CHAPTER 1

1 INTRODUCTION

1.1 Totipotency and preimplantation mouse embryo development

Waddington's epigenetic landscape model conceptualises development of a cell towards terminal differentiation as a ball descending along a series of branching valleys and ridges (Waddington 1940). During its path down the mountain, the variations in the landscape limit the number of trails the ball can ultimately follow. Under normal conditions, once in its final valley, the ball cannot easily cross to neighbouring valleys (trans-differentiation) or climb back to preceding states (reprogramming). In this metaphor, the ball at the top of the mountain represents a totipotent cell and development corresponds to the series of progressive fate restrictions along the descending valleys, that result in a complete organism.

Mouse preimplantation development comprises the stages from fertilization to uterine implantation, at embryonic day 4.5 (Figure 1.1).

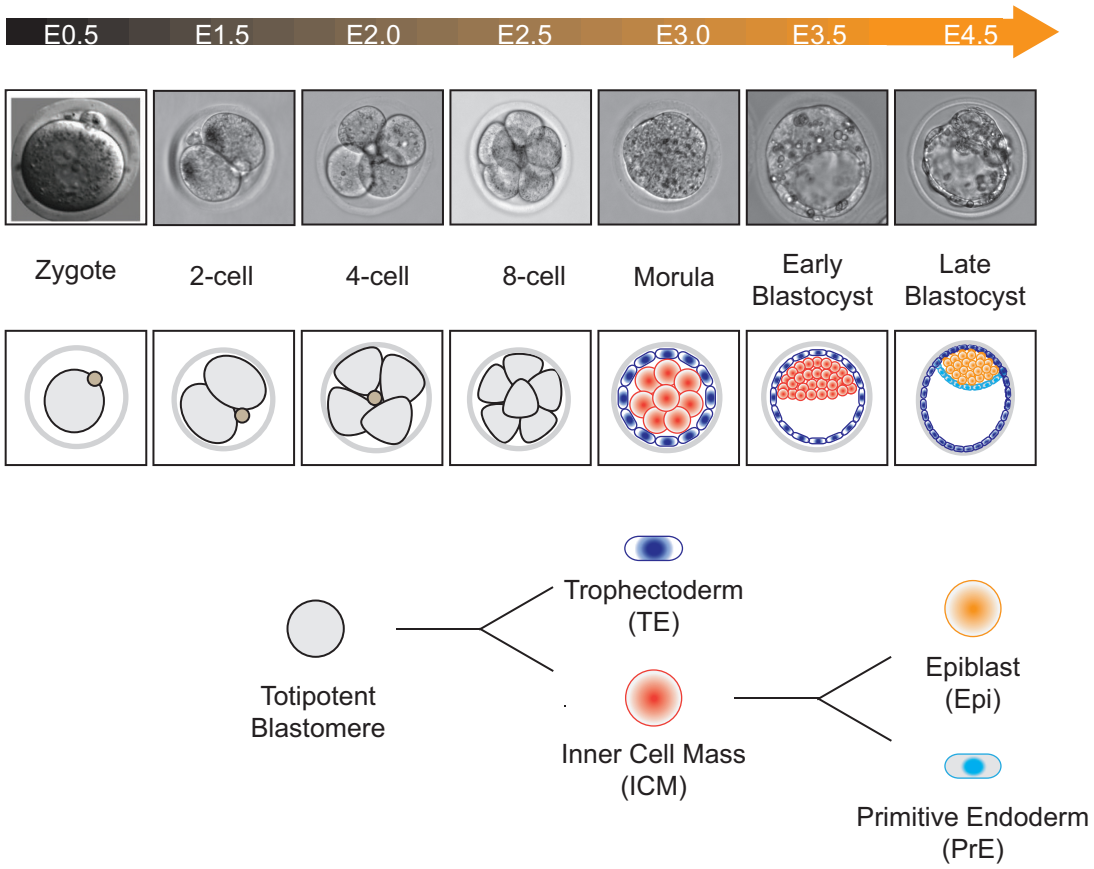

Figure 1.1 | Preimplantation mouse embryo development. Upon fertilisation, fusion of the egg and sperm results in formation of the totipotent zygote. The zygote undergoes a series of symmetric cleavage divisions, sequentially producing 2-cell, 4-cell and 8-cell embryos. After the 8-cell stage, the first cell lineage decision takes place and outer cells in the morula become biased towards forming the trophectoderm (TE) whereas inside cells are biased to form inner cell mass (ICM). As the blastocyst progresses through early, mid and late stages, ICM cells undergo the second lineage decision and further specify into either epiblast (Epi) or primitive endoderm (PrE). Adapted from Niakan et al (Niakan, Schrode et al. 2013).

1.1.1 Totipotency and epigenetic dynamics in early preimplantation

Mammalian embryonic development starts with the fusion of two highly specialised and haploid cells, the sperm and egg, producing the fertilized 1-cell diploid zygote. During preimplantation development, the zygote follows serial divisions in which the cells progressively lose potency and acquire fate specialization. A stringent definition of totipotency is the ability of a single-cell to develop to a full organism on its own (Edwards and Beard 1997) and this way only the mouse zygote and 2-cell blastomeres fit the

requirements (Tarkowski 1959). Considering a less stringent definition, individual mouse blastomeres of up to the 8-cell stage embryo are considered totipotent (Ishiuchi and Torres-Padilla 2013) in the sense that each can contribute to all cell lineages, embryonic and extra-embryonic, when supported by carrier blastomeres (Rossant 1976, Kelly 1977, Balakier and Pedersen 1982).

Totipotency of early embryos is specified by unique epigenetic status including histone variants, histone modifications and patterns of DNA hypomethylation (Zhou and Dean 2015). A complex epigenetic reprogramming is triggered upon fertilization and is believed to be the basis of reacquisition of totipotency. Immediately after fertilization, the paternal genome is subject to active demethylation and hydroxymethylation, while the maternal genome is exposed to slower demethylation (Santos, Hendrich et al. 2002, Iqbal, Jin et al. 2011, Wossidlo, Nakamura et al. 2011), in order to ensure embryonic gene activation (EGA), also referred to as zygotic gene activation (ZGA).

