregulating *Sox1* is not detectable by the upregulation of the Scarlet reporter.

The degradation of RA seemed also a very good candidate for the mechanism by which $Sox1^+$ cells enforce the conclusion of the RA response, also because the degrading enzyme Cyp26a1 is a direct target of the pathway, and it is known to establish a negative feedback loop on the RA response. However, if that were the case, the knock-out of Cyp26a1 would prolong the RA response even after the cells express

Sox1. I found that removing Cyp26a1 renders the RA response more penetrant, with more cells expressing the RA reporter and at higher level when first exposed to a given RA concentration, but it did not extend the activation of RA signaling (Fig3.20a). On the contrary, it appears that after the first more efficient response to RA in the Cyp26a1 KO, the reporter levels decrease much more rapidly, and a reduced fraction of cells shows Scarlet fluorescence at the last time points analyzed. I interpret the results proposing that the removal of the degrading enzyme facilitates the first response to RA both in terms of its maximum intensity in a given cell and of number of cells responding, but it does not affect the possibility to tune down RA signaling once the neuroectoderm program begins.

A further possible explanation for the observed trend of the RA signaling activation is that the pathway is attenuated during neuroectoderm specification owing to the degradation of RA receptors (RARs). RA-induced proteasomal degradation of its receptors has been proposed as a negative feedback mechanism at least for RARα and RARγ (Giannì et al., 2002; Zhu et al., 1999). The degradation of the receptors responsible for Scarlet expression would explain the progressive reduction of the reporter levels, although a different combination of receptors, which does no trigger Scarlet expression, would be required to allow *Sox1* downregulation. To explore this possibility, I attempted the mutations of single RARs, but owing to the extensively reported redundancy of these receptors, no clear phenotype was identified, rendering the investigation in this direction particularly challenging. Even though the combined mutation of the three RARs is not helpful to study the dynamics of the RA response, it demonstrates that RA actions as the downregulation of *Brachyury* and the expression of the signaling reporter depend on the presence of functional RARs. The expression of *Sox1* in the triple RAR KO appears independent of RA signaling and does not follow the normal trend. In other words, observing the triple RAR KO it seems that the activation of the $Sox1^+$ state can be achieved by other means, but the subsequent differentiation of the progenitors is impaired. More thorough analyses, with time courses and pulses of RA, together with the RNA sequencing of the $Sox1^+$ population emerging in the KO, will reveal more about the dynamics of neural specification in this line, perhaps also improving our understanding about the role of RAR-mediated gene repression in this process.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

Despite the tremendous advancements in our understanding of the events occurring during gastrulation and the onset of the germ layers, proved by the ability of mimicking more and more complex and advanced stages of development, a mechanistic comprehension of the processes requiring the interplay between multiple signals and cell types, as the formation of the neural lineage is not fully reached. The intent of this project was, therefore, to use mESCs as a model system to bring together some of the pieces going from the Primitive Streak formation to the acquisition of neural fate, often investigated as independent phenomena, and moving from there to revisit some paradigms about neuroectoderm formation, as the role of retinoic acid. In order to do so, the knowledge underlying the formulation of many lineage specification protocols was at the same time the starting point and the focus of the analyses of the project. In this report, I presented the individual contributions of Wnt and TGF-β pathways in the Primitive Streak formation, and the effects of their inhibition on the specification to the neural lineage, highlighting the interdependence of these two fate decisions both from pluripotent mESCs and the epiblast/Primitive Streak. The characterization of a culture system mimicking the formation of the Primitive Streak and the onset of the three germ layers, moreover, enabled to investigate their capability to regulate each other in absence of a precise spatial distribution. As a result of the complex signaling environment present, multiple neuroectoderm cell types arise, and I proposed the putative stimuli underlying their generation. Furthermore, a significant contribution of this work, is the re-evaluation of the role assigned to retinoic acid signaling in neuroectoderm specification, owing to the possibility of studying its functions in relation to the other neural inductive cues, and in a scenario closely resembling physiological conditions. Much of the information we have about the RA functions, in fact, comes from studies performed *in vitro*, often using overexpression approaches and exogenous supply of RA, likely concealing subtle effects of the endogenous system. The generation of a cell line ad hoc for monitoring the RA response, combined with pharmacological and genetic perturbation of the pathway argues in favor of a clear contribution of RA in the neural fate decision in our context. Remarkably, the data also support an unprecedented involvement of RA even in the mechanisms of induction by Wnt or TGF-β inhibition. Moreover, the simultaneous

analysis of the response to RA and of the acquisition of the neural fate gave new insights into the complex mechanisms of action of the retinoid signaling. Finally, the observation made in the Aldh1a2 KO line, showing that, in this genetic background, RA could still be produced and its signaling still present, allows to question the main argument against the role of RA in the neural specification: the presence of seemingly normal neural induction in the presumed absence of RA signaling in the *Aldh1a2*$^{-/-}$ mice.

Overall, this work and the molecular tools generated represent a basis for further investigation about the neuroectoderm lineage specification. In the near future, the achievement of conditions completely devoid of RA synthesis will be crucial to define more correctly the functions of this pathway. The transcriptome analysis of Aldh1a2, Cyp26a1 and triple RAR KOs will allow the comparison of the neural cell types emerging in these various conditions. The use of these and other lines, like the Chordin and Noggin double mutant line, in experiments similar to the combination of the Cyp26a1 KO and Wnt or TGF-β inhibition, have also a great potential for the analysis of the interplay between the major neural inductive cues. Ultimately, the transcription profile of the cells marked by the novel RA reporter will disclose the nature of the additional cells responding to RA, with the attractive possibility to also generate a reporter mouse line.

6 MATERIALS AND METHODS

6.1 Mouse embryonic stem cell culture

6.1.1 General cell culture practice

Every cell culture experiment carried out in this project was performed in a laminar flow hood according to the following general practice. The sterility of the hood was ensured by spraying with 70% Ethanol the working surface prior and after any experiment and by periodical germicidal UV cycles. Any material entering the flow hood was sprayed with 70 % Ethanol and wiped clean. Sterile cell culture plasticware as well as any material containing reagents which would come in contact with cells were handled exclusively under the laminar air flow. During cell culture operations, bottles, tubes, and plates were kept closed whenever possible. Glassware and any in-house liquid material as D-PBS and H_2O were autoclaved. Gelatin and all the stem cells media solutions were sterile filtered after mixing the required components, using a 0.22 μm Stericup Filter Unit (Millipore, SCGPU05RE), aliquoted in 500 mL bottles and stored at 4 °C for no longer than 6 weeks. Before use, the media were warmed to 37°C in a water bath. Bioactive molecules particularly sensitive to repeated warming and cooling cycles were added only prior use to the solutions aliquoted in 50mL or 15mL falcon tubes, and in the case of light-sensitive material tubes were protected with aluminum foil. The experiments involving this last kind of substances were carried out in dark conditions. For all the different kind of experiments included in this work, cells were maintained in humidified incubators, providing a stable environment consisting of a temperature of 37 °C and 5% CO_2. The tray containing the water granting humidity was disinfected with 70% Ethanol before every refill.

6.1.2 Preparation of cell culture plates with coating agents

When not stated otherwise, the cell culture was carried using plates (Nunc™) coated with 0.1% (w/v) gelatin from porcine skin (SIGMA, G1890). The coating solution was prepared according to the following procedure: 0.8g of gelatin were weighed in a sterile beaker intended for cell culture use only, and 500mL of autoclaved water were added. Upon swirling, the solution was warmed till reaching its boiling point, necessary to ensure the dissolution of the gelatin in water. The volume was then

increased till 800mL by adding room temperature (RT) water , and when the solution reached a temperature of approximately 50 °C, it was sterile filtered in the laminar air hood with a 0.22 µm Stericup Filter Unit (Millipore; SCGPU05RE) and later stored at 4 °C. To prepare the plates for cell culture, a volume of 0.1% (w/v) gelatin sufficient to completely cover their surface was added at latest 15 minutes before the use. Dishes were stored in the humidified incubator at 37 °C, and the gelatin solution was completely aspirated just before adding medium to the plate.

For the experiments concerning the re-seeding of the neural progenitors, Laminin_511 (Sigma, L2020) was used as coating agent instead of gelatin. The coating solution, kept at -20 °C and stabilized at +4 °C prior use, was prepared by diluting the stock to a final concentration of 5µg/mL with D-PBS. Coating the plates in this case required one hour of incubation and one wash with D-PBS. The solution containing the laminin dilution can be stored at +4 °C for a few weeks.

6.1.3 Maintaining and propagating mESC

The medium allowing maintenance and propagation of mESCs, abbreviated as ES complete, consists of the following components:

Component	Comment	Supplier (Number)	Final concentration
DMEM	high glucose, w/o glutamine	Invitrogen (11960)	80 % (v/v)
Fetal Bovine Serum	ES-qualified	Millipore (ES-009-B)	15 % (v/v)
100x Non-Essential Amino Acids	10 mM	Invitrogen (11140)	1x (100 µM)
100x L-Glutamine	200 mM	Invitrogen (25030)	1x (2 mM)
100x Sodium Pyruvate	100 mM	Invitrogen (11360)	1x (100 µM)
100x Penicillin-Streptomycin	10,000U/ mL 10,000U/ mL	Invitrogen (15140)	1x (100 U/ mL each)
2-Mercaptoethanol	55 mM in D-PBS	Invitrogen (21985)	100 µM
Murine LIF	Recombinant	EMBL PEPCore	10 ng/mL

To ensure stable nutrient conditions, ES complete medium was changed every day, and the cells were passaged every other day, two to three hours after refreshing the media, according to the following procedure. The medium was removed, and the

dish was rinsed once with D-PBS at RT. A solution of 0.05 % Trypsin-EDTA (Invitrogen, 25300) was added in a dropwise manner (~ 25-30 µL/cm^2), the plate was tilted to allow the solution to cover the entire surface, and then incubated for 4-5 min at 37 °C. To inactivate trypsin digestion, at least four volumes of medium containing 10-15% serum were added to one volume of trypsin. A single cell suspension was generated by pipetting up and down approximately 10-12 times, while flushing the surface of the well, and a fraction of the resuspension was transferred into fresh medium in a gelatin coated dish, with a dilution ratio decided as such that the culture would reach a 70 to 90% confluency within two days. Although the cells were typically not counted during routine maintenance, this procedure allows the seeding of approximatively 30 to 40000 cells/cm^2.

6.1.4 Counting ES cells using the Neubauer chamber

In all the cases that required cell counting, and for all the differentiation protocols, 50µL of the cell suspension after trypsinization were mixed with 50µL of 0,4 % (w/v) Trypan blue (Invitrogen, 15250), and 10µL of the mix were transferred to a Neubauer improved chamber, mounted according to manufacturer recommendations. The chamber was then positioned on an inverted DM IL LED microscope from Leica with an HI PLAN 10x/0.22 PH1 objective. Only transparent/live cells were counted in all the four quadrants, each composed of 16 single squares. The obtained number was then divided by 4 to have the average cell number per quadrant and the result multiplied by 2 (to account for the 1:2 dilution with Trypan blue), and by 10000, factor deriving from the volume of one quadrant (1mm x1mm x 0.1mm = 0.1mm^3 = 10^{-4} mL). The general equation to calculate the number of cells in each mL of suspension is:

$$\text{Concentration } \left[\frac{cells}{mL}\right] = \frac{\sum Number\ of\ cells \times 10{,}000\ \times \text{dilution factor}}{\text{Number of quadrants}}$$

6.1.5 Freezing mouse embryonic stem cells

Before freezing the cells, the culture was grown till a confluency of at least 80% and the medium refreshed two to three hours before the starting the procedure. The same steps of passaging the cells apply also for freezing, but when the single cell suspension was obtained, in this case, the cells were spun down (5 minutes, 500 rpm

at RT) in 1.5mL microtubes or 15mL falcon according to the volume. The supernatant was then removed, and the pellet resuspended with a volume of freezing solution (10 % DMSO (v/v) (SIGMA, D8418) in ES-qualified EmbyroMax Fetal Calf Serum (Millipore, ES-009-B)) dependent on the amount of cells/size of the pellet to freeze, typically 1mL for a well of a 6-well dish. The suspension was then moved to a 2mL cryo-vial and placed in a polystyrene sandwich at -80 °C, which allows for gradual decreasing of the temperature. The day after the vials were transferred in liquid nitrogen for longer term storage.

6.1.6 Thawing mouse embryonic stem cells

As opposite to the freezing process, which needs to occur gradually to avoid formation of crystals, thawing must be as fast as possible considering the toxicity of high concentrations of DMSO at room temperature. For this reason, all the necessary for the procedure was prepared before removing the vials from the freezer, including labeled 15mL falcon tubes containing pre-warmed ES complete medium. After all the preparations were complete, the cryo-vials were transferred from the freezer or liquid nitrogen to the cell culture room. Here, to speed-up the process of thawing, the vials were submersed by half in the water bath and gently shaken. When only a small fragment of frozen cells remained, the vial was brought to the laminar flow hood and 1mL of ES complete medium was added, before transferring the entire mix into the pre-prepared falcon tube. After spinning down the tube and aspirating the supernatant, the pellet was resuspended in a volume of fresh medium depending on the number of cells to seed and the surface of the plate.

6.1.7 Transfection of mESCs with FuGene

The same general procedure was followed for transfection of linearized expression vectors for reporter cell line or plasmid containing gRNA for knock-out experiments. In all the cases, the cells were seeded 18 to 22 hours before the transfection in gelatin coated dishes, in a way to reach at least 60% of confluency the day after. The next day ES complete medium was refreshed, at latest two hours before the starting of the procedure. To perform the transfection in a 10cm dish, 25µg of DNA were mixed in 1.2mL of pre-warmed Opti-MEM medium (Gibco™) (or 4µg of DNA in 200µL of Opti-MEM for a well of a 6-well). Subsequently, 75µL of FuGene

(Promega, E2312) -12µL in the case of transfection of 4µg of DNA- equilibrated at RT for 10 minutes, were added directly in the solution, and after gently pipetting up and down twice, the mix was incubated at RT for 14 minutes. In the meantime, the medium in the plate to be transfected was replaced again with fresh one and, thus, the transfecting solution was added dropwise with 1 mL Gilson pipette, and the plate gently shaken before placing it in the incubator to stay overnight.

6.1.8 Generation of fluorescent reporter cell lines

All stable transgenic cell lines generated in this project derived from a *Sox1*_GFP, *Brachyury*_BFP double knock-in (2KI) line, previously developed in our group, and also adopted for originating the triple KI line bearing the mCherry reporter under the control of the *Eomes* promoter published in Sladitschek and Neveu 2019. Briefly, starting from the well-established *Sox1*_GFP line by Aubert et al. 2003, the H2B-3xTagBFP reporter was targeted to the *Brachyury* locus by homologous recombination. Subsequently, in the case of the *Eomes*_mCherry triple reporter line, mCherry was heterozigously integrated into the *Eomes* locus with the same strategy (Sladitschek and Neveu 2019).

6.1.8.1 Transcriptional activation reporter.

The canonical Wnt sensor and the retinoic acid reporter lines were both originated starting from the *Sox1*_GFP, *Brachyury*_BFP double knock-in (2KI) line, by stable integration in the genome of a construct driving the expression of an NLS-Scarlet-PEST2D reporter. In the case of the canonical Wnt sensor line, seven TCF/LEF binding sites together with a minimal CMV promoter were adopted as cis-regulatory sequences of the Scarlet reporter, to detect β-catenin mediated transcriptional activation, as demonstrated by SuperTop flash luciferase system (Veeman et al., 2003). For the retinoic acid sensor line instead, the Scarlet reporter expression was driven by six retinoic acid Responsive elements (RAREs) derived from the RARβ promoter, combined with the minimal CMV promoter, similarly to the RARE-*lacZ* reporter applied by Rossant and colleagues in the classical mouse line (Rossant et al., 1991). The seven TCF/LEF binding sites were ordered and cloned as two separated pairs of DNA oligos with the MXS chaining method, explained in the dedicated section. More in detail, two pairs of oligos containing three and four contiguous TCF/LEF binding

sites in the case of the Wnt reporter, and one pair of oligos containing six consecutive RAREs, were ordered including at the 5'-end an XhoI restriction site and the overhang deriving from MluI digestion, and at the 3'-end the overhang characteristic of a SalI digestion. The oligos were then annealed, as explained in the dedicated paragraph, and cloned in MluI-SalI digested backbones. The two portions of the Wnt regulatory region were combined in this backbone and subsequently excised as one sequence with MluI-SalI digestion, and cloned upstream of the expression cassette minCMV::Kozak_NLS-Scarlet-PEST2D_BGHpA (polyA sequence)_PGK::HygroR-bGHpA, generated separately by MXS cloning. To generate the stable cell lines, the constructs were linearized before the transfection with an AseI overnight digestion. A list of the oligos used to generate the regulatory regions of the signaling reporter can be found in the Appendix section.

With a similar strategy for the novel *Mafa*_RARE reporter line, three copies of the RARE sequence identified in the *Mafa* locus (Moutier et al., 2012) were multimerized and cloned as one block upstream of the CMV promoter.

Alternative versions of the sable reporter lines in which all nuclei were marked by the iRFP670 reporter, were developed by stable integration in the genome of the construct CAG::H2B-iRFP-BGHpA, granting constitutive expression in every cell.

6.1.8.2 Generation of the mESC sensor lines

For the generation of reporter lines, $8x10^5$ – $1x10^6$ cells of the *Sox1*_GFP, *Brachyury*_BFP 2KI line were seeded, according to the transfection protocol above, in gelatin coated 10cm dish. The transfection mix was composed of 25 μg of DNA linearized by overnight digestion and inactivated for 20 minutes at 60-80 °C and without purification, 1.2 mL of OpiMEM medium and 75 μL of FuGene reagent. The day following the transfection, the antibiotic selection was started by changing the medium with ES complete containing the appropriate antibiotic concentration: for plasmids bearing an FRT-flanked PGK::Neomycin/Kanamycin resistance, 1 mg/mL of G418 (InvivoGen, ant-gn-5, 100mg/mL stock) was adopted, whereas in the case of PGK::HygroR-bGHpA cassette, Hygromycin B (InvivoGen, ant-hg-5) was added at a concentration of 125μg/mL. The selection medium was changed every day after washing once with D-PBS to remove dead cells. On the seventh day of selection the medium was removed, the cells rinsed twice with D-PBS, and then 15mL of D-PBS

were added. The plate was then transferred to the inverted DM IL LED Leica microscope with the HI PLAN 5x/0.22 PH1 objective. Here, using a 20µL Gilson pipetman, individual colonies were dislodged and transferred into single wells of U-bottom 96-well plates. A number of colonies able to fill from one to three 96-well plates was typically selected. In any case, full 96-well plates were processed individually by adding 25µL of 0.05 % Trypsin-EDTA solution to each well with a multichannel pipet, and incubating 5 minutes at 37 °C. Trypsin was then inactivated using 100µL of ES complete medium without antibiotic selection and after several resuspensions with the multichannel pipet, the cells were transferred into flat-bottom 96-well plates containing already 100uL of fresh ES complete in each well. After two or three days of incubation, the 96-well plates containing the growing clones were duplicated by passaging the cells, so to screen the expression of the desired reporter in the replica plates by flow cytometry. When positive clones were identified, they were expanded first into 24-well and then into 6-well plates and finally frozen.

6.1.9 Generation of knock-out (KO) cell lines

6.1.9.1 Single guide RNAs (sgRNA) design and cloning

All the Knock-out (KO) lines were generated by using the CRISPR spCas9 genome editing approach published by Zhang and colleagues and described as follows (Cong et al., 2013). For each gene to KO, a single guide RNA (sgRNA) was designed using the E-CRISPR software curated by the Boutros group (Heigwer et al., 2014). Only gRNAs starting with a G nucleotide, preferentially targeting the central region of large exons and not located at a polymorphic site were considered, and the overhangs 5'-CACC-3' and 5'-ACCC-3' were added, respectively to the forward and reverse complement sequences of the 20bp protospacer, in order to get:

5'- **CACC**Gnnnnnnnnnnnnnnnnnn -3'

3'- Cnnnnnnnnnnnnnnnnnn**CAAA** -5'

A list of all the gRNAs used in this project can be found in the Appendix. The oligos where annealed and phosphorylated according to the procedure described in the annealing oligos section and cloned into the pX330-U6-Chimeric-BBCBh-hSpCas9 backbone, digested with BpiI restriction enzyme. A digestion test with BamHI, which cuts only the backbone without insert, was then used to select the edited plasmids, and

the presence of one copy of the gRNA sequence was confirmed by Sanger sequencing using the primer 5'- TGCTCACATGTGAGGGC -3'.

6.1.9.2 Introducing null mutations in stable lines

For the KO procedure, the cells were passaged the day before the transfection in 6-well dish, at a ratio only slightly lower than normally used. The day after, the plasmid containing the gRNA and Cas9 coding sequence was mixed with another plasmid coding for puromycin resistance in 200µL of Opti-Mem media. The stoichiometry of the two plasmid was at least 9:1 for a total of 4 µg - typically 3.6 µg of gRNA plasmid, with 400ng of puromycin resistance-coding plasmid- to increase the chances that the cells acquiring the puromycin resistance also contained the genome editing construct. Using 12µL of FuGene reagent, the transfection was then performed according to the standard protocol. The day after the transfection, the culture was switched to ES complete medium containing 1µg/mL puromycin (InvivoGen, ant-pr-5, 10mg/mL), and the medium changed daily. Following two days of puromycin selection, cells were passaged, at a ratio from 1:1 to 1:3, to increase the death of non-resistant cells. After a total of four days from the beginning of the selection and five from the transfection, cells were passaged a second time, but in this case in wells containing ES complete medium without antibiotic selection, for additional one or two days. Subsequently, the surviving cells were trypsinized, single cell suspended, and counted to seed 1000 cells in a 10cm dish. Seven days after the single cell seeding, during which no change of medium was performed, each cell formed a colony that could be picked with the same general procedure anticipated for the generation of reporter lines. In this case, 24-96 colonies where collected to screen the edited clones. Following the usual three days of incubation in 96-well, all the clones were expanded into 24-well plates. When reaching a confluency of at least 80% each clone was frozen, while a fraction was re-plated to extract genomic DNA after additional two or three days. Alternatively, upon the first passaging in 96-wells the mother plates or replica plates were used for the DNA extraction, whereas the cells were always frozen from 24-well plates.

6.1.9.3 Identification of the KO clones

The genomic DNA extraction of the putative KO clones was performed as described below in the dedicated section, and the DNA pellet suspended in 100-150µL

of EB buffer. 1 µL of each of this suspension was used to test by PCR the editing of the desired locus as follows. A couple of primers was designed as such to obtain a 600-800bp fragment from PCR of the wild type locus, trying to have the sequence targeted by the gRNA equally spaced between the two primers. A list of the PCR primers used for the KO screening can be found in the Appendix. The PCR reactions were then analyzed by agarose gel electrophoresis. The DNA bands corresponding to a clone, whose locus was considered edited compared to the wild type reference, were gel purified with the QIAquick Gel Extraction Kit (Qiagen), and sent to be sequenced by Sanger method using one of the two genomic PCR primers. The clones showing the desired type of editing in both the alleles, preferentially insertion or deletion causing a frameshift or also deletion of splicing sites, were considered as lead candidates and were further tested by differentiation. However, in absence of those, also clones in which the sequence of only one allele was retrieved were considered for further testing, because being likely either homozygous for the same mutation, or heterozygous with a deletion on one allele larger than the PCR product.

6.1.10 mESCs germ layers differentiation

6.1.10.1 PS-like/Endoderm differentiation

The differentiation procedure here reported to generate cells of the tree primary germ layers, definitive endoderm, mesoderm, and neuroectoderm, through a pluripotent state of epiblast/Primitive Streak like cells derives from the protocol developed by Borowiak and colleagues to generate endoderm progenitors by the use of the small molecules IDE1/IDE2 (Borowiak et al., 2009). The standard IDE1 differentiation was carried out in 6-well plates according to the following procedure. After trypsinization, mESCs were single cell suspended and counted. The seeding density applied for this differentiation was of 2700-3000 cells/cm^2, and in the case of 6-well plates the required number of cells was diluted in 500µL of medium per each well to seed. In this way, the cells were seeded ~ 18 hours before the supply of the differentiation cues, adding 500µL of the diluted cell suspension to 1.5mL of ES complete medium already prepared in every well, and they were incubated overnight. The morning after, the medium was removed and every well was rinsed once with D-PBS (RT) to remove traces of LIF and serum, before adding pre-warmed IDE1 differentiation medium, with the following composition:

Component	Comment	Supplier (Number)	Final concentration
Advanced RPMI	w/o HEPES, w/o L-Glutamine	Invitrogen (12633)	98 % (v/v)
Fetal Bovine Serum	ES-qualified	Millipore (ES-009-B)	0.2% (v/v)
100x L-Glutamine	200 mM	Invitrogen (25030)	1x (2 mM)
100x Penicillin-Streptomycin	10,000U/ mL 10,000U/ mL	Invitrogen (15140)	1x (100 U/ mL each)
IDE1	10mM in DMSO	Tocris Bioscience (4015)	1 µM

The cells were then incubated for 48 hours, after which the medium was changed every day, with a D-PBS wash in case of large number of dead cells. When the differentiation was carried out in 96-well plates, after trypsinization and suspension of the cells of every well in 200µL of ES complete, 6-9µL of cells were transferred to a new 96-well plate containing 150µL of medium per well.

6.1.10.2 Neuroectoderm differentiation

The monolayer differentiation directed towards neuroectoderm cell types was performed according to the indication of Ying and colleagues (Ying et al., 2003; Ying & Smith, 2003). The standard differentiation was carried out in 6-well plates according to the following procedure. After trypsinization, mESCs were single cell suspended and counted. The seeding density applied for this differentiation was of 10000-11000 cells/cm^2, and the required number of cells was diluted in 500µL of medium per each well of 6-well plate to seed. The cells were seeded ~ 18 hours before the supply of the differentiation cues, adding 500µL of the diluted cell suspension to 1,5mL of ES complete already prepared in every well and incubating overnight. The following morning, the medium was removed, and every well was rinsed once with D-PBS (RT) to remove LIF and serum, before adding pre-warmed N2B27 differentiation medium prepared as follows:

Component	Comment	Supplier (Number)	Final concentration
Neurobasal medium	w/o L-Glutamine	Invitrogen (21103)	48,4 % (v/v)
DMEM/F12	w/o HEPES with L-Glutamine	Invitrogen (21041)	48,4 % (v/v)
100x L-Glutamine	200 mM	Invitrogen (25030)	1x (1,25 mM)
100x Penicillin-Streptomycin	10,000U/ mL 10,000U/ mL	Invitrogen (15140)	1x (100 U/ mL each)
B27 Supplement	(50x) with vitamin A	Invitrogen (17504)	0,5 x
N2 Supplement	(100x) 1mM transferrin, 86 µM insulin, 2 µM progesterone, 10 mM putrescine, 10 µM selenite	Invitrogen (17502)	0,5 x
BSA fraction V	20 mg/mL	SIGMA (A3311)	10 µg/mL
Insulin	20 mg/mL, recombinant human	SIGMA (91077C)	10 µg/mL
Retinoic Acid	10mM in DMSO	SIGMA (R2625)	1 µM

The standard neuroectoderm differentiation requires N2B27 containing vitamin A (N2B27 vit.A); when the formulation of N2B27 not containing vitamin A was used instead, it was explicitly stated as N2B27/N2B27 without vitamin A. The switch from ES complete to N2B27 is considered the time 0 of the experiment, and 24 hours later (day 1 from the switch), N2B27 vit.A containing retinoic acid at a concentration of 1µM was added. When possible, the cells were incubated for 48 hours, after which the medium containing RA was changed every day, following a wash with D-PBS to remove the dead cells. For the experiments of transient exposition to RA, or RA pulse, the cells were kept for 24 hours in N2B27 vit.A, 24 hours in N2B27 vit.A with RA, and from then onwards (that is from day 2 from the switch) N2B27 without vitamin A was used to continue the differentiation in absence of RA.

In the case of neuroectoderm differentiation in 96-well plates, after trypsinization and suspension of the cells of every well in 200µL of ES complete, 15-20µL of cells were transferred to a new 96-well plate containing 150µL of medium per well.