Fertilization also induces the paternal genome to exchange protamines for maternal histones and acquire histone modifications, while the maternal genome appears epigenetically stable (Zhou and Dean 2015). This histone organisation generates an epigenetic asymmetry between paternal and maternal genomes that is correlated with asymmetric transcriptional activity and may account for the different reprogramming capacities of the two pronuclei (Liu, Yin et al. 2014). In cleavage stage mouse embryos, this epigenetic asymmetry is associated with cell fate decision and can be detected by the 4-cell stage (VerMilyea, Maneck et al. 2011). Consistent with the key roles for preimplantation epigenome remodelling, the spatiotemporal timeline of preimplantation development can be reconstructed exclusively based on the expression patterns of chromatin modifiers (Burton, Muller et al. 2013).

Transcriptionally, until the 8-cell stage of mouse development, blastomeres co-express determinants of all three main lineages – TE, Epi and PrE (Dietrich and Hiiragi 2007). On the basis of lineage segregation, totipotent cells would be the ones that co-express embryonic and extraembryonic determinants (Morgani and Brickman 2014), such as *Pou5f1 (Oct4)*, *Gata6*, *Nanog* and *Cdx2* (Dietrich and Hiiragi 2007, Plusa, Piliszek et al. 2008).

There is much to be learned regarding the complex process of acquisition of totipotency, such as defining key reprogramming factors in the nucleus and cytoplasm required for totipotency or determining the triggering events for activation of totipotent genome/epigenome (Zhou and Dean 2015). Recently, there were several reports of totipotent-like subpopulations within ESCs (reviewed in Baker *et al* (Baker and Pera 2018)). Moreover, two groups reported establishment of stable stem cell *in vitro* cultures

with expanded developmental potential, defined as the capacity to contribute to both embryonic and extraembryonic tissues (Yang, Ryan et al. 2017, Yang, Liu et al. 2017). These represent exciting new model systems for future biochemical and molecular interrogations of the totipotent-like state.

1.1.2 The first cell lineage decision: inner cell mass versus trophectoderm specification

1.1.2.1 Classical models for the first cell lineage specification

Following the developmental course, the 8-cell embryo undergoes compaction, blastomeres acquire apical polarity and subsequent asymmetric divisions result in outer polar and inner apolar cells (Zernicka-Goetz 2005). By the 16-cell stage, outer and inner cells are biased towards extraembryonic trophectoderm (TE) and inner cell mass (ICM), respectively, and these two lineages become fully segregated by the blastocyst stage (Morgani and Brickman 2014). The molecular mechanisms behind this first cell lineage specification remain elusive and the exact timing for lineage bias is uncertain. Nevertheless, there are two main theories proposed to explain this process: the inside/outside model, and the cell polarity model (Figure 1.2).

The observation that cells in the early embryo can switch cell fate when moved to a new position (Mintz 1965) led to the idea that cell positioning is the key determinant for ICM or TE cell fate. This was further supported by the fact that in pre-compaction morula, outside blastomeres had a preference to colonise the TE layer of the embryo, whereas inside blastomeres were biased towards ICM (Tarkowski and Wroblewska 1967). These observations led to the classical inside/outside model (Figure 1.2A).

Blastomeres become polarised by the 8-cell stage, so that the outward facing region (apical) become distinct from the inward-facing region (basolateral) (Ziomek and Johnson 1980, Johnson and Ziomek 1981). The apical domain is established and enriched for microvilli, F-actin and an apical protein complex composed of PAR3, PAR6 and atypical protein kinase C (aPKC) (Yamanaka, Ralston et al. 2006). The cell polarity model was proposed to address how inside/outside differences are established (Johnson and Ziomek 1981) (Figure 1.2B). It assumes that unequal distribution of polarity in daughter cells resulting from the cleavage division following the 8-cell stage determines the inner and outer positioning and thus ICM and TE fate. If the cleavage occurs parallel to the polarity axis, this produces a symmetric division where the two

daughter cells inherit equal polarity. If the cleavage occurs perpendicular to the polarity axis, then division is asymmetric and produces one polar and one apolar daughter cell. Polar cells position in the outer layer and are biased towards TE whereas apolar cells position inwards and will form ICM. The unequal distribution of polarity proteins likely results in asymmetric distribution of RNA, histones and other signalling molecules. For instance, mRNA coding for the TE lineage specifier Cdx2, localises in the apical region of the 8- and 16-cell stage blastomeres (Jedrusik, Parfitt et al. 2008). Consistently, polar cells are more likely to inherit this transcription factor and activate the TE transcriptional programme (Skamagki, Wicher et al. 2013). Additionally, it has been shown that Cdx2 can regulate cell polarity in such a way that its depletion leads to downregulation and mislocalisation of aPKC polarity protein, whereas Cdx2 overexpression results in aPKC upregulation and increased TE cell numbers (Jedrusik, Parfitt et al. 2008).

These two classical models provide conceptual guidance to explain the first cell lineage decision. However, it has been shown that outer morula cells can overwrite this positional information when aggregated with 8-cell embryos and contribute to ICM (Rossant and Vijh 1980). This is an example of high cellular plasticity in preimplantation development with some rare cells are still capable to transition between lineages even after cell fate establishment (reviewed in Morgani et al (Morgani and Brickman 2014)). Moreover, how polarity is first determined and at which precise stage do blastomeres become biased towards one lineage remains unknown.

Studies using Cre-loxP labelling or rainbow reporter mice strategies, suggest that TE/ICM bias might be present as early as the 4-cell stage (Fujimori, Kurotaki et al. 2003, Tabansky, Lenarcic et al. 2013). Furthermore, heterogeneity of different factors at the 4-cell stage has been functionally connected to developmental bias (Torres-Padilla, Parfitt et al. 2007, Plachta, Bollenbach et al. 2011, Burton, Muller et al. 2013).

Levels of Histone H3 arginine 26 methylation (H3R26me) are higher in 4-cell blastomeres that later contribute to either ICM or polar TE, whereas low H3R26me at this early stage seems to result in a bias towards mural TE (Torres-Padilla, Parfitt et al. 2007). By overexpressing CARM1, an H3-specific arginine methyltransferase, in one of the 2-cell blastomeres, it could be demonstrated these would contribute preferentially to ICM (Torres-Padilla, Parfitt et al. 2007). Similarly, PRDM14 is heterogeneously expressed in 4-cell stage blastomeres and its overexpression in one of the 2-cell blastomeres showed this modified cell displayed a preference for ICM integration (Burton, Muller et al. 2013). Additionally, a link between Oct4 kinetics in early blastomeres and TE/ICM cell fate has also been established, such that blastomeres with “slow” or “fast” Oct4 kinetics are biased to ICM and TE, respectively (Plachta, Bollenbach et al. 2011).

Recent reports showing single-cell transcriptomic characteristics of preimplantation embryos have also demonstrated high heterogeneity in early blastomeres (Biase, Cao et al. 2014, Deng, Ramskold et al. 2014, Piras, Tomita et al. 2014, Shi, Chen et al. 2015). These will certainly contribute to future functional elucidation of this early decision.

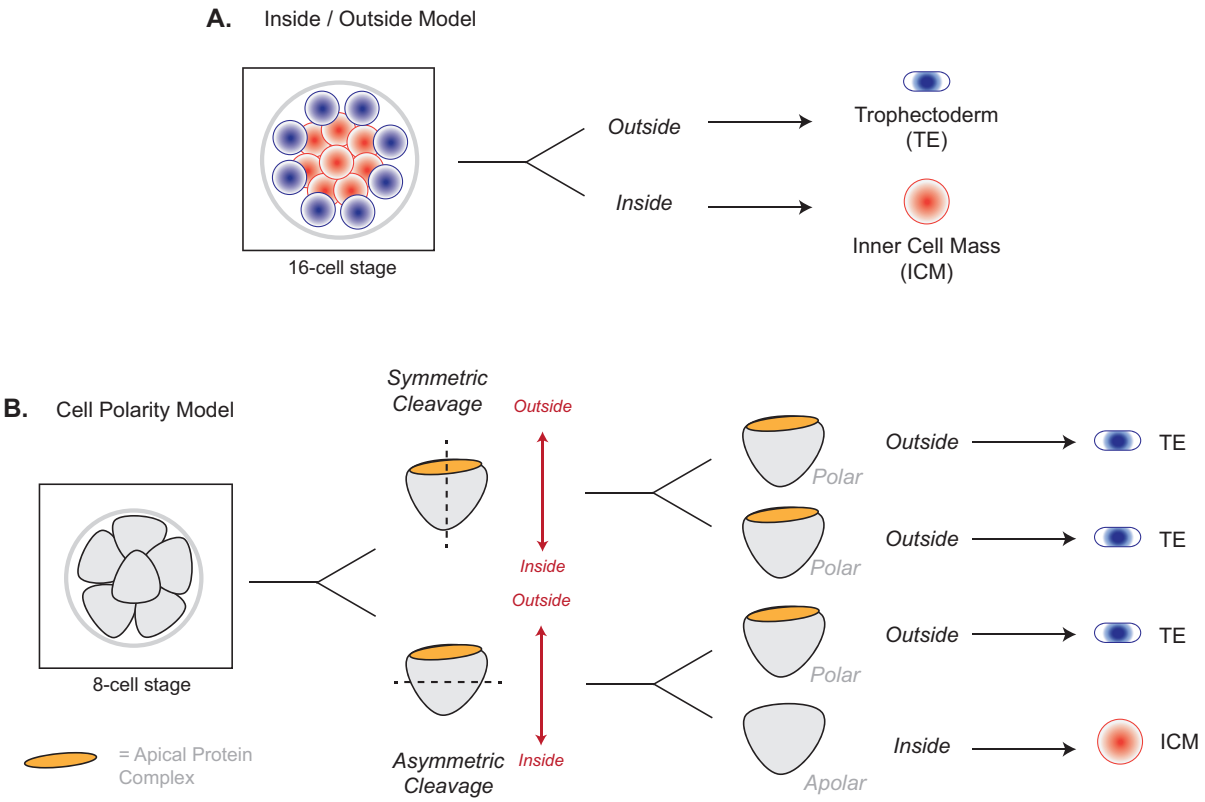

Figure 1.2 | Models for the first cell lineage decision. A | The inside/outside model considers that by the 16-stage, cell position determines fate decision, in such a way that outside cells contribute preferentially to TE and inner cells to ICM (Tarkowski and Wroblewska 1967). **B |** In the cell polarity model, cell polarity and cleavage pattern by the 8-cell stage determine inside and outside positioning by 16-cell and later stages. A symmetric division plane parallel to the polarity axis produces two polar outer daughter cells. In contrast, asymmetric cleavage with a division plane perpendicular to polarity axis results in one polar outer and one apolar inner daughter cell (Johnson and Ziomek 1981).

1.1.2.2 Transcriptional networks associated with TE/ICM fate decision

Cdx2 is a key transcription factor expressed in the TE and is essential for its establishment (Strumpf, Mao et al. 2005). Cdx2 null embryos initiate TE commitment and formation of the blastocoel cavity, but by the late blastocyst stage outer cells fail to

acquire epithelial identity and express reduced levels of TE markers, resulting in a compromised trophoblast compartment (Strumpf, Mao et al. 2005). As a result, mutant blastocysts arrest their development prior to implantation (Chawengsaksophak, James et al. 1997).

Cdx2 zygotic mutants show ectopic expression of *Nanog* and *Oct4* in TE, suggesting *Cdx2* mediates repression of these two pluripotent genes in the TE compartment (Strumpf, Mao et al. 2005). Additionally, Cdx2 and Oct4 directly interact, acting as co-repressors and this way contributing to ICM and TE segregation (Niwa, Toyooka et al. 2005). In fact, overexpression of *Cdx2* in embryonic stem cells can convert them to trophoblast stem cells (Niwa, Toyooka et al. 2005). Conversely, downregulation of *Oct4* produces the same effect (Niwa, Miyazaki et al. 2000). In the TE compartment, *Cdx2* cooperates with *Gata3, Eomes* and *Elf5* to establish the trophoblast transcriptional programme (Russ, Wattler et al. 2000, Ng, Dean et al. 2008, Ralston, Cox et al. 2010) (Figure 1.3B).