6.2 Flow cytometry analysis and cell sorting

6.2.1 Flow cytometry analysis by DB LSRFortessa™ Analyser

To prepare the samples to be analyzed by flow cytometry, the cells were trypsinized as in the passaging procedure, and the digestion stopped with 1mL of harvesting medium (10% FCS DMEM, with FCS from GIBCO) in the case of differentiation in 6-well plate. The cells were then resuspended pipetting up and down and transferred in 1.5mL microtubes to be spun down 2 min at 2000rpm or 5 min at 500rpm (RT). The supernatant was then aspirated, and the cell pellet was resuspended in 200µL of D-PBS or more depending on the number of cells. To ensure the separation of every cell in the suspension, 200µL of cells were then strained through a 40µm filter (BD Biosciences) into polystyrene flow tubes (Falcon, 352058). The samples were then measured according to manufacturer recommendations, diluting too dense suspensions with additional D-PBS. An average number of 2 $x10^5$ of cells were usually acquired, or at least $5x10^4$ when not possible, based on hierarchical gating of cells (FSC-A and SSC-A), and singlets (SSC-H and SSC-A). The same lasers and detectors settings were applied throughout the project. For the measurement of fluorescent reporters, the following laser-detectors parameters were used: BFP (405-450/50-A), GFP (488-530/30-A), mCHERRY and SCARLET (561-610/20-A), iRFP (640-670/14-A). When entire 96-well plates had to be analyzed, the sample were prepared as for passaging, but inactivating the trypsin digestion with 120µL of harvesting medium. Cells were resuspended with a multichannel pipette and transferred into U-bottom 96-well plates containing 80µL of D-PBS, without further filtration of the cells. The samples were thus measured with the BD™ HTS plate reader, automatically stopping the acquisition in each well when 5000 events were recorded in the single cell gate (SSC-H vs SSC-A), or when the elapsing time was reached, based on the following settings:

- Standard throughput
- Sample flow rate: 2.0 µL/sec
- Sample volume: 40µL
- Mixing volume: 100µL
- Mixing speed: 180µL/sec
- Number of Mixes: 2 cycles
- Wash Volume: 200µL

6.2.1.1 Data analysis

All the flow cytometry data were exported as FCS3 files and analyzed with the FlowJo ™ v10 software. Every analysis consisted first in selecting the cells, according to their size and complexity, compared to debris and fragments of membranes, plotting FSC vs SSC. Then, only the single cells were gated in the SSC-H vs SSC-A plot, and the resulting sub-populations were visualized as histograms to represent the distribution of the fluorescence detected in one channel in all the cells, or pseudocolor density plots to compare simultaneously two different fluorescence intensities. When further gates were applied to quantify the fraction of cells expressing a given fluorescent reporter, the gate position was set according to Ab2.2 WT controls analyzed at the same day of differentiation to avoid autofluorescence background.

6.2.2 Bulk cell sorting by MoFlo XDP-Beckman Coulter, DB FACS Aria Fusion and Melody

To sort the cells expressing fluorescent reporters, either for RNA sequencing or re-seeding in culture, the samples were prepared similarly to the flow cytometry analysis described above. Only in the case of re-plating of the sorted cells, trypsin was blocked with ES complete medium, whereas in the other cases the in-house harvesting medium was used. Cells were resuspended and spun down, and the pellet mixed in a volume of sorting buffer (1-2% FBS, 2.5mM EDTA, 1% Pen-Strep in D-PBS) according to the number of cells. The suspension was filtered with the 40μm cell strainer into 5mL polypropylene flow tube (BD Falcon, 352063), to reduce cell adhesion to the tube. The same tubes were also used for sample collection, adding 1mL of medium in each to increase the viability during the sorting process. For samples intended for RNA sequencing the sorting was performed keeping the collection tubes at the constant temperature of 4 °C, whereas the collection of cells to re-seed was carried out at RT to avoid changes of temperature. When allowed by the availability of fluorescent channels not occupied by any reporter expressed by the cells, DAPI or DRAQ7 ™ staining were used to sort only live cells, otherwise FSC vs SSC gating was used to exclude debris and membrane fragments. The gate to select the population expressing the desired fluorescent reporter was set based on Ab2.2 WT controls differentiated in parallel with the sample to analyze. The sorting was performed in Purity or 4-way purity modes, at a maximum flow rate of 4000 events/sec, isolating a

minimum of 10000 cells for populations with frequency lower than 5%, or at least 50000 cells in the other cases. When the sample collection was complete, the cells were transferred into 15mL Falcon tubes to be spun down. In case of re-seeding, the cells were resuspended directly in the culture medium and transferred in the coated dishes, whereas for samples destined to RNA sequencing, another intermediate round of centrifugation was required to pellet the cells in 1.5mL microtube and freeze the pellet at -80 °C, after removing the supernatant.

6.3 RNA sequencing

6.3.1 Total RNA extraction with mirVana ™ kit

Total RNA destined for RNA sequencing was extracted by mirVana RNA column purification kit (Ambion), according to manufacturer recommendations. All the solutions dedicated to RNA extraction were handled with filter tips only, and bench and pipette wiped clean with 70% ethanol. To start the procedure, the cell pellet was thawed on ice and the 1.5mL microtube gently flicked. 600 µL of Lysis solution were then added, and cell fragmentation was favored by ~10 passages through sterile 1mL syringes with 20-gauge needle. 60 µL of miRNA Homogenate were mixed with the cell lysate by inverting the tubes several times. After leaving the mixture 10 min of ice, 600 µL of Acid-Phenol: Chloroform (from 4°C) were withdrawn avoiding the upper aqueous phase, added into the microtubes, and these were inverted for 30-60 sec. Following 5 min of centrifugation at maximum speed and RT, ~ 500 µL aqueous upper phase containing the RNA were then transferred to a fresh tube, avoiding any contact with the protein interphase. 1.25 volumes of 100% ethanol were mixed with the aqueous phase, and 700 µL of the lysate/ethanol solution were applied on the columns provided by the kit. Two centrifugation steps for 30 sec at 10000 x g were required to pass the entire lysate through the filter, discarding the flow-through after each centrifugation. Three washing steps were performed, applying to the column first 700 µL of Wash solution 1 and twice 500 µL of Wash solution 2/3, spinning down every time for 30sec at 10000 x g and discarding the flow-through. An additional centrifugation for 1 min, following the last wash, allowed the removal of residual fluid from the filter. The column was finally transferred into a fresh collection tube, and the RNA was the eluted by applying 100 µL of pre-heated (95°C) Elution solution onto

the filter and spinning down at maximum speed for 30-60 sec. The RNA was then stored at -80°C.

6.3.2 RNA preparation for sequencing

All RNA samples to sequence were quantified and quality checked by Nanodrop and capillary electrophoresis with Agilent 2100 Bioanalyzer System. Samples with initial RNA concentrations greater than 5ng/µL (~20-50000 cells) were analyzed by the Agilent RNA 6000 Nano chip according to manufacturer recommendations. When the initial concentration was lower than 5ng/µL or a first measurement on the Nano chip indicated low amount of RNA, the samples were analyzed with the Agilent RNA 6000 Pico chip. For the cDNA libraries preparation using TruSeq RNA Sample Preparation (Illumina), 150ng of every RNA samples were prepared in a total volume of 50µL RNase-free water. Samples with concentrations lower than 3ng/µL were concentrated through Eppendorf Vacufuge concentrator. The libraries were run on Illumina NextSeq 500 in the 75 bp single-end regime, with 18-20 sample multiplexing.

6.3.3 RNA-seq analysis

The analysis of the RNA sequencing results was performed as described in Sladitschek and Neveu 2019. Briefly, the reads were aligned using the Bowtie software and the mRNAs read counts of each ENSEMBL ID were established applying a custom Phyton script. The normalization of the read counts was done in a way to allow comparison of gene expression changes across different samples, by adjusting for the total number of aligned reads, and matching to the identity line read counts of highly expressed genes. All the heatmaps presented in this report were generated using R studio and the Pheatmap package. To determine the relative expression values between different samples, for every gene analyzed it was calculated the $\log_2$ of the fold change between the read counts across the samples tested and the mean count for that gene.

6.4 Confocal imaging

All the confocal imaging experiments were carried out in gelatin coated 8 well µ-slides (ibidi), using an inverted Leica SP8 confocal microscope with a 40x PL Apo 1.1 W objective 8, and with the stage for imaging mounted inside a humidified incubation chamber set at 37 °C and 5 % CO_2.

6.4.1 Live cell imaging of reporter cell line

Every imaging experiment started by obtaining a single cell suspension of the reporter lines to differentiate, according to the general differentiation procedure described above. Four different seeding densities were usually adopted for the same line, ranging from 6000 to 12000 cells/cm^2 for the Primitive Steak-like/IDE1 differentiation, and 10000 to 20000 cells/cm^2 for the N2B27 RA differentiation. After seeding, the differentiation was carried out as indicated by the respective protocols, being however more careful to cell detachment when washing with D-PBS or changing the media. On the desired day of differentiation, the differentiation medium was renewed prior to the imaging experiment, with an intermediate D-PBS wash to remove floating dead cells. For the imaging procedure, sequential acquisition was preferred to increase the imaging processivity and avoid crosstalk, with TagBFP and mCHERRY/Scarlet detected in the first sequence, and GFP (with eventually iRFP whenever present) in the second. The lasers wavelengths to excite the fluorophores and their intensities were: 405 nm (5-10%) for BFP; 488 nm (5-10%) for GFP; 561 nm (10-15%) for mCHERRY or (1-5%) for (SCARLET); 633 nm (0.5-2%) for iRFP. The Argon laser was always set to 20% of its maximum intensity. HyD detectors were used for any acquisition with voltages ranging from 350% to 500%, and the emission spectra detected were: 415-480 nm (TagBFP), 495-560 nm (GFP), 585-670 (mCherry), 575-660 nm (Scarlet), 650-720nm (iRFP). A resolution of 1024x1024 was always applied with frame averaging of six to reduce imaging noise. The single channel images were analyzed and merged using ImageJ (Fiji).

6.4.2 Time-lapse confocal imaging of reporter cell line

For the time-lapse experiments, the same procedures of the standard live imaging described above were essentially applied, with the following differences. The Leica MatrixScreen software was used to set imaging lapses every 15 minutes in 8 different fields of the same well, and on three different position on the z axis (spaced 5µm from each other). Regular pumping of water to the objective was enabled by setting automatic refilling (100Hz). To ensure constant humidity and avoid draining out of medium from the wells, beakers placed in the incubation chamber, and containing sponges and Whatman 3MM paper, were regularly refilled with water. The autofocus map, required to set the starting z position for the imaging in every field was based on

the constitutive nuclear marker H2B-iRFP, using a 633 nm laser intensity of 0.5%. All the other excitation lasers intensities were also decreased in these set of experiments compared to standard live imaging, to enable cell survival upon repeated imaging for more than 24 hours: BFP (405nm, 2%), GFP (488nm, 2%), Scarlet (561nm, 0.5%). To reduce the imaging time and increase cell viability also the resolution was lowered to 512x512, with four frames averaging. The analysis of the single time frames in every position to generate a time-course was carried out with a custom Python script.

6.5 Molecular biology techniques

6.5.1 Genomic DNA extraction

The cells from which to extract the DNA were collected and typically spun down in 1.5mL microtube and, after removing the supernatant, the pellet was frozen at -20 °C. To start the extraction procedure from a well of a 6-well dish, 500 µL of Spooling Buffer, containing 5 µL of Proteinase K (20mg/mL), were added to the cell pellet, and the microtube was incubated overnight in a thermomixer at 55 °C, with gentle shaking (300-400rpm). For extraction from 24-well or 96-well plates, the pellet of each individual well was incubated with 200 µL of Spooling Buffer, with 2 µL of Proteinase K. The day after, when the pellet was dissolved, 125 µL of Brine Buffer were added to the 500 µL of Spooling Buffer from the previous step (50 µL of Brine were used instead for 200 µL of Spooling,). The microtube was then gently shaken in the thermomixer at 55 °C, until the solution became clear and free of debris. To precipitate the DNA, one volume of isopropanol (625 µL for extraction from 6-well dish) at RT was added and the tubes inverted repeatedly, till a white precipitate formed, containing the genomic DNA. The tubes were spun down at RT for 10 min at maximum speed, the supernatant discarded by inverting the tubes, and the DNA pellet washed with 1mL of 70% Ethanol. After inverting the tubes briefly, the pellet was spun down again by centrifugation at maximum speed for 5-10 min at RT. The supernatant was discarded completely, first by inverting the tubes, then by pipetting out the remaining volume. The pellet was then air-dried by placing the tubes with the lid open in the thermomixer for 5 min at 55 °C. Avoiding the complete drying out of the pellet, the genomic DNA was then resuspended by adding an appropriate volume of EB buffer (QIAGEN) for the size of the pellet (100 µL for a confluent well of a 96-well, 100-200 µL for 24-well

plate), and incubating in the thermomixer at 55 °C with gentle shaking overnight. The day after, 1 µL of the DNA solution could be directly used as template for genomic PCR. Recipes of the Spooling and Brine buffers can be found in section 6.5.7.

6.5.2 PCR protocol

All the primers used in this project were designed manually using the DNA cloning software "Geneious", in a way to have length of 20-25 nucleotides, excluding eventual overhangs, melting temperature (Tm) predicted by the online "oligo analyzer tool" (idtdna.com) comprised between 60-64 °C (the Tm predicted by Geneious was typically ~8 °C lower), with a difference not greater than ~2-3 °C between the Tms of forward and reverse primer. Absence of secondary structure at the Tm of each primer, and of heterodimers formation were also checked, as well as it was avoided the design in a genomic region containing polymorphisms or repetitive sequences. The primers were then ordered from Sigma as desalted, lyophilized oligonucleotides, and resuspended with EB buffer at a concentration of 100µM, incubating the tubes in a thermomixer at 55°C and 500rpm for at least one hour. All the PCR reactions were performed with the Phusion High-Fidelity DNA polymerase, in the following reaction mix:

Component	**Volume (1x)**
5X HF/GC buffer	5µL
DMSO	0.75µL
dNTP (10mM)	0. 5µL
forward primer (100µM)	0.25µL
reverse primer (100µM)	0. 25µL
DNA template (genomic)	1µL (100-200ng)
Phusion DNA Polymerase	0.2µL (0.4U)
Nuclease-Free H_2O	Up to 25µL

The PCR cycles were set according to the Phusion DNA Polymerase product information, using an extension time of 30sec/kb. In most cases, the temperature of annealing of new couple of primers was empirically tested by PCR with gradient of temperatures, but usually it could be determined adding 2°C to the Tm of the primers.

The PCR reaction was performed in a thermocycler (Bio-Rad, S-1000) adopting the following program.

Cycle step	Temperature	Time	Cycles
Initial denaturation	98°C	2 min	1
Denaturation	98°C	20 sec	(25-) 39
Annealing	60-65°C	30 sec	
Extension	72°C	30 sec/Kb	
Final extension	72°C	2 min	1

The PCR reactions were analyzed through DNA gel electrophoresis on 0.8-2% agarose (Sigma) gels. In some cases the entire PCR reaction was loaded into the gel and the product extracted and purified using the QIAquick Gel Extraction Kit (Qiagen), whereas in others only 4µL of the PCR volume were analyzed and, when the product was showing one clear band, the remaining PCR reaction was purified directly on column with the Qiaquick PCR purification kit (Qiagen). For the instances where the PCR fragment was not intended for cloning inside a vector, as for the genomic PCR to test CRISPR editing, the entire purified PCR, eluted in 30 µL, was sent for sanger sequencing (Eurofins GATC Biotech).