Different signalling pathways have been shown to affect the first cell lineage decision. The Hippo signalling pathway is perhaps one of the best characterised players in TE/ICM specification (Figure 1.3A). The transcription factor *Tead4* is essential for TE lineage specification as *Tead4* null mutation are embryonically lethal, with mutant embryos lacking TE formation (Yagi, Kohn et al. 2007). *Tead4* is expressed quite early in development, and was shown to induce *Cdx2* expression, this way initiating TE specification (Yagi, Kohn et al. 2007, Nishioka, Yamamoto et al. 2008). The role of *Tead4* is differentially regulated in inner and outer cells due to differential activity of the Hippo pathway. Tead4 needs its transcriptional co-activators YAP and TAZ (also known as WWTR1) in order to upregulate TE-specific genes (Nishioka, Yamamoto et al. 2008). YAP and TAZ have different subcellular localisation depending on their phosphorylation status. In TE, non-phosphorylated YAP/TAZ can translocate to the nucleus where they associate with Tead4 and activate TE-specific genes such as *Cdx2* and *Gata3* (Nishioka, Inoue et al. 2009, Ralston, Cox et al. 2010). On the other hand, active Hippo signalling in ICM cells results in phosphorylation of YAP/TAZ by LATS kinases (Lats1/2), resulting in cytoplasmic localisation for these co-activators and preventing the activation of TE transcriptional programme in the pluripotent layer (Nishioka, Inoue et al. 2009).

Further studies have identified other regulators of Hippo signalling in ICM and TE cells. Nf2 was shown to affect YAP subcellular localisation as maternal and zygotic *Nf2* mutant embryos displayed YAP and Cdx2 expression in ICM cells (Cockburn, Biechele et al. 2013). The role of Nf2 was specific to ICM cells, where it acts upstream of Lats1/2

to promote YAP phosphorylation and cytoplasmic localisation (Cockburn, Biechele et al. 2013).

RHO-ROCK signalling affects ICM and TE segregation due to an essential role on both Hippo signalling and cell polarity (Kono, Tamashiro et al. 2014). Inhibition of ROCK signalling affected segregation of several apical and basal polarity regulators (Kono, Tamashiro et al. 2014). In addition, in TE cells, ROCK acts upstream of Lats1/2 inhibiting these two kinases and therefore ensuring proper YAP nuclear localisation.

Another connection between Hippo signalling and cell polarity was recently established due to identification of junction-associated scaffolding Angiomotin (AMOT), which is necessary to maintain active Hippo signalling in ICM cells (Hirate, Hirahara et al. 2013, Leung and Zernicka-Goetz 2013). In the apolar ICM cells, phosphorylation of AMOT results in its localisation next to adherens junctions (Hirate, Hirahara et al. 2013). AMOT cooperates with Nf2 and Lats1/2 to promote YAP phosphorylation. In TE polar cells, non-phosphorylated AMOT is restrained at the apical domain, ensuring Hippo signalling remains inactive.

Other signalling pathways have been shown to regulate or be necessary for the first cell lineage specification, although their precise role remains elusive. BMP signalling was shown to be required for proper extraembryonic lineage specification (Graham, Wicher et al. 2014). Notch signalling is differentially expressed between inner and outer cells, in a reverse fashion to Hippo signalling. Notch is active in TE cells where it cooperates with Tead4 to promote Cdx2 expression, whereas it remains inactive in ICM (Rayon, Menchero et al. 2014). Interestingly, overexpression of Notch resulted in inner cells relocating to an outside position and adopting a TE fate (Rayon, Menchero et al. 2014).

The Ras/MAPK pathway has also been implicated in TE/ICM segregation as expression of a constitutively active form of Ras results in Cdx2 upregulation and delays in blastocyst development, in addition to promoting differentiation of ESCs to TSCs (Saba-El-Leil, Vella et al. 2003, Lu, Yabuuchi et al. 2008). Inhibition of Ras/MAPK hampers Cdx2 expression and results in embryonic lethality due to TE defects (Saba-El-Leil, Vella et al. 2003, Lu, Yabuuchi et al. 2008).

Overall, differential signalling pathway activities between inner and outer cells may ensure proper lineage segregation as exemplified by the central role of Hippo signalling. Hippo provides a working model for the integration of cell polarity, cell positioning and transcriptional regulation of ICM/TE specification. Nevertheless, the mechanisms dictating acquisition of polarity and blastomere positioning remain unclear. Furthermore, other signalling pathways independent of Hippo might also play a role in this early event.

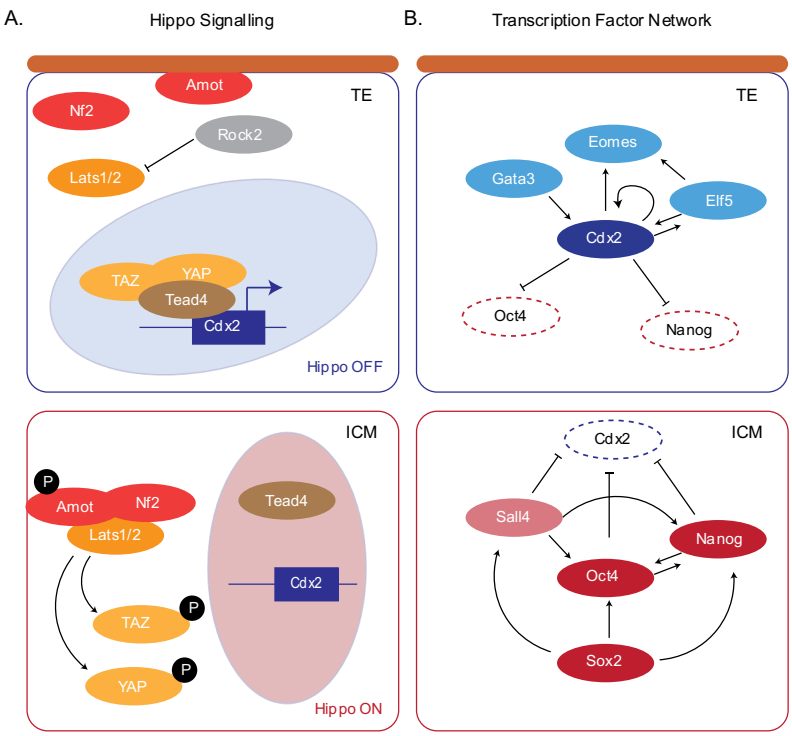

Figure 1.3 | Major signalling and transcriptional networks controlling TE and ICM segregation. A | Hippo Signalling is differentially controlled in outer TE cells versus inner ICM cells. Inactivation of Hippo signalling in outside cells is maintained by AMOT sequestration in the apical domain and inhibition of Lats1/2 kinases by Rock2. This way, YAP and TAZ remain non-phosphorylated and translocate to the nucleus where they associate with Tead4 to promote expression of TE-genes such as Cdx2. In ICM cells, Hippo signalling is active and AMOT, Nf2 and Lats1/2 associate to phosphorylate Tead4 co-activators YAP and TAZ, preventing their translocation to the cell nucleus and consequent Cdx2 expression in the ICM. **B |** Upon fate determination, specific transcriptional networks are established in the ICM and TE to reinforce lineage commitment. In TE cells, Cdx2, Eomes, Gata3 and Elf5 cooperate to maintain TE identity and repress ICM-specific genes. Conversely, in the ICM, Oct4, Nanog, Sox2 and Sall4 cooperate to establish the pluripotency programme while repressing TE-specifiers. Image based on (Oron and Ivanova 2012, Chazaud and Yamanaka 2016).

1.1.3 The second cell lineage decision: Epiblast versus Primitive Endoderm

Shortly after formation of the ICM in the early blastocyst-stage embryo, the second cell lineage specification event begins, resulting in the formation of the epiblast (Epi) and a second extraembryonic layer, the primitive endoderm (PrE) (Figure 1.1). By the late blastocyst at E4.5, cells at the surface of the ICM are in direct contact with the blastocoel cavity and form the PrE layer, with a polarized epithelium and a basement membrane.

The Epi apolar cells are located deeper in the structure and protected from extraembryonic signalling.

ICM cells isolated from E4.5 embryos reconstitute a PrE layer on their surface (Rossant 1975), an observation suggesting that cell positioning also played a role in Pre/Epi specification. Thus, early models of PrE/Epi cell fate decision were conceptually similar to inside/outside model for TE/ICM specification (Figure 1.4A). However, identification of Epi and PrE markers, Nanog and Gata6, respectively, (Fujikura, Yamato et al. 2002, Mitsui, Tokuzawa et al. 2003, Chazaud, Yamanaka et al. 2006), and use of live imaging revealed that lineage commitment occurred earlier than previously thought and questioned the validity of the positioning model *per se*. Nanog and Gata6 were both expressed in E3.5 ICM, prior to positional segregation of the two lineages (Plusa, Piliszek et al. 2008). By the 64-cell stage, Gata6 and Nanog expression in the ICM is mutually exclusive, resulting in an apparently random "salt and pepper" distribution of PrE and Epi progenitors (Chazaud, Yamanaka et al. 2006, Kurimoto, Yabuta et al. 2006, Plusa, Piliszek et al. 2008). Furthermore, lineage tracing showed that at E3.5 the lineage was already restricted as single ICM cells gave rise to either Epi or PrE in E5.5 chimeras (Chazaud, Yamanaka et al. 2006). The current working model for PrE/Epi specification entails at least two phases, with an initial lineage commitment stage that is completed by E3.5, followed by cell assortment and positioning in the E4.5 blastocyst (Figure 1.4A).

Lineage specification has been connected to cleavage history within ICM cells. This model claims that cells internalised during the first wave of asymmetric cell division (8 – 16 cells) have a bias towards forming Epi, whereas cells internalised during the following two waves (16 – 32 and 32 – 64 cells) displayed a preference for PrE (Rossant, Chazaud et al. 2003, Yamanaka, Ralston et al. 2006). This spatial organisation would shield the primary inner cells, resulting from the first wave of cell division, from extraembryonic and external environment signalling. This hypothesis was experimentally addressed by two independent groups, yielding conflicting results. In a lineage tracing experiment in unmanipulated embryos, Morris *et al* confirmed this model (Morris, Teo et al. 2010). However, using a chimera approach, Yamanaka *et al* concluded both primary and secondary inner cells in ICM could contribute to Epi and Pre in similar proportions (Yamanaka, Lanner et al. 2010). These results indicate that cell history, at least on its own, is not sufficient to determine developmental fate. A model for progressive PrE and Epi commitment through a stochastic process (Yamanaka, Lanner et al. 2010) is currently the most accepted and will be discussed in the following sub-section.

Independently of how lineage is specified, upon fate determination, the ICM cells eventually reorganise to position the PrE layer at the surface in contact with the

blastocoel. This process is referred to as cell sorting and involves actin-dependent cell movement as well as retention of position by PrE-committed cells already in contact with the blastocoel (Plusa, Piliszek et al. 2008, Meilhac, Adams et al. 2009). The embryo then has a plasticity degree shown by the fact that some PrE-committed cells that fail to sort, keeping an inside location, can downregulate PrE markers and possibly acquire Epi fate; alternatively, these are eliminated through apoptosis (Plusa, Piliszek et al. 2008, Meilhac, Adams et al. 2009, Morris, Teo et al. 2010). PrE cells in their final position form a polarized epithelial layer (Gerbe, Cox et al. 2008) and show expression differences relative to pre-sorted cells, as illustrated by the specific activation of Sox7 in sorted cells (Artus, Piliszek et al. 2011).