6.5.3 MXS cloning strategy

The strategy adopted for all the DNA cloning necessary in this project is the MXS chaining system, previously published by our group (Sladitschek & Neveu, 2015). The strategy is essentially based on generating cloning blocks presenting MluI (M) and XhoI (X) sites upstream of the sequence of interest, and SalI (S) restriction site downstream. Thus, every block to be cloned can be digested with MluI-SalI and inserted in an empty vector digested in the same way. The other fundamental aspect is that the overhangs generated by the XhoI and SalI digestion are compatible and, when ligating together an insert digested MS in a backbone digested MX, they generate a translatable scar, containing no restriction site. In this way, to combine cloning blocks one with the other it is sufficient to digest the vector containing the block that will be downstream with MluI-XhoI, and the vector or the PCR product that will go to the 5'-end with MluI and SalI, in such a way to obtain: 5'-MluI-XhoI-Block2-SCAR (from

X-S ligation)-Block1-SalI-3'. Another functional unit can then be cloned upstream of this new block in the same way, and so on.

6.5.4 Restriction digestion

All the restriction digestions necessary for this project were performed using NEB restriction enzymes and the provided buffers according to manufacturer recommendations. In the case of simultaneous digestion with two or more enzymes the compatible digestion buffers to use were selected from the NEB catalog. The standard plasmid digestion for cloning was carried out at 37°C for at least 3 hours in the following mix:

Component	**Volume (1x)**
DNA	2µg
10X NEB buffer	3.5µL
Restriction Enzyme	1-2 µL
Nuclease-Free H_2O	Up to 35µL

Even for double digestions the total volume of enzymes added to the mix was never exceeding 1/10 of the entire volume, to avoid inhibition of the reaction by the glycerol contained in the enzyme solution. Particularly in the case of backbone digestion, or linearization of plasmids for transfection, the total amount of the enzyme was added in two steps to increase the efficiency of the digestion, and when possible, the reaction was incubated overnight. To prepare the DNA for transfection and stable cell line generation, the amount of DNA linearized depended on the surface of the plate to transfect, as such to digest 4µg for a well of a 6-well dish and 25µg for a 10 cm dish. To clone PCR products, instead, all the DNA eluted after the purification was used in the restriction enzyme digestion. The digestions intended for cloning were analyzed by Agarose gel electrophoresis, and the fragments of interest were excised and purified with the QIAquick Gel Extraction Kit (Qiagen).

6.5.5 Annealing oligos procedure

All the DNA oligos used in this project were ordered from Sigma. The oligos were solubilized at a 100µM concentration adding the amount of EB buffer indicated by the manufacturer, and shaking the tubes for 30 min-1 hour at 500-700 rpm in a

thermomixer set at 55°C. To anneal forward and reverse strand 50 µL of each of the two solutions were mixed in a 1.5mL microtube, which was submersed into boiling water and let cool down at RT. The annealed oligos were then diluted by adding 1.4mL of EB buffer, and 1 µL of this dilution was then used for a phosphorylation reaction composed as follows:

Component	**Volume (1x)**
Annealed oligos (1.5mL)	1µL
10X T4 Ligase buffer	1µL
T4 Polynucleotide Kinase (PNK)	0. 5µL
Nuclease-Free H_2O	Up to 10µL

The reaction was then incubated for 40min – 1 hour at 37°C, and the product was ligated in the digested receiving vector, using a 2:1 phosphorylated insert: backbone ratio.

6.5.6 DNA Ligation and competent bacteria transformation

Purified DNA fragments digested with restriction enzymes were ligated with a digested receiving vector by mixing in a 1.5mL microtube the following components:

Component	**Volume (1x)**
DNA Insert	8µL
DNA Vector	1µL
10x T4 DNA Ligase buffer	1µL
T4 DNA ligase	0.25µL

The reaction was incubated for 15-20 min at RT. In the meantime, 10 min before the end of the ligation 100 µL aliquots of Mach T1® transformation competent bacteria were transferred from -80°C into ice, avoiding excessive handling of the tubes. When the reaction was terminated, 50 µL of bacteria suspension were pipetted out after gentle mixing and added directly to the ligation tube. The tubes were then incubated 2 min on ice, then placed for 1 min into a thermomixer pre-heated at 42°C, after which they were immediately chilled on ice for at least 30 sec. For transformation of bacteria with plasmid containing Ampicillin resistance, no further recovery was necessary, and the bacteria could be directly pipetted onto the LB-Agar plate with

ampicillin and spread out with the use of sterile glass beads. The plates were then incubated overnight at 37°C to let the bacteria grow, and the day after single colonies could be inoculated in 2-3mL of 2xYT medium. After a minimum incubation of six hours at 37°C with shaking at 180-200rpm, typically extended overnight, the bacteria could be pelleted by 1 min centrifugation at maximum speed and the plasmid recovered using Qiagen Miniprep kit following manufacturer instructions. The cloning step was tested by screening at least four different colonies with digestion test. When the successfully edited clone was selected, a new overnight inoculation was started using 15 µL of the original mini-inoculation in 2mL of fresh 2xYT. The morning after, 500 µL of the inoculation were mixed with 500 µL of 85% glycerol in a labeled cryovial, to make a permanent stock of the plasmid at -80°C.

6.5.7 Buffers and solutions

D-PBS: 2.7 mM KCL, 1.47 mM KH_2PO_4, 137 mM NaCl, 8.1 mM N_2HPO_4, pH 7.4

Spooling buffer: 75mM NaCl (= 15mL 5M NaCl), 25mM EDTA ph8 (=50mL 0.5M EDTA), 1% SDS (10g), H_2O up to 1L

Brine buffer: 5M NaCl

TAE-Buffer (50x): 40 mM Tris-acetate, 1 mM EDTA pH 8,3

LB media: 1 % w/v Tryptone, 0,5 % w/v Yeast Extract, 1 % NaCl, pH 7,0

YT media: 1,6 % w/v Tryptone, 1 % w/v Yeast Extract, 0,5 % NaCl, pH 7,0

LB agar plates with antibiotics: LB medium was supplemented with 1,5 % w/v agar, autoclaved and warmed up in the microwave till boiling point. The solution was then let cool down at RT and heat sensitive antibiotics were added just before pouring into Petri dishes.

Ampicillin: Stock solution 100mg/mL (SIGMA); used at 50 µg/mL in liquid media, such as 2xYT, and 100 µg/mL in LB-agar plates.

7 APPENDIX

7.1 List of signaling molecules

Signaling molecule	Stock concentration	Supplier (Number)	Final concentration
CHIR99021	20mM	Tocris (4423)	3µM
XAV939	10mM	Tocris (3748)	1µM
SB431542	50mM in DMSO	Tocris (1614)	10µM
Noggin	100 µg/mL (0.1%BSA) in H_2O	Peprotech (250-38)	50 ng/mL
s-FRP1	100 µg/mL (0.1%BSA) in H_2O	Peprotech (120-29)	50 ng/mL
Dkk1	100 µg/mL (0.1%BSA) in H_2O	Peprotech (120-30)	50 ng/mL
Wnt3a	100 µg/mL in H_2O	Peprotech (315-20)	25-50 ng/mL
Activin A	200 µg/mL	Peprotech (120-14E)	50 ng/mL
Bmp4	10 µg/mL	Peprotech (120-05)	10 ng/mL
AGN193109	5mM in DMSO	Tocris (5758)	200nM-1µM
Vitamin A (Retinol)	250µg/mL in DMSO	SIGMA (R7632)	20-50ng/mL
Ch55	10 mM in DMSO	Tocris (2020)	50nM
LG100268	5mM in DMSO	Tocris (5920)	100nM
Retinoic Acid	10mM in DMSO	SIGMA (R2625)	1 µM
IDE1	10mM in DMSO	Tocris Bioscience (4015)	1 µM

7.2 List of primers used to clone the gRNAs for knock-outs

Gene	Primer	Sequence (5'-3')
Cyp26a1	Forward	CACCGCGCCCATCACCCGCACCGT
	Reverse	AAACACGGTGCGGGTGATGGGCGC
Aldh1a2 (gRNA1)	Forward	CACCGGACAAGCTTGCAGACTTGG
	Reverse	AAACCCAAGTCTGCAAGCTTGTCC
Aldh1a2 (gRNA2)	Forward	CACCGGAATGGCTTACCGCCATTT
	Reverse	AAACAAATGGCGGTAAGCCATTCC
RARα	Forward	CACCGGTGGGCGAGCTCATTGAGA
	Reverse	AAACTCTCAATGAGCTCGCCCACC
RARβ	Forward	CACCGCGTGGTGTATTTACCCAGC
	Reverse	AAACGCTGGGTAAATACACCACGC
RARγ	Forward	CACCGTGGGACAAGTTCAGCGAGC
	Reverse	AAACGCTCGCTGAACTTGTCCCAC
Chordin	Forward	CACCGACAGGACCGAGGTCGCAGG
	Reverse	AAACCCTGCGACCTCGGTCCTGTC
Noggin	Forward	CACCGGAAGTTACAGATGTGGCTG
	Reverse	AAACCAGCCACATCTGTAACTTCC

7.3 List of primers used to test CRISPR genome editing

Gene	Primer	Sequence (5'-3')
Cyp26a1	Forward	TGCTTCAGGTAAGGGAGCTCG
	Reverse	AGCTAAAGTGGATCTCAGATAGAGAGG
Aldh1a2 (gRNA1)	Forward	TGTGAAGTTCAAGAAGCAGACAAG
	Reverse	CCTTAGTAGTGATGCAATGCTATGTC
Aldh1a2 (gRNA2)	Forward	GTACAGGCAGAGAATAGCACATG
	Reverse	CGGTCTAACTTAATTCTCAGCAGC
RARα	Forward	AATAGGATTTAGGATATGAGTCAGACAGG
	Reverse	ATGGTGAGGGTGGTGAAGCCG
RARβ	Forward	GATCTGTAGCTGGATTGATATTTAAAAGATG
	Reverse	TGCACACTAGGAAAGGCAGGC
RARγ	Forward	CCAGTAAGGAAAGCCTGAAC
	Reverse	CCGACAGCTATGAACTGAGT
Chordin	Forward	TGGAAGGCTCTACACTAGGTC
	Reverse	GCAGGGTGTTCAAACAGGATG
Noggin	Forward	GATCCATCAAGTGTCTGGGC
	Reverse	TGGTGGACCTCATCGAACATC

7.4 List of oligos to generate the regulatory sequences of the transcriptional activation reporters

Gene	Primer	Sequence (5'-3')
Canonical Wnt sensor	Forward (3xTCF/LEF)	CGCGTAACTCGAGAAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGG
	Reverse (3xTCF/LEF)	TCGACCCCTTTGATCTTACCCCCTTTGATCTTACCCCCTTTGATCTTCTCGAGTTA
	Forward (4xTCF/LEF)	CGCGTAACTCGAGAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGGCGCGAGATCAAAGGGG
	Reverse (4xTCF/LEF)	TCGACCCCTTTGATCTCGCGCCCCTTTGATCTTACCCCCTTTGATCTTACCCCCTTTGATCTCTCGAGTTA
Retinoic Acid sensor	Forward (6xRAREs)	CGCGTAACTCGAGGAAGGGTTCACCGAAAGTTCACTCGCGAAGGGTTCACCGAAAGTTCACTCGCGAAGGGTTCACCGAAAGTTCACTCGCG
	Reverse (6xRAREs)	TCGACGCGAGTGAACTTTCGGTGAACCCTTCGCGAGTGAACTTTCGGTGAACCCTTCGCGAGTGAACTTTCGGTGAACCCTTCCTCGAGTTA
Mafa/ Retinoic Acid sensor	Forward	CGCGTAACTCGAGAGGTCAGAAGTTCAAGGTCAGAAGGGGAGCCCAGGTCAGAAGTTCAAGGTCAGAAGGGGAGCCCAGGTCAGAAGTTCAAGGTCATCGCG
	Reverse	TCGACGCGATGACCTTGAACTTCTGACCTGGGCTCCCCTTCTGACCTTGAACTTCTGACCTGGGCTCCCCTTCTGACCTTGAACTTCTGACCTCTCGAGTTA

8 ACKNOWLEDGEMENTS

When four years ago I moved to Germany to start my PhD, I could have hardly imagined that I was about to start the most important experience of my life. I am eternally grateful to EMBL for giving me the opportunity to grow up as a scientist and maybe even more importantly as a human being. One of the thing I will miss the most of my PhD here is the environment allowing everyone to work according to their needs, being only limited by your own imagination, together with the possibility to grow just by osmosis of science deriving from the interaction with brilliant minds of so many various nationalities. On my arrival, I did not feel adequate to be in such a place and, even though I gave 100% of what I could for four year, I am still not sure I took full advantage of this opportunity. If I am here writing these words, the first person to thank is of course my supervisor, Pierre. I still remember vividly our first meetings in the interview week and the connection I felt was immediately established, which convinced me to emigrate barely speaking English. The greatest fear I had when I first considered to go abroad for a PhD was to be left alone with a project to deal with and all the related problems, but I can really say you were always two steps away to help me even with the smallest issues. During this time, we had the possibility to reach a profound comprehension and respect of each other. Thank you again for granting me this opportunity.

I would like to thank all my TAC members, and in particular Jamie and Ana for all their useful feedbacks and the professionality they showed in being additional supervisors of my project. Although we did not have many chances to meet you fueled my energy with your encouragements towards a continuous improvement.

All over my PhD I was adopted by the flow cytometry facility, that helped my work throughout the countless hours of flow cytometry. A special thanks to Malte and Diana, who were always available to support me since the beginning of the project and gave me the possibility to learn how to use so many different instruments.

A special thanks to the members of the Neveu group. I think we can really say we were lucky to be in a small group and to have the possibility to get so close and become friends. Sebastian, you have always been a bit as a big brother in the lab helping the new rookie. I could always count on you also with all the issues outside of

the lab, regarding the German language or problems of grow up men. I hope you will remember a bit of everything I thought you about Italy as I will for your teachings about the life here in Germany. Thanks also for the translation of the summary that costed you valuable swiss hours. Amit, thank you for being present at any moment of day and night and worrying for my over-working. You are one of the kindest persons I know, and I think I owe you an apology for not being as present and as a good friend as you are. Guys, I hope we will always keep in touch wherever our lives will lead us.