1.1.3.1 Transcriptional networks and signalling associated with PrE/Epi commitment

An interplay between mutually exclusive factors Nanog and Gata6, controlled by differential FGF/ERK signalling in PrE and Epi precursors is the known basis of lineage specification and commitment (Figure 1.4B). Nanog expression is first detected at the 8-cell stage, being specific to ICM after the first cell lineage segregation and to Epi upon the second cell lineage decision (Chambers, Colby et al. 2003, Chazaud, Yamanaka et al. 2006). Nanog-null embryos arrest in preimplantation development, do not form Epi and overexpress the PrE marker Gata6 in the ICM (Mitsui, Tokuzawa et al. 2003). These results highlight the key role for Nanog in the Epi layer, but also as a negative regulator of Gata6.

Gata6 and Gata4 are individually not pivotal for PrE specification, but their knockout severely impairs PrE differentiation (Soudais, Bielinska et al. 1995, Morrisey, Tang et al. 1998, Koutsourakis, Langeveld et al. 1999). Gata6 and Gata4 might have redundant functions in preimplantation embryo and compensate for each other in PrE specification. In line with this, overexpression of dominant-negative Gata6 relocates outer cells to an inner location within the ICM (Meilhac, Adams et al. 2009). Overexpression of Gata6 is not sufficient to shift inner ICM cells to the surface, but this could be achieved by co-expression of another factor expressed in surface ICM cells, Wnt9a (Meilhac, Adams et al. 2009). These results suggest Gata6 might actually have a role in cell sorting during PrE formation.

Interestingly, at the 8-cell stage, blastomeres co-express both Nanog and Gata6, but these become mutually exclusive as PrE and Epi specify (Plusa, Piliszek et al. 2008, Guo,

Huss et al. 2010). This observation resembles the Oct4/Cdx2 co-repression and its role for TE/ICM segregation. Other factors have been identified that act downstream of Gata6 and cooperate with it for the definitive establishment of PrE fate, such as Sox17, Sox7 and Gata4 (Artus, Piliszek et al. 2011).

FGF/ERK signalling plays a key role in PrE specification as mutants of several genes in this pathway exhibit impairment in PrE formation. These include the Fgf receptor FgfR2 (Arman, Haffner-Krausz et al. 1998), its ligand Fgf4 (Feldman, Poueymirou et al. 1995, Goldin and Papaioannou 2003), and the adaptor molecule Grb2 (Cheng, Saxton et al. 1998, Chazaud, Yamanaka et al. 2006). Small molecule inhibition of FgfR2 or MEK effector prevents PrE formation (Nichols, Silva et al. 2009, Yamanaka, Lanner et al. 2010). Conversely, supply of high Fgf4 doses bias the majority of ICM cells towards a PrE fate (Yamanaka, Lanner et al. 2010). In the blastocyst, Fgf signalling is differentially regulated in such a way that Fgf4 is secreted from Epi cells as a direct target of Oct4 and Sox2 (Yuan, Corbi et al. 1995). In contrast, its receptor is downregulated in ICM cells that form Epi progenitors while specifically maintained in TE and PrE progenitor cells (Guo, Huss et al. 2010). Guo et al analysed the expression profiles of single blastomeres, and revealed that FgfR2 and Fgf4 expression onset occurs at the 32-cell stage and its mutually exclusive pattern precedes that of Nanog/Gata6 (Guo, Huss et al. 2010). This finding suggests FGF signalling acts upstream of the lineage restriction through Nanog/Gata6. The mechanisms for Fgf4 and FgfR2 regulation remain however unknown.