The polar star guiding me through my PhD was Lucia. You are the reason why no matter what will be of my life I will be always grateful to EMBL for giving us the opportunity to meet. We have started this hard path together as friends, and conquering your heart was truly an epic labor, but I am glad I did it. Thanks to you I never stop improving as a person and opening my mind. You are one of the brightest persons I know, I really admire you, and I wish I could become only half as a good scientist as you are. Thank you for all your scientific inputs, the help you gave me with preparation of presentations, TAC reports, discussion about the project; you should be acknowledged in all of it. There are not enough words to express all I think about you.

Thanks to all the friends I met here in Heidelberg. Thanks to Shahraz, Jacob, Ewa and Kuba; the kicker brought us together, but our bond goes beyond it. Shahraz, I really like you except from when you want to be too arrogant. I think I can learn from your way of living. Jacob, we immediately felt connected and we could be very good friends if only you would decide to work a bit less. Thanks to Sergio, Alessio, Matteo and Christian for all the things we did together. I hope we built a friendship over these years that will last wherever we will live in the future. Sergio, you helped me since the beginning of this adventure. We lived together, which makes you the first and likely only roommate of my life, and we were always there for each other in the good and often in the bad times. I wish we will both learn to live in a more equilibrated way helping each other in that direction.

Grazie a mio padre e mia madre che hanno sempre supportato la mia scelta di svolgere un dottorato all'estero, senza mai mettere al primo posto il dolore di genitori il cui figlio va a vivere così lontano da loro per lavoro. Spero che siate orgogliosi di me e che questa piccola soddisfazione possa ripagare almeno in parte tutte le lacrime che avete versato dopo ogni mia partenza da casa.

9 BIBLIOGRAPHY

Acampora, D. et al. (2016). Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast. *Cell Reports*, *15*(12), 2651–2664. https://doi.org/10.1016/j.celrep.2016.05.041

Ambrosetti, D. C., Basilico, C., & Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Molecular and Cellular Biology*, *17*(11), 6321–6329. https://doi.org/10.1128/mcb.17.11.6321

Andoniadou, C. L., & Martinez-Barbera, J. P. (2013). Developmental mechanisms directing early anterior forebrain specification in vertebrates. *Cellular and Molecular Life Sciences*, *70*(20), 3739–3752. https://doi.org/10.1007/s00018-013-1269-5

Andrews, P. W. (1984). Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Developmental Biology*. https://doi.org/10.1016/0012-1606(84)90316-6

Ang, H. L., Deltour, L., Žgombić-Knight, M., Wagner, M. A., & Duester, G. (1996). Expression patterns of class I and class IV alcohol dehydrogenase genes in developing epithelia suggest a role for alcohol dehydrogenase in local retinoic acid synthesis. *Alcoholism: Clinical and Experimental Research*, *20*(6), 1050–1064. https://doi.org/10.1111/j.1530-0277.1996.tb01946.x

Arnold, S. J. et al. (2000). Brachyury is a target gene of the Wnt/β-catenin signaling pathway. *Mechanisms of Development*, *91*(1–2), 249–258. https://doi.org/10.1016/S0925-4773(99)00309-3

Arnold, S. J., Hofmann, U. K., Bikoff, E. K., & Robertson, E. J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development*, *135*(3), 501–511. https://doi.org/10.1242/dev.014357

Arnold, S. J., & Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, *10*(2), 91–103. https://doi.org/10.1038/nrm2618

Aubert, J. et al. (2003). Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(SUPPL. 1), 11836–11841. https://doi.org/10.1073/pnas.1734197100

Aubert, J., Dunstan, H., Chambers, I., & Smith, A.. (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nature Biotechnology*, *20*(12), 1240–1245. https://doi.org/10.1038/nbt763

Aubin, J., Davy, A., & Soriano, P. (2004). In vivo convergence of BMP and MAPK signaling pathways: Impact of differential Smad1 phosphorylation on development and homeostasis. *Genes and Development*, *18*(12), 1482–1494. https://doi.org/10.1101/gad.1202604

Avantaggiato, V., Acampora, D., Tuorto, F., & Simeone, A. (1996). Retinoic acid induces stage-specific repatterning of the rostral central nervous system. *Developmental Biology*, *175*(2), 347–357. https://doi.org/10.1006/dbio.1996.0120

Bachiller, D. et al. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature*, *403*(6770), 658–661. https://doi.org/10.1038/35001072

Bain, G., & Gottlieb, D. I. (1994). Expression of retinoid X receptors in P19 embryonal carcinoma cells and embryonic stem cells. *Biochemical and Biophysical Research Communications*, *200*(3), 1252–1256. https://doi.org/10.1006/bbrc.1994.1585

Bain, G., Ray, W. J., Yao, M., & Gottlieb, D. I. (1996). Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture. *Biochemical and Biophysical Research Communications*, *223*(3), 691–694. https://doi.org/10.1006/bbrc.1996.0957

Balmer, J. E., & Blomhoff, R.. (2002). Gene expression regulation by retinoic acid. *Journal of Lipid Research*, *43*(11), 1773–1808. https://doi.org/10.1194/jlr.R100015-JLR200

Beddington, R. S. P., Rashbass, P., & Wilson, V. (1993). Brachyury- A gene affecting mouse gastrulation and early organogenesis. *Journal of Neuroscience*, *13*(6), 157–165.

Ben-Shushan, E., Sharir, H., Pikarsky, E., & Bergman, Y. (1995). A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. *Molecular and Cellular Biology*, *15*(2), 1034–1048. https://doi.org/10.1128/mcb.15.2.1034

Beppu, H. et al. (2000). BMP Type II Receptor Is Required for Gastrulation and Early Development of Mouse Embryos. *Developmental Biology*, *221*(1), 249–258. https://doi.org/10.1006/dbio.2000.9670

Berg, R. W., & McBurney, M. W. (1990). Cell density and cell cycle effects on retinoic acid-induced embryonal carcinoma cell differentiation. *Developmental Biology*, *138*(1), 123–135. https://doi.org/10.1016/0012-1606(90)90182-I

Bibel, M., Richter, J., Lacroix, E., & Barde, Y. A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols*, *2*(5), 1034–1043. https://doi.org/10.1038/nprot.2007.147

Billings, S. E. et al. (2013). The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. *The FASEB Journal*, *27*(12), 4877–4889. https://doi.org/10.1096/fj.13-227967

Boroviak, T. et al. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Developmental Cell*, *35*(3), 366–382. https://doi.org/10.1016/j.devcel.2015.10.011

Borowiak, Malgorzata et al. (2009). Small Molecules Efficiently Direct Endodermal Differentiation of Mouse and Human Embryonic Stem Cells. *Cell Stem Cell*, *4*(4), 348–358. https://doi.org/10.1016/j.stem.2009.01.014

Böttcher, R. T., & Niehrs, C. (2005). Fibroblast growth factor signaling during early vertebrate development. In *Endocrine Reviews* (Vol. 26, Issue 1, pp. 63–77). https://doi.org/10.1210/er.2003-0040

Bourillot, P. Y. et al. (2009). Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. *Stem Cells*, *27*(8), 1760–1771. https://doi.org/10.1002/stem.110

Bouwmeester, T., Kim, S. H., Sasai, Y., Lu, B., & De Robertis, E. M. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature*, *382*(6592), 595–601. https://doi.org/10.1038/382595a0

Brennan, J. et al. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature*, *411*(6840), 965–969. https://doi.org/10.1038/35082103

Brons, I. G. M. et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, *448*(7150), 191–195. https://doi.org/10.1038/nature05950

Cai, C., & Grabel, L. (2007). Directing the differentiation of embryonic stem cells to neural stem cells. *Developmental Dynamics*, *236*(12), 3255–3266. https://doi.org/10.1002/dvdy.21306

Camus, A., Perea-Gomez, A., Moreau, A., & Collignon, J. (2006). Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Developmental Biology*, *295*(2), 743–755. https://doi.org/10.1016/j.ydbio.2006.03.047

Chambers, I. et al. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, *113*(5), 643–655. https://doi.org/10.1016/S0092-8674(03)00392-1

Chambers, I. et al. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature*, *450*(7173), 1230–1234. https://doi.org/10.1038/nature06403

Chambers, S. M. et al. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*, *27*(3), 275–280. https://doi.org/10.1038/nbt.1529

Chen, X. et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*, *133*(6), 1106–1117. https://doi.org/10.1016/j.cell.2008.04.043

Chesley, P. (1935). Development of the short-tailed mutant in the house mouse. *Journal of Experimental Zoology*, *70*(3), 429–459. https://doi.org/10.1002/jez.1400700306

Chng, Z., Teo, A., Pedersen, R. A., & Vallier, L. (2010). SIP1 Mediates Cell-Fate Decisions between Neuroectoderm and Mesendoderm in Human Pluripotent Stem Cells. *Cell Stem Cell*, *6*(1), 59–70. https://doi.org/10.1016/j.stem.2009.11.015

Choi, J. et al. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature*, *548*(7666), 219–223. https://doi.org/10.1038/nature23274

Clagett-Dame, M., & DeLuca, H. F. (2002). The role of Vitamin A in mammalian reproduction and embryonic development . *Annual Review of Nutrition*, *22*(1), 347–381. https://doi.org/10.1146/annurev.nutr.22.010402.102745e

Cockburn, K., & Rossant, J. (2010). Making the blastocyst: lessons from the mouse. *The Journal of Clinical Investigation*, *120*(4). https://doi.org/10.1172/JCI41229

Collignon, J., Varlet, I., & Robertson, E. J. (1996). Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature*, *381*(6578), 155–158. https://doi.org/10.1038/381155a0

Cong, L. et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Conlon, F. L. et al. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development*, *120*(7), 1919–1928. https://doi.org/10.1016/0168-9525(94)90026-4

Cunningham, T. J., & Duester, G. (2015). Mechanisms of retinoic acid signalling and its roles in organ and limb development. In *Nature Reviews Molecular Cell Biology* (Vol. 16, Issue 2, pp. 110–123). https://doi.org/10.1038/nrm3932

De Robertis, E. M., Larraín, J., Oelgeschläger, M., & Wessely, O. (2000). The establishment of Spemann's organizer and patterning of the vertebrate embryo. *Nature Reviews Genetics*, *1*(3), 171–181. https://doi.org/10.1038/35042039

Del Barco Barrantes, I., Davidson, G., Gröne, H. J., Westphal, H., & Niehrs, C. (2003). Dkk1 and noggin cooperate in mammalian head induction. *Genes and Development*, *17*(18), 2239–2244. https://doi.org/10.1101/gad.269103

Del Corral, R. D. et al. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron*, *40*(1), 65–79. https://doi.org/10.1016/S0896-6273(03)00565-8

Deng, C. X. et al. (1994). Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes and Development*, *8*(24), 3045–3057. https://doi.org/10.1101/gad.8.24.3045

Derynck, R., & Zhang, Y. E. (2003). Smad-dependent and Smad-independent pathways in TGF-β family signalling. In *Nature* (Vol. 425, Issue 6958, pp. 577–584). Nature Publishing Group. https://doi.org/10.1038/nature02006

Di-Gregorio, A. et al. (2007). BMP signalling inhibits premature neural differentiation in the mouse embryo. *Development*, *134*(18), 3359–3369. https://doi.org/10.1242/dev.005967

Ding, J. et al. (1998). Cripto is required for correct orientation of the anterior–posterior axis in the mouse embryo. *Nature*, *395*(6703), 702–707. https://doi.org/10.1038/27215

Dollé, P. (2009). Developmental expression of retinoic acid receptors (RARs). *Nuclear Receptor Signaling*, *7*(1), nrs.07006. https://doi.org/10.1621/nrs.07006

Dorey, K. & Amaya, E. (2010). FGF signalling: Diverse roles during early vertebrate embryogenesis. In *Development* (Vol. 137, Issue 22, pp. 3731–3742). https://doi.org/10.1242/dev.037689

Duester, G. (2008). Retinoic Acid Synthesis and Signaling during Early Organogenesis. *Cell*, *134*(6),921–931. https://doi.org/10.1016/j.cell.2008.09.002

Dunn, N. R., Vincent, S. D., Oxburgh, L., Robertson, E. J., & Bikoff, E. K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development*, *131*(8), 1717–1728. https://doi.org/10.1242/dev.01072

Dunn, S. J., Martello, G., Yordanov, B., Emmott, S., & Smith, A. G. (2014). Defining an essential transcription factor program for naïve pluripotency. *Science*, *344*(6188), 1156–1160. https://doi.org/10.1126/science.1248882

Eivers, E., Fuentealba, L. C., & De Robertis, EM. (2008). Integrating positional information at the level of Smad1/5/8. *Current Opinion in Genetics & Development*, *18*(4), 304–310. https://doi.org/10.1016/j.gde.2008.06.001

Ema, M., Takahashi, S., & Rossant, J. (2006). Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. *Blood*, *107*(1), 111–117. https://doi.org/10.1182/blood-2005-05-1970

Engberg, N., Kahn, M., Petersen, D. R., Hansson, M., & Serup, P. (2010). Retinoic Acid Synthesis Promotes Development of Neural Progenitors from Mouse Embryonic Stem Cells by Suppressing Endogenous, Wnt-Dependent Nodal Signaling. *STEM CELLS*, *28*(9), 1498–1509. https://doi.org/10.1002/stem.479

Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, *292*(5819), 154–156. http://www.ncbi.nlm.nih.gov/pubmed/7242681

Fehling, H. J. et al. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development*, *130*(17), 4217–4227. https://doi.org/10.1242/dev.00589

Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M., & Goldfarb, M. (1995). Requirement of FGF-4 for postimplantatior mouse development. *Science*, *267*(5195), 246–249. https://doi.org/10.1126/science.7809630

Feng, L., Hernandez, R. E., Waxman, J. S., Yelon, D., & Moens, C. B. (2010). Dhrs3a regulates retinoic acid biosynthesis through a feedback inhibition mechanism. *Developmental Biology*, *338*(1), 1–14. https://doi.org/10.1016/j.ydbio.2009.10.029

Fuentealba, L. C. et al. (2007). Integrating Patterning Signals: Wnt/GSK3 Regulates the Duration of the BMP/Smad1 Signal. *Cell*, *131*(5), 980–993. https://doi.org/10.1016/j.cell.2007.09.027

Fujii, H. et al. (1997). Metabolic inactivation of retinoic acid by a novel P450 differentially expressed in developing mouse embryos. *EMBO Journal*, *16*(14), 4163–4173. https://doi.org/10.1093/emboj/16.14.4163

Fujiwara, T., Dehart, D. B., Sulik, K. K., & Hogan, B. L. M. (2002). Distinct requirements for extra-embryonic and embryonic bone morphogenetic protein 4 in the formation of the node and primitive streak and coordination of left-right asymmetry in the mouse. In *Development* (Vol. 129, Issue 20). https://dev.biologists.org/content/develop/129/20/4685.full.pdf

Funa, N. S. et al. (2015). β-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell*, *16*(6), 639–652. https://doi.org/10.1016/j.stem.2015.03.008

Gadue, P., Huber, T. L., Paddison, P. J., & 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. *Proceedings of the National Academy of Sciences*, *103*(45), 16806–16811. https://doi.org/10.1073/pnas.0603916103

Gadue, P., Huber, T. L., Nostro, M. C., Kattman, S., & Keller, G. M. (2005). Germ layer induction from embryonic stem cells. *Experimental Hematology*, *33*(9 SPEC. ISS.), 955–964. https://doi.org/10.1016/j.exphem.2005.06.009

Galceran, J., Hsu, S. C., & Grosschedl, R. (2001). Rescue of a Wnt mutation by an activated form of LEF-1: Regulation of maintenance but not initiation of Brachyury expression. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(15), 8668–8673. https://doi.org/10.1073/pnas.151258098

Gaulden, J., & Reiter, J. F. (2008). Neur-ons and neur-offs: regulators of neural induction in vertebrate embryos and embryonic stem cells. *Human Molecular Genetics*, *17*(R1), R60–R66. https://doi.org/10.1093/hmg/ddn119

Ghyselinck, N. B., & Duester, G. (2019). Retinoic acid signaling pathways. *Development (Cambridge)*, *146*(13). https://doi.org/10.1242/dev.167502

Giannì, M., Bauer, A., Garattini, E., Chambon, P., & Rochette-Egly, C. (2002). Phosphorylation by p38MAPK and recruitment of SUG-1 are required for RA-induced RARγ degradation and transactivation. *EMBO Journal*, *21*(14), 3760–3769. https://doi.org/10.1093/emboj/cdf374

Glinka, A., Wu, W., Onichtchouk, D., Blumenstock, C., & Niehrs, C. (1997). Head induction by simultaneous repression of Bmp and Wnt signalling in Xenopus. *Nature*, *389*(6650), 517–519. https://doi.org/10.1038/39092

Gluecksohn-Schoenheimer, S. (1944). The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick. *Proceedings of the National Academy of Sciences*, *30*(6), 134–140. https://doi.org/10.1073/pnas.30.6.134

Gottlieb, D. I., & Huettner, J. E. (1999). An in vitro pathway from embryonic stem cells to neurons and glia. *Cells Tissues Organs*, *165*(3–4), 165–172. https://doi.org/10.1159/000016696

Gouon-Evans, V. et al. (2006). BMP-4 is required for hepatic specification of mouse

embryonic stem cell-derived definitive endoderm. *Nature Biotechnology*, *24*(11), 1402–1411. https://doi.org/10.1038/nbt1258

Gouti, M. et al. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Developmental Cell*, *41*(3), 243-261.e7. https://doi.org/10.1016/j.devcel.2017.04.002

Grunz, H., & Tacke, L. (1989). Neural differentiation of Xenopus laevis ectoderm takes place after disaggregation and delayed reaggregation without inducer. *Cell Differentiation and Development*. https://doi.org/10.1016/0922-3371(89)90006-3

Guo, G. et al. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development*, *136*(7), 1063–1069. https://doi.org/10.1242/dev.030957

Guo, G., Huang, Y., Humphreys, P., Wang, X., & Smith, A. (2011). A piggybac-based recessive screening method to identify pluripotency regulators. *PLoS ONE*, *6*(4), e18189. https://doi.org/10.1371/journal.pone.0018189

Hackett, J. A., Kobayashi, T., Dietmann, S., & Surani, M. A. (2017). Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. *Stem Cell Reports*, *8*(6), 1645–1658. https://doi.org/10.1016/j.stemcr.2017.05.014

Haegel, H. et al. (1995). Lack of β-catenin affects mouse development at gastrulation. *Development*, *121*(11), 3529–3537. https://dev.biologists.org/content/develop/121/11/3529.full.pdf

Hanna, J. H., Saha, K., & Jaenisch, R. (2010). Pluripotency and cellular reprogramming: Facts, hypotheses, unresolved issues. In *Cell* (Vol. 143, Issue 4, pp. 508–525). Elsevier. https://doi.org/10.1016/j.cell.2010.10.008

Hayashi, K., Lopes, S. M. C., Tang, F., & 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. https://doi.org/10.1016/j.stem.2008.07.027

Heigwer, F., Kerr, G., & Boutros, M. (2014). E-CRISP: Fast CRISPR target site identification. In *Nature Methods* (Vol. 11, Issue 2, pp. 122–123). Nature Publishing Group. https://doi.org/10.1038/nmeth.2812

Hemmati-Brivanlou, A. Kelly, O. G., & Melton, D. A. (1994). Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell*, *77*(2), 283–295. https://doi.org/10.1016/0092-8674(94)90320-4

Henrique, D., Abranches, E., Verrier, L., & Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development (Cambridge)*, *142*(17), 2864–2875. https://doi.org/10.1242/dev.119768

Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L., & Moens, C. B. (2007). Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development*, *134*(1), 177–187. https://doi.org/10.1242/dev.02706

Huelsken, J. et al. (2000). Requirement for β-catenin in anterior-posterior axis formation in mice. *Journal of Cell Biology*, *148*(3), 567–578. https://doi.org/10.1083/jcb.148.3.567

Iulianella, A., Beckett, B., Petkovich, M., & Lohnes, D. (1999). A molecular basis for retinoic acid-induced axial truncation. *Developmental Biology*, *205*(1), 33–48. https://doi.org/10.1006/dbio.1998.9110

James, D., Levine, A. J., Besser, D., & Hemmati-Brivanlou, A. (2005). TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development*, *132*(6), 1273–1282. https://doi.org/10.1242/dev.01706

Kalkan, T. et al. (2017). Tracking the embryonic stem cell transition from ground state

pluripotency. *Development*, *144*(7), 1221–1234. https://doi.org/10.1242/dev.142711

Kalkan, T., & Smith, A.. (2014). Mapping the route from naive pluripotency to lineage specification. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1657), 20130540. https://doi.org/10.1098/rstb.2013.0540

Kalmar, T. et al. (2009). Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biology*, *7*(7), e1000149. https://doi.org/10.1371/journal.pbio.1000149

Kamiya, D. et al. (2011). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature*, *470*(7335), 503–510. https://doi.org/10.1038/nature09726

Kan, L. et al. (2004). Sox1 acts through multiple independent pathways to promote neurogenesis. *Developmental Biology*, *269*(2), 580–594. https://doi.org/10.1016/j.ydbio.2004.02.005

Kastner, P. et al. (1997). Genetic evidence that the retinoid signal is transduced by heterodimeric RXR/RAR functional units during mouse development. *Development*, *124*(2).

Kataoka, H. et al. (1997). Expressions of PDGF receptor alpha, c-Kit and Flk1 genes clustering in mouse chromosome 5 define distinct subsets of nascent mesodermal cells. *Development Growth and Differentiation*, *39*(6), 729–740. https://doi.org/10.1046/j.1440-169X.1997.t01-5-00009.x

Kawaguchi, R. et al. (2007). A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. *Science*, *315*(5813), 820–825. https://doi.org/10.1126/science.1136244

Kelly, O. G., Pinson, K. I., & Skarnes, W. C. (2004). The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. *Development*, *131*(12), 2803–2815. https://doi.org/10.1242/dev.01137

Khokha, M. K., Yeh, J., Grammer, T. C., & Harland, R. M. (2005). Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Developmental Cell*, *8*(3), 401–411. https://doi.org/10.1016/j.devcel.2005.01.013

Kiecker, C., & Lumsden, A. (2005). Compartments and their boundaries in vertebrate brain development. In *Nature Reviews Neuroscience* (Vol. 6, Issue 7, pp. 553–564). Nature Publishing Group. https://doi.org/10.1038/nrn1702

Kim, M. et al. (2009). Regulation of mouse embryonic stem cell neural differentiation by retinoic acid. *Developmental Biology*, *328*(2), 456–471. https://doi.org/10.1016/j.ydbio.2009.02.001

Kimelman, D. (2016). Tales of Tails (and Trunks). Forming the Posterior Body in Vertebrate Embryos. In *Current Topics in Developmental Biology* (Vol. 116, pp. 517–536). Academic Press. https://doi.org/10.1016/bs.ctdb.2015.12.008

Kimura-Yoshida, C. et al. (2005). Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm. *Developmental Cell*, *9*(5), 639–650. https://doi.org/10.1016/j.devcel.2005.09.011

Klingensmith, J., Ang, S. L., Bachiller, D., & Rossant, J. (1999). Neural induction and patterning in the mouse in the absence of the node and its derivatives. *Developmental Biology*, *216*(2), 535–549. https://doi.org/10.1006/dbio.1999.9525

Kojima, Y. et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell*, *14*(1), 107–120. https://doi.org/10.1016/j.stem.2013.09.014

Krezel, W. et al. (1996). RXRγ null mice are apparently normal and compound RXRα+/-/RXRβ-/-/RXRγ-/- mutant mice are viable. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(17), 9010–9014. https://doi.org/10.1073/pnas.93.17.9010

Kubiczkova, L., Sedlarikova, L., Hajek, R., & Sevcikova, S. (2012). TGF-β - an excellent servant but a bad master. In *Journal of Translational Medicine* (Vol. 10, Issue 1). https://doi.org/10.1186/1479-5876-10-183

Kubo, A. et al. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development*, *131*(7), 1651–1662. https://doi.org/10.1242/dev.01044

Kunath, T. et al. (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. https://doi.org/10.1242/dev.02880

Lamb, T. M. et al. (1993). Neural induction by the secreted polypeptide noggin. *Science*, *262*(5134), 713–718. https://doi.org/10.1126/science.8235591

Lanner, F. et al. (2010). Heparan sulfation-dependent fibroblast growth factor signaling maintains embryonic stem cells primed for differentiation in a heterogeneous state. *Stem Cells*, *28*(2), 191–200. https://doi.org/10.1002/stem.265

Lawson, K. A. et al. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes and Development*, *13*(4), 424–436. https://doi.org/10.1101/gad.13.4.424

Lenka, N., & Ramasamy, S. K. (2007). Neural induction from ES cells portrays default commitment but instructive maturation. *PLoS ONE*, *2*(12), e1349. https://doi.org/10.1371/journal.pone.0001349

Leung, C. Y., & Zernicka-Goetz, M. (2015). Mapping the journey from totipotency to lineage specification in the mouse embryo. *Current Opinion in Genetics and Development*, *34*(Figure 1), 71–76. https://doi.org/10.1016/j.gde.2015.08.002

Levine, A. J., & Brivanlou, A. H. (2007). Proposal of a model of mammalian neural induction. In *Developmental Biology* (Vol. 308, Issue 2, pp. 247–256). Academic Press. https://doi.org/10.1016/j.ydbio.2007.05.036

Lewis, S. L. et al. (2008). Dkk1 and Wnt3 interact to control head morphogenesis in the mouse. *Development*, *135*(10), 1791–1801. https://doi.org/10.1242/dev.018853

Lewis, S. L., & Tam, P. P. L. (2006). Definitive endoderm of the mouse embryo: Formation, cell fates, and morphogenetic function. In *Developmental Dynamics* (Vol. 235, Issue 9, pp. 2315–2329). John Wiley & Sons, Ltd. https://doi.org/10.1002/dvdy.20846

Lindsley, R. C., Gill, J. G., Kyba, M., Murphy, T. L., & Murphy, K. M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development*, *133*(19), 3787–3796. https://doi.org/10.1242/dev.02551

Liu, P. et al. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nature Genetics*, *22*(4), 361–365. https://doi.org/10.1038/11932

Loebel, D. A. .., Watson, C. M., De Young, R. A., & Tam, P. P. L. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. In *Developmental Biology* (Vol. 264, Issue 1, pp. 1–14). https://doi.org/10.1016/S0012-1606(03)00390-7

Loh, K. M., & Lim, B. (2011). A Precarious Balance: Pluripotency Factors as Lineage Specifiers. *Cell Stem Cell*, *8*(4), 363–369. https://doi.org/10.1016/j.stem.2011.03.013

Lowe, L. A., Yamada, S., & Kuehn, M. R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. *Development (Cambridge, England)*, *128*(10), 1831–

1843. papers2://publication/uuid/CD62F955-59DA-4024-AE28-73236EA85671

Macarthur, B. D., & Lemischka, I. R. (2013). Statistical mechanics of pluripotency. In *Cell* (Vol. 154, Issue 3, pp. 484–489). Elsevier. https://doi.org/10.1016/j.cell.2013.07.024

Mahony, S. et al. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biology*, *12*(1). https://doi.org/10.1186/gb-2011-12-1-r2

Mark, M., Ghyselinck, N. B., & Chambon, P. (2009). Function of retinoic acid receptors during embryonic development. In *Nuclear receptor signaling* (Vol. 7). SAGE Publications. https://doi.org/10.1621/nrs.07002

Marks, H. et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, *149*(3), 590–604. https://doi.org/10.1016/j.cell.2012.03.026

Marlétaz, F., Holland, L. Z., Laudet, V., & Schubert, M. (2006). Retinoic acid signaling and the evolution of chordates. *International Journal of Biological Sciences*, *2*(2), 38–47. https://doi.org/10.7150/ijbs.2.38

Marshall, H., Morrison, A., Studer, M., Pöpperl, H., & Krumlauf, R. (1996). Retinoids and Hox genes . In *The FASEB Journal* (Vol. 10, Issue 9). https://doi.org/10.1096/fasebj.10.9.8801179

Marson, A. et al. (2008). Wnt Signaling Promotes Reprogramming of Somatic Cells to Pluripotency. In *Cell Stem Cell* (Vol. 3, Issue 2, pp. 132–135). Elsevier. https://doi.org/10.1016/j.stem.2008.06.019

Martello, G., Bertone, P., & Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO Journal*, *32*(19), 2561–2574. https://doi.org/10.1038/emboj.2013.177

Martello, G., & Smith, A. (2014). The Nature of Embryonic Stem Cells. *Annual Review of Cell and Developmental Biology*, *30*(1), 647–675. https://doi.org/10.1146/annurev-cellbio-100913-013116

Martin, B. L., & Kimelman, D. (2008). Regulation of Canonical Wnt Signaling by Brachyury Is Essential for Posterior Mesoderm Formation. *Developmental Cell*, *15*(1), 121–133. https://doi.org/10.1016/j.devcel.2008.04.013

Martin, B. L., & Kimelman, D. (2010). Brachyury establishes the embryonic mesodermal progenitor niche. *Genes & Development*, *24*(24), 2778–2783. https://doi.org/10.1101/gad.1962910