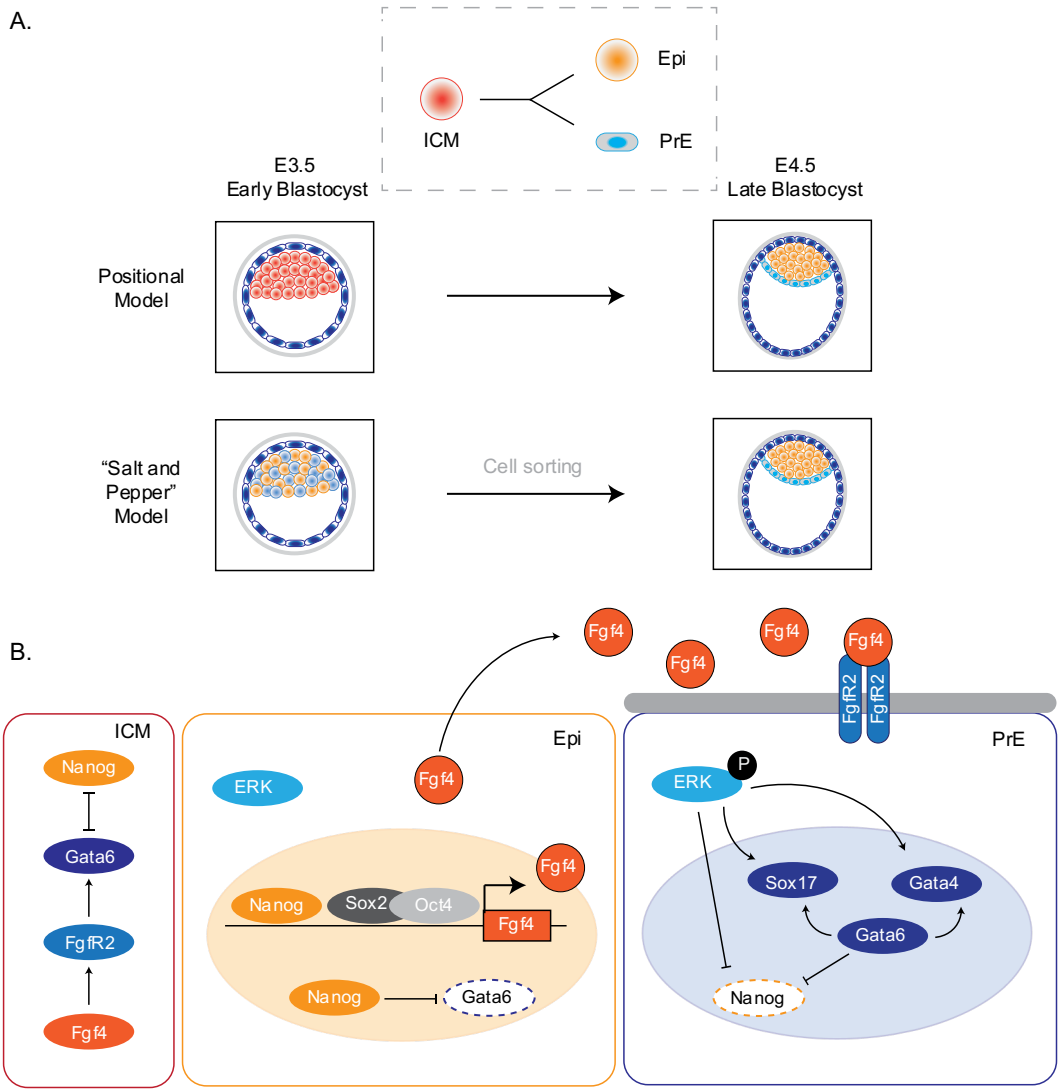

Figure 1.4 | The second cell lineage specification: Epiblast (Epi) versus Primitive Endoderm (PrE). A | Models explaining the second lineage decision in mouse embryos. The top schematic represents a classical view of an inside/outside model, in which cells positioned in the outside layer of the ICM, facing the blastocoel cavity, will differentiate towards PrE, whereas inside cells will form the Epi layer. A second model (bottom schematic) was later developed based on the observation of a mutually exclusive "salt and pepper" distribution of lineage determinants Gata6 and Nanog in the ICM. This model purports there is a first stage for lineage commitment completed by E3.5, followed by cell sorting and positioning of Gata6-PrE cells in the external layer facing the blastocoel cavity at E4.5. The inside layer forms the Epi, which is physically protected from extraembryonic signalling. **B |** The major pathway involved in this process is FGF/MAPK signalling, responsible for the differential expression of Nanog and Gata6. In the ICM, both Fgf4 and its receptor FgfR2 activate expression of Gata6, which inhibits Nanog. In Epi progenitor cells, FgfR2 expression is lost. A higher level of Nanog, in cooperation with Sox2 and Oct4 drive the expression of Fgf4 while Nanog continues to repress Gata6. On the other hand, PrE progenitors retain expression of FgfR2, which is activated by Fgf4 secreted from Epi neighbouring cells and result in phosphorylation of Erk. Activated Erk and Gatat6 both cooperate to inhibit Nanog expression and promote activation of Sox17 and Gata4, which together with FgfR2 and Gata6 form the transcriptional network required to maintain PrE identity. Figure recreated based on (Chazaud and Yamanaka 2016).

1.2 Stem cell lines derived from preimplantation mouse embryos

The blastocyst stage embryo is composed of three different lineages: trophectoderm (TE), epiblast (Epi) and primitive endoderm (PrE). The trophectoderm layer of the embryo is multipotent and restricted to form the extraembryonic tissues that support embryo development, such as the placenta. The epiblast is the pluripotent compartment of the blastocyst and has the potency to generate all tissues comprising the embryo-proper. Primitive endoderm is a second extraembryonic layer, and gives rise to the visceral and parietal endoderm layers of the yolk sac. Extensive research allowed determining culture conditions to isolate and maintain stem cell lines representative of these three embryonic lineages (Figure 1.5). Embryonic stem cells (ESCs) can be derived from the Epi (Evans and Kaufman 1981, Martin 1981) whereas trophoblast stem cells (TSCs) can be isolated from the TE (Tanaka, Kunath et al. 1998) and extraembryonic endoderm (XEN) cells from the PrE (Kunath, Arnaud et al. 2005). The term "stem cell" implies these cell lines can i) self-renew and proliferate for long periods in culture while ii) remaining in an undifferentiated state under appropriate culture conditions, and iii) retain the developmental potential of their embryonic counterparts differentiating to equivalent cell lineages both *in vivo* and *ex vivo*.

These different stem cell lines provide valuable *in vitro* models to study mechanisms involved in lineage specification and maintenance in early embryos.

1.2.1 Embryonic Stem Cells (ESCs)

A major breakthrough in stem cell biology occurred when two independent groups established that culturing the epiblast from mouse blastocysts on a mitotically inactivated mouse embryonic fibroblast (MEF) feeder layer and in the presence of serum allowed the *in vitro* derivation of ESCs (Evans and Kaufman 1981, Martin 1981). These ESCs were shown to retain the epiblast pluripotency by their ability to form teratocarcinomas when injected into mice, their potential to differentiate to all somatic lineages, and their capacity to contribute to chimeras with germline transmission (Evans and Kaufman 1981, Martin 1981, Bradley, Evans et al. 1984, Beddington and Robertson 1989).

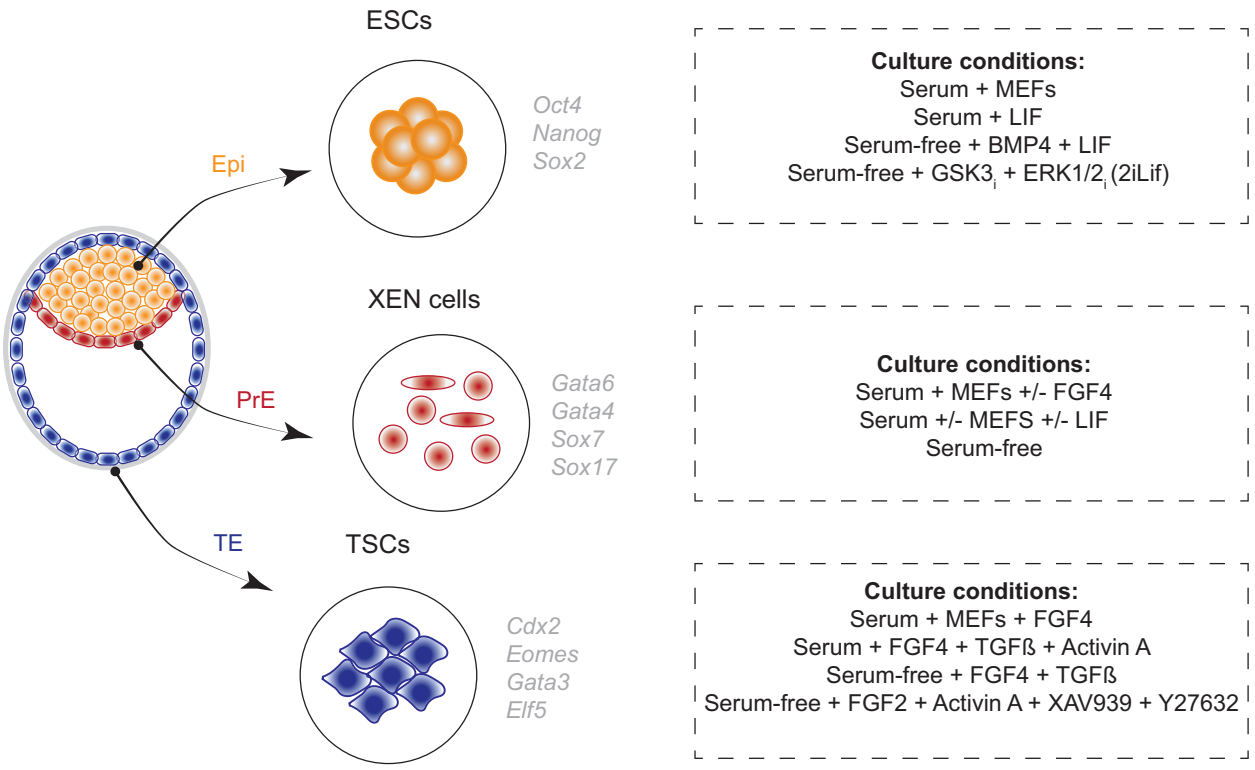

Figure 1.5 | Stem cell lines can be established from blastocyst stage mouse embryos and cultured indefinitely *in vitro*. Stem cell lines representative of the three layers that compose a blastocyst stage embryo can be isolated and maintained *in vitro*. Embryonic stem cells (ESCs) are derived from the epiblast (Epi), trophoblast stem cells (TSCs) from the trophectoderm (TE) and extraembryonic endoderm (XEN) cell from the primitive endoderm (PrE). These stem cell lines are characterised by different morphologies, expression of different lineage-specific transcription factors (highlighted in grey) and culture conditions. Importantly, they retain the developmental potential of their embryonic counterparts in chimeras. Figure recreated based on Artus *et al* (Artus and Hadjantonakis 2012).

The discovery of ESCs has revolutionised our understanding of pluripotency, self-renewal and the mechanisms controlling differentiation in development. Furthermore, ESCs allowed germline transmission of genetic modifications introduced *in vitro*, transforming our capacity to generate mouse models with defined genetic perturbations (Robertson, Bradley et al. 1986).

The initial culture condition to derive ESCs contained unknown factors secreted by the feeder layer. Dissection of individual culture components identified Leukaemia inhibitory factor (LIF) as the key element secreted by feeders, thus allowing the culture of ESCs in feeder-free conditions when supplementing serum-containing medium with LIF (Smith, Heath et al. 1988, Williams, Hilton et al. 1988). LIF signalling occurs through LIFR and GP130 membrane receptors, resulting in phosphorylation of the transcription factor STAT3 (signal transducer and activator of transcription 3) (Niwa, Burdon et al.

1998, Matsuda, Nakamura et al. 1999), which then translocates into the nucleus where it regulates gene expression of factors such as *Myc* and *Klf4* (Cartwright, McLean et al. 2005, Hall, Guo et al. 2009). In fact, ESCs expressing a stabilized form of MYC protein can be maintained with LIF-free serum-medium (Cartwright, McLean et al. 2005).

Bone morphogenic protein 4 (BMP4) was later identified as the pivotal component in serum, such that ESCs can be derived and maintained in an undifferentiated state in serum-free media supplemented with LIF and BMP4 (Ying, Nichols et al. 2003). Additionally, it has been shown that rather than exogenous culture factors promoting self-renewal, this state is well preserved through inhibition of intrinsic differentiation signalling (Burdon, Stracey et al. 1999). In particular, the combined inhibition of glycogen synthase kinase 3 (GSK3) and fibroblast growth factor (FGF) / extracellular-signal-related-kinase (ERK) signalling – a culture condition termed 2iLif – allowed the establishment of ground state ESCs (also known as naïve ESCs) (Ying, Wray et al. 2008).

In ESCs, FGF4 acts as an intrinsic factor to mediate exit from the pluripotent state and initiation of lineage commitment through stimulation of ERK (Kunath, Saba-El-Leil et al. 2007). GSK3 on the other hand acts via the canonical Wnt pathway, regulating ubiquitination and degradation of β-catenin. The β-catenin destruction complex is composed of Axin, GSK3-β, APC (adenomatous polyposis coli) and CK-1 (casein kinase 1). In the cytoplasm, GSK promotes phosphorylation of β-catenin, tagging it for degradation. Upon inhibition of GSK3, β-catenin remains unphosphorylated and translocates to the nucleus where it binds to TCF3 transcriptional repressor, blocking it from binding to key pluripotency *loci* and thereby promoting ESC self-renewal (Wray, Kalkan et al. 2011).

2iLif culture conditions allowed efficient derivation of ESCs from strains previously non-permissive to ESC establishment (Ying, Wray et al. 2008, Nichols, Silva et al. 2009). In addition, compared to serum-cultured ESCs, 2iLif cultures have a more homogeneous expression of pluripotency factors, minimal expression of lineage specifiers and reduced amounts of both bivalent domains and repressive histone marks (Marks, Kalkan et al. 2012). It was recently shown that global transcriptome in naïve ESCs strongly resemble that of E4.5 epiblast cells (Boroviak, Loos et al. 2015).

The core transcriptional network promoting pluripotency and maintaining prolonged