Martin, B. L. L., & Kimelman, D. (2012). Canonical Wnt Signaling Dynamically Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation. *Developmental Cell*, *22*(1), 223–232. https://doi.org/10.1016/j.devcel.2011.11.001

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(12), 7634–7638. http://www.ncbi.nlm.nih.gov/pubmed/6950406

Martyn, I., Kanno, T. Y., Ruzo, A., Siggia, E. D., & Brivanlou, A. H. (2018). Self-organization of a human organizer by combined Wnt and Nodal signaling. *Nature*, *558*(7708), 132–135. https://doi.org/10.1038/s41586-018-0150-y

Massagué, J. (2012). TGFβ signalling in context. In *Nature Reviews Molecular Cell Biology* (Vol. 13, Issue 10, pp. 616–630). Nature Publishing Group. https://doi.org/10.1038/nrm3434

Masui, S. et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology*, *9*(6), 625–635.

https://doi.org/10.1038/ncb1589

Matsuda, T. et al. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO Journal*, *18*(15), 4261–4269. https://doi.org/10.1093/emboj/18.15.4261

Mendoza-Parra, M. A., Walia, M., Sankar, M., & Gronemeyer, H. (2011). Dissecting the retinoid-induced differentiation of F9 embryonal stem cells by integrative genomics. *Molecular Systems Biology*, *7*(1), 538. https://doi.org/10.1038/msb.2011.73

Mic, F. A., Molotkov, A., Benbrook, D. M., & Duester, G. (2003). Retinoid activation of retinoic acid receptor but not retinoid X receptor is sufficient to rescue lethal defect in retinoic acid synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12), 7135–7140. https://doi.org/10.1073/pnas.1231422100

Mishina, Y., Suzuki, A., Ueno, N., & Behringer, R. R. (1995). Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes and Development*, *9*(24), 3027–3037. https://doi.org/10.1101/gad.9.24.3027

Molotkov, A., Deltour, L., Foglio, M. H., Cuenca, A. E., & Duester, G. (2002). Distinct retinoid metabolic functions for alcohol dehydrogenase genes Adh1 and Adh4 in protection against vitamin A toxicity or deficiency revealed in double null mutant mice. *Journal of Biological Chemistry*, *277*(16), 13804–13811. https://doi.org/10.1074/jbc.M112039200

Molotkov, A., Molotkova, N., & Duester, G. (2006). Retinoic acid guides eye morphogenetic movements via paracrine signaling but is unnecessary for retinal dorsoventral patterning. *Development*, *133*(10), 1901–1910. https://doi.org/10.1242/dev.02328

Molotkova, N., Molotkov, A., & Duester, G. (2007). Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone. *Developmental Biology*, *303*(2), 601–610. https://doi.org/10.1016/j.ydbio.2006.11.035

Molotkova, N., Molotkov, A., Sirbu, I. O., & Duester, G. (2005). Requirement of mesodermal retinoic acid generated by Raldh2 for posterior neural transformation. *Mechanisms of Development*, *122*(2), 145–155. https://doi.org/10.1016/j.mod.2004.10.008

Moutier, E. et al. (2012). Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *The Journal of Biological Chemistry*, *287*(31), 26328–26341. https://doi.org/10.1074/jbc.M112.361790

Mukhopadhyay, M. et al. (2001). Dickkopf1 Is Required for Embryonic Head Induction and Limb Morphogenesis in the Mouse. *Developmental Cell*, *1*(3), 423–434. https://doi.org/10.1016/S1534-5807(01)00041-7

Murry, C. E., & Keller, G. (2008). Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell*, *132*(4), 661–680. https://doi.org/10.1016/j.cell.2008.02.008

Naito, A. T. et al. (2006). Developmental stage-specific biphasic roles of Wnt/β-catenin signaling in cardiomyogenesis and hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(52), 19812–19817. https://doi.org/10.1073/pnas.0605768103

Najm, F. J. et al. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell*, *8*(3), 318–325. https://doi.org/10.1016/j.stem.2011.01.016

Ng, E. S. et al. (2005). The primitive streak gene Mixl1 is required for efficient haematopoiesis and BMP4-induced ventral mesoderm patterning in differentiating ES cells.

Development, *132*(5), 873–884. https://doi.org/10.1242/dev.01657

Nichols, J. et al. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, *95*(3), 379–391. https://doi.org/10.1016/S0092-8674(00)81769-9

Nichols, J., & Smith, A. (2009). Naive and Primed Pluripotent States. *Cell Stem Cell*, *4*(6), 487–492. https://doi.org/10.1016/j.stem.2009.05.015

Nichols, J., & Smith, A. (2012). Pluripotency in the Embryo and in Culture. *Cold Spring Harbor Perspectives in Biology*, *4*(8), a008128–a008128. https://doi.org/10.1101/cshperspect.a008128

Niederreither, K. et al. (2002). *Genetic evidence that oxidative derivatives of retinoic acid are not involved in retinoid signaling during mouse development*. https://doi.org/10.1038/ng876

Niederreither, K., Subbarayan, V., Dollé, P., & Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature Genetics*, *21*(4), 444–448. https://doi.org/10.1038/7788

Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P., & Dollé, P. (2000). Retinoic acid synthesis and hindbrain patterning in the mouse embryo. *Development*, *127*(1), 75–85.

Nieuwkoop, P. D. (1952). Activation and organization of the central nervous system in amphibians. Part II. Differentiation and organization. *Journal of Experimental Zoology*, *120*(1), 33–81. https://doi.org/10.1002/jez.1401200103

Nishikawa, S. I., Jakt, L. M., & Era, T. (2007). Embryonic stem-cell culture as a tool for developmental cell biology. In *Nature Reviews Molecular Cell Biology* (Vol. 8, Issue 6, pp. 502–507). https://doi.org/10.1038/nrm2189

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

Niwa, H., Miyazaki, J. I., & Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, *24*(4), 372–376. https://doi.org/10.1038/74199

Niwa, H., Ogawa, K., Shimosato, D., & Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature*, *460*(7251), 118–122. https://doi.org/10.1038/nature08113

Nostro, M. C., Cheng, X., Keller, G. M., & Gadue, P. (2008). Wnt, Activin, and BMP Signaling Regulate Distinct Stages in the Developmental Pathway from Embryonic Stem Cells to Blood. *Cell Stem Cell*, *2*(1), 60–71. https://doi.org/10.1016/j.stem.2007.10.011

Nusse, R. (2008). Wnt signaling and stem cell control. In *Cell Research* (Vol. 18, Issue 5, pp. 523–527). https://doi.org/10.1038/cr.2008.47

Okada, Y., Shimazaki, T., Sobue, G., & Okano, H. (2004). Retinoic-acid-concentration-dependent acquisition of neural cell identity during in vitro differentiation of mouse embryonic stem cells. *Developmental Biology*, *275*(1), 124–142. https://doi.org/10.1016/j.ydbio.2004.07.038

Okamoto, K. et al. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell*, *60*(3), 461–472. https://doi.org/10.1016/0092-8674(90)90597-8

Olivera-Martinez, I., Harada, H., Halley, P. A., & Storey, K. G. (2012). Loss of FGF-Dependent Mesoderm Identity and Rise of Endogenous Retinoid Signalling Determine Cessation of Body Axis Elongation. *PLoS Biology*, *10*(10), e1001415.

https://doi.org/10.1371/journal.pbio.1001415

Otero, J. J., Fu, W., Kan, L., Cuadra, A. E., & Kessler, J. A. (2004). β-catenin signaling is required for neural differentiation of embryonic stem cells. *Development*, *131*(15), 3545–3557. https://doi.org/10.1242/dev.01218

Ozair, M. Z., Kintner, C., & Brivanlou, A. H. (2013). Neural induction and early patterning in vertebrates. In *Wiley Interdisciplinary Reviews: Developmental Biology* (Vol. 2, Issue 4, pp. 479–498). John Wiley & Sons, Ltd. https://doi.org/10.1002/wdev.90

Park, C. et al. (2004). A hierarchical order of factors in the generation of FLK1- and SCL-expressing hematopoietic and endothelial progenitors from embryonic stem cells. *Development*, *131*(11), 2749–2762. https://doi.org/10.1242/dev.01130

Peng, G. et al. (2019). Molecular architecture of lineage allocation and tissue organization in early mouse embryo. In *Nature* (Vol. 572, Issue 7770, pp. 528–532). Nature Publishing Group. https://doi.org/10.1038/s41586-019-1469-8

Pennimpede, T. et al. (2010). The role of CYP26 enzymes in defining appropriate retinoic acid exposure during embryogenesis. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, *88*(10), 883–894. https://doi.org/10.1002/bdra.20709

Pera, E. M., Ikeda, A., Eivers, E., & De Robertis, E. M. (2003). Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction. *Genes and Development*, *17*(24), 3023–3028. https://doi.org/10.1101/gad.1153603

Perea-Gomez, A. et al. (2002). Nodal antaginists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Developmental Cell*, *3*(5), 745–756. https://doi.org/10.1016/S1534-5807(02)00321-0

Pereira, L., Yi, F., & Merrill, B. J. (2006). Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal. *Molecular and Cellular Biology*, *26*(20), 7479–7491. https://doi.org/10.1128/mcb.00368-06

Piccolo, S. et al. (1999). The head inducer cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature*, *397*(6721), 707–710. https://doi.org/10.1038/17820

Pijnappel, W. W. M. et al. (1993). The retinoid ligand 4-oxo-retinoic acid is a highly active modulator of positional specification. *Nature*, *366*(6453), 340–344. https://doi.org/10.1038/366340a0

Przyborski, S. A., Morton, I. E., Wood, A., & Andrews, P. W. (2000). Developmental regulation of neurogenesis in the pluripotent human embryonal carcinoma cell line NTERA-2. *European Journal of Neuroscience*, *12*(10), 3521–3528. https://doi.org/10.1046/j.1460-9568.2000.00230.x

Ramkumar, N., & Anderson, K. V. (2011). SnapShot: Mouse primitive streak. In *Cell* (Vol. 146, Issue 3). Cell Press. https://doi.org/10.1016/j.cell.2011.07.028

Rhinn, M. et al. (1998). Sequential roles for Otx2 in visceral endoderm and neuroectoderm for forebrain and midbrain induction and specification. *Development*, *125*(5).

Rhinn, M., & Dolle, P. (2012). Retinoic acid signalling during development. *Development*, *139*(5), 843–858. https://doi.org/10.1242/dev.065938

Ribes, V., Fraulob, V., Petkovich, M., & Dollé, P. (2007). The oxidizing enzyme CYP26a1 tightly regulates the availability of retinoic acid in the gastrulating mouse embryo to ensure proper head development and vasculogenesis. *Developmental Dynamics*, *236*(3), 644–653. https://doi.org/10.1002/dvdy.21057

Ribes, V., Le Roux, I., Rhinn, M., Schuhbaur, B., & Dolle, P. (2009). Early mouse caudal development relies on crosstalk between retinoic acid, Shh and Fgf signalling pathways.

Development, *136*(4), 665–676. https://doi.org/10.1242/dev.016204

Rijli, F. M., Gavalas, A., & Chambon, P. (1998). Segmentation and specification in the branchial region of the head: The role of the Hox selector genes. In *International Journal of Developmental Biology* (Vol. 42, Issue 3, pp. 393–401). UPV/EHU Press. https://doi.org/10.1387/ijdb.9654024

Rivera-Pérez, J. A., & Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. *Developmental Biology*, *288*(2), 363–371. https://doi.org/10.1016/j.ydbio.2005.09.012

Robb, L., & Tam, P. P. L. (2004). Gastrula organiser and embryonic patterning in the mouse. *Seminars in Cell & Developmental Biology*, *15*(5), 543–554. https://doi.org/10.1016/j.semcdb.2004.04.005

Rosa, A., & Brivanlou, A. H. (2011). A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation. *EMBO Journal*, *30*(2), 237–248. https://doi.org/10.1038/emboj.2010.319

Rossant, J., Zirngibl, R., Cado, D., Shago, M., & Giguere, V. (1991). Expression of a retinoic acid response element-hsplacZ transgene defines specific domains of transcriptional activity during mouse embryogenesis. *Genes & Development*, *5*(8), 1333–1344. https://doi.org/10.1101/gad.5.8.1333

Rossant, J. (2008). Stem Cells and Early Lineage Development. In *Cell* (Vol. 132, Issue 4, pp. 527–531). Elsevier. https://doi.org/10.1016/j.cell.2008.01.039

Rossant, J., & Tam, P. P. L. (2017). New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. In *Cell Stem Cell* (Vol. 20, Issue 1, pp. 18–28). https://doi.org/10.1016/j.stem.2016.12.004

Russ, A. P. et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature*, *404*(6773), 95–99. https://doi.org/10.1038/35003601

Rydeen, A. et al. (2015). Excessive feedback of Cyp26a1 promotes cell non-autonomous loss of retinoic acid signaling. *Developmental Biology*, *405*(1), 47–55. https://doi.org/10.1016/j.ydbio.2015.06.008

Sakai, Y. et al. (2001). The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. *Genes and Development*, *15*(2), 213–225. https://doi.org/10.1101/gad.851501

Samarut, E., & Rochette-Egly, C. (2012). Nuclear retinoic acid receptors: Conductors of the retinoic acid symphony during development. *Molecular and Cellular Endocrinology*, *348*(2), 348–360. https://doi.org/10.1016/j.mce.2011.03.025

Sandell, L. L. et al. (2007). RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. *Genes and Development*, *21*(9), 1113–1124. https://doi.org/10.1101/gad.1533407

Sasai, Y. et al. (1994). Xenopus chordin: A novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell*, *79*(5), 779–790. https://doi.org/10.1016/0092-8674(94)90068-X

Sasai, Y., Lu, B., Steinbeisser, H., & De Robertis, E. M. (1995). Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in Xenopus. *Nature*, *376*(6538), 333–336. https://doi.org/10.1038/376333a0

Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H., & Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *The EMBO Journal*, *9*(7), 2185–2195. https://doi.org/10.1002/j.1460-2075.1990.tb07388.x

Schubert, F. R., Fainsod, A., Gruenbaum, Y., & Gruss, P. (1995). Expression of the novel murine homeobox gene Sax-1 in the developing nervous system. *Mechanisms of Development*, *51*(1), 99–114. https://doi.org/10.1016/0925-4773(95)00358-8

Semrau, S. et al. (2017). Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01076-4

Silva, J. et al. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biology*, *6*(10), 2237–2247. https://doi.org/10.1371/journal.pbio.0060253

Silva, J. et al. (2009). Nanog Is the Gateway to the Pluripotent Ground State. *Cell*, *138*(4), 722–737. https://doi.org/10.1016/j.cell.2009.07.039

Sirbu, I. O., & Duester, G. (2006). Retinoic-acid signalling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. *Nature Cell Biology*, *8*(3), 271–277. https://doi.org/10.1038/ncb1374

Sirbu, I. O., Gresh, L., Barra, J., & Duester, G. (2005). Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development*, *132*(11), 2611–2622. https://doi.org/10.1242/dev.01845

Sirbu, I. O., Zhao, X., & Duester, G. (2008). Retinoic acid controls heart anteroposterior patterning by down-regulating Isl1 through the Fgf8 pathway. *Developmental Dynamics*, *237*(6), 1627–1635. https://doi.org/10.1002/dvdy.21570

Sladitschek, H. L., & Neveu, P. A. (2015). The bimodally expressed microRNA miR-142 gates exit from pluripotency. *Molecular Systems Biology*, *11*(12), 850–850. https://doi.org/10.15252/msb.20156525

Sladitschek, H. L., & Neveu, P. A. (2015). MXS-chaining: A highly efficient cloning platform for imaging and flow cytometry approaches in mammalian systems. *PLoS ONE*, *10*(4), 1–20. https://doi.org/10.1371/journal.pone.0124958

Sladitschek, H. L., & Neveu, P. A. (2019). A gene regulatory network controls the balance between mesendoderm and ectoderm at pluripotency exit. *Molecular Systems Biology*, *15*(12). https://doi.org/10.15252/msb.20199043

Smith, A. (2013). Nanog heterogeneity: Tilting at windmills? In *Cell Stem Cell* (Vol. 13, Issue 1, pp. 6–7). Elsevier. https://doi.org/10.1016/j.stem.2013.06.016

Smith, J. R. et al. (2008). Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Developmental Biology*, *313*(1), 107–117. https://doi.org/10.1016/j.ydbio.2007.10.003

Smith, W. C., & Harland, R. M. (1992). Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. *Cell*, *70*(5), 829–840. https://doi.org/10.1016/0092-8674(92)90316-5

Smith, W. C., Knecht, A. K., Wu, M., & Harland, R. M. (1993). Secreted noggin protein mimics the Spemann organizer in dorsalizing Xenopus mesoderm. *Nature*, *361*(6412), 547–549. https://doi.org/10.1038/361547a0

Smukler, S. R., Runciman, S. B., Xu, S., & Van Der Kooy, D. (2006). Embryonic stem cells assume a primitive neural stem cell fate in the absence of extrinsic influences. *Journal of Cell Biology*, *172*(1), 79–90. https://doi.org/10.1083/jcb.200508085

Sonntag, K. C. et al. (2005). Context-dependent neuronal differentiation and germ layer induction of Smad4-/- and Cripto-/- embryonic stem cells. *Molecular and Cellular Neuroscience*, *28*(3), 417–429. https://doi.org/10.1016/j.mcn.2004.06.003

Spemann, H., & Mangold, H. (1924). über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Archiv Für Mikroskopische Anatomie Und Entwicklungsmechanik*. https://doi.org/10.1007/BF02108133

Stavridis, M. P., Simon L., J., Collins, B. J., & Storey, K. G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development*, *134*(16), 2889–2894. https://doi.org/10.1242/dev.02858

Steinhart, Z., & Angers, S. (2018). Wnt signaling in development and tissue homeostasis. In *Development (Cambridge, England)* (Vol. 145, Issue 11). https://doi.org/10.1242/dev.146589

Stern, C. D. (2005). Neural induction: old problem, new findings, yet more questions. *Development*, *132*(9), 2007–2021. https://doi.org/10.1242/DEV.01794

Storey, K. G., Crossley, J. M., De Robertis, E. M., Norris, W. E., & Stern, C. D. (1992). Neural induction and regionalisation in the chick embryo. *Development*, *114*(3), 729–741.

Strickland, S., & Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. *Cell*. https://doi.org/10.1016/0092-8674(78)90008-9

Sun, X., Meyers, E. N., Lewandoski, M., & Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes and Development*, *13*(14), 1834–1846. https://doi.org/10.1101/gad.13.14.1834

Tada, S. et al. (2005). Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. *Development*, *132*(19), 4363–4374. https://doi.org/10.1242/dev.02005

Takada, S. et al. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes and Development*, *8*(2), 174–189. https://doi.org/10.1101/gad.8.2.174

Tam, P. P. L., & Loebel, D. A. F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nature Reviews Genetics*, *8*(5), 368–381. https://doi.org/10.1038/nrg2084

Tam, P. P. L., Loebel, D. A. F., & Tanaka, S. S. (2006). Building the mouse gastrula: signals, asymmetry and lineages. *Current Opinion in Genetics and Development*, *16*(4), 419–425. https://doi.org/10.1016/j.gde.2006.06.008

Teo, A. K. K. et al. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes and Development*, *25*(3), 238–250. https://doi.org/10.1101/gad.607311

Tesar, P. J. et al. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, *448*(7150), 196–199. https://doi.org/10.1038/nature05972

Thomas, P., & Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Current Biology*, *6*(11), 1487–1496. https://doi.org/10.1016/S0960-9822(96)00753-1

Thomson, M. et al. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*, *145*(6), 875–889. https://doi.org/10.1016/j.cell.2011.05.017

Tonge, P. D., & Andrews, P. W. (2010). Retinoic acid directs neuronal differentiation of human pluripotent stem cell lines in a non-cell-autonomous manner. *Differentiation*, *80*(1), 20–30. https://doi.org/10.1016/j.diff.2010.04.001

Torres, J., Prieto, J., Durupt, F. C., Broad, S., & Watt, F. M. (2012). Efficient differentiation of Embryonic stem cells into mesodermal precursors by BMP, Retinoic Acid and Notch signalling. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0036405

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., & Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*, *135*(5), 909–918. https://doi.org/10.1242/dev.017400

Tropepe, V. et al. (2001). Direct neural fate specification from embryonic stem cells: A primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron*, *30*(1), 65–78. https://doi.org/10.1016/S0896-6273(01)00263-X

Trott, J., & Martinez Arias, A. (2013). Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. *Biology Open*, *2*(10), 1049–1056. https://doi.org/10.1242/bio.20135934

Tsakiridis, A. et al. (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development (Cambridge)*, *141*(6), 1209–1221. https://doi.org/10.1242/dev.101014

Turner, D. A. et al. (2017). Anteroposterior polarity and elongation in the absence of extraembryonic tissues and of spatially localised signalling in gastruloids: Mammalian embryonic organoids. *Development (Cambridge)*, *144*(21), 3894–3906. https://doi.org/10.1242/dev.150391

Turner, D. A., Rué, P., Mackenzie, J. P., Davies, E., & Martinez Arias, A. (2014). Brachyury cooperates with Wnt/β-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. *BMC Biology*, *12*(1), 63. https://doi.org/10.1186/s12915-014-0063-7

Turner, D. A., Trott, J., Hayward, P., Rué, P., & Martinez Arias, A. (2014). An interplay between extracellular signalling and the dynamics of the exit from pluripotency drives cell fate decisions in mouse ES cells. *Biology Open*, *3*(7), 614–626. https://doi.org/10.1242/bio.20148409

Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V., & Nicolas, J. F. (2009). Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Developmental Cell*, *17*(3), 365–376. https://linkinghub.elsevier.com/retrieve/pii/S1534580709003396

Ueno, S. et al. (2007). Biphasic role for Wnt/β-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(23), 9685–9690. https://doi.org/10.1073/pnas.0702859104

Vallier, L. et al. (2009). 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. https://doi.org/10.1371/journal.pone.0006082

Vallier, L., Alexander, M., & Pedersen, R. A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *Journal of Cell Science*, *118*(19), 4495–4509. https://doi.org/10.1242/jcs.02553

Vallier, L., Reynolds, D., & Pedersen, R. A. (2004). Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Developmental Biology*, *275*(2), 403–421. https://doi.org/10.1016/j.ydbio.2004.08.031

van den Brink, S. C. et al. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature*. https://doi.org/10.1038/s41586-020-2024-3

Van Den Brink, S. C. et al. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development (Cambridge)*, *141*(22), 4231–4242. https://doi.org/10.1242/dev.113001

Varlet, I., Collignon, J., & Robertson, E. J. (1997). Nodal expression in the primitive endoderm

is required for specification of the anterior axis during mouse gastrulation. *Development*, *124*(5), 1033–1044.

Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. Ha., & Moon, R. T. (2003). Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Current Biology*, *13*(8), 680–685. https://doi.org/10.1016/S0960-9822(03)00240-9

Verani, R. et al. (2007). Expression of the Wnt inhibitor Dickkopf-1 is required for the induction of neural markers in mouse embryonic stem cells differentiating in response to retinoic acid. *Journal of Neurochemistry*, *100*(1), 242–250. https://doi.org/10.1111/j.1471-4159.2006.04207.x

Vincent, S. D., Dunn, N. R., Hayashi, S., Norris, D. P., & Robertson, E. J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes and Development*, *17*(13), 1646–1662. https://doi.org/10.1101/gad.1100503

Vonica, A., & Gumbiner, B. M. (2002). Zygotic Wnt activity is required for Brachyury expression in the early Xenopus laevis embryo. *Developmental Biology*, *250*(1), 112–127. https://doi.org/10.1006/dbio.2002.0786

Waldrip, W. R., Bikoff, E. K., Hoodless, P. A., Wrana, J. L., & Robertson, E. J. (1998). Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. *Cell*, *92*(6), 797–808. https://doi.org/10.1016/S0092-8674(00)81407-5

Wang, J., Sinha, T., & Wynshaw-Boris, A. (2012). Wnt signaling in mammalian development: Lessons from mouse genetics. *Cold Spring Harbor Perspectives in Biology*, *4*(5), 6. https://doi.org/10.1101/cshperspect.a007963

Weston, A. D., Blumberg, B., & Underhill, T. M. (2003). Active repression by unliganded retinoid receptors in development: Less is sometimes more. *Journal of Cell Biology*, *161*(2), 223–228. https://doi.org/10.1083/jcb.200211117

Wilkinson, D. G., Bhatt, S., & Herrmann, B. G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature*, *343*(6259), 657–659. https://doi.org/10.1038/343657a0

Wilson, J. G., Roth, C. B., & Warkany, J. (1953). An analysis of the syndrome of malformations induced by maternal vitamin a deficiency. Effects of restoration of vitamin a at various times during gestation. *American Journal of Anatomy*, *92*(2), 189–217. https://doi.org/10.1002/aja.1000920202

Wilson, P. A., & Hemmati-Brivanlou, A. (1995). Induction of epidermis and inhibition of neural fate by Bmp-4. *Nature*, *376*(6538), 331–333. https://doi.org/10.1038/376331a0

Winnier, G., Blessing, M., Labosky, P. A., & Hogan, B. L. M. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes and Development*, *9*(17), 2105–2116. https://doi.org/10.1101/gad.9.17.2105

Wood, H. B., & Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mechanisms of Development*, *86*(1–2), 197–201. https://doi.org/10.1016/S0925-4773(99)00116-1

Yamaguchi, T P, Takada, S., Yoshikawa, Y., Wu, N., & McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes & Development*, *13*(24), 3185–3190. https://doi.org/10.1101/gad.13.24.3185

Yamaguchi, T. P., Conlon, R. A., & Rossant, J. (1992). Expression of the fibroblast growth factor receptor FGFR-1/flg during gastrulation and segmentation in the mouse embryo. *Developmental Biology*, *152*(1), 75–88. https://doi.org/10.1016/0012-1606(92)90157-C

Yamaguchi, T. P, Harpal, K., Henkemeyer, M., & Rossant, J. (1994). fgfr-1 is required for

embryonic growth and mesodermal patterning during mouse gastrulation. *Genes and Development*, *8*(24), 3032–3044. https://doi.org/10.1101/gad.8.24.3032

Yamamoto, M. et al. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature*, *428*(6981), 387–392. https://doi.org/10.1038/nature02418

Yang, J. et al. (2010). Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell*, *7*(3), 319–328. https://doi.org/10.1016/j.stem.2010.06.022

Yang, Y. P., & Klingensmith, J. (2006). Roles of organizer factors and BMP antagonism in mammalian forebrain establishment. *Developmental Biology*, *296*(2), 458–475. https://doi.org/10.1016/j.ydbio.2006.06.014

Yasunaga, M. et al. (2005). Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nature Biotechnology*, *23*(12), 1542–1550. https://doi.org/10.1038/nbt1167

Ye, S., Li, P., Tong, C., & Ying, Q. L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO Journal*, *32*(19), 2548–2560. https://doi.org/10.1038/emboj.2013.175

Yi, F., Pereira, L., & Merrill, B. J. (2008). Tcf3 Functions as a Steady-State Limiter of Transcriptional Programs of Mouse Embryonic Stem Cell Self-Renewal. *Stem Cells*, *26*(8), 1951–1960. https://doi.org/10.1634/stemcells.2008-0229

Ying, Q. L., Nichols, J., Chambers, I., & Smith, A. (2003). BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. *Cell*, *115*(3), 281–292. https://doi.org/10.1016/S0092-8674(03)00847-X

Ying, Q. L. et al. (2008). The ground state of embryonic stem cell self-renewal. *Nature*, *453*(7194), 519–523. https://doi.org/10.1038/nature06968

Ying, Q. L., & Smith, A. G. (2003). Defined Conditions for Neural Commitment and Differentiation. *Methods in Enzymology*, *365*, 327–341. https://doi.org/10.1016/S0076-6879(03)65023-8

Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., & Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnology*, *21*(2), 183–186. https://doi.org/10.1038/nbt780

Yoshikawa, Y., Fujimori, T., McMahon, A. P., & 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. https://doi.org/10.1006/dbio.1997.8502

Young, R. A. (2011). Control of the embryonic stem cell state. *Cell*, *144*(6), 940–954. https://doi.org/10.1016/j.cell.2011.01.032

Zhang, M., Chen, W., Smith, S. M., & Napoli, J. L. (2001). Molecular Characterization of a Mouse Short Chain Dehydrogenase/Reductase Active with All-trans-retinol in Intact Cells, mRDH1. *Journal of Biological Chemistry*, *276*(47), 44083–44090. https://doi.org/10.1074/jbc.M105748200

Zhong, Z., Wen, Z., & Darnell, J. E. (1994). Stat3: A STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. *Science*, *264*(5155), 95–98. https://doi.org/10.1126/science.8140422

Zhu, J. et al. (1999). Retinoic acid induces proteasome-dependent degradation of retinoic acid receptor α (RARα) and oncogenic RARα fusion proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(26), 14807–14812. https://doi.org/10.1073/pnas.96.26.14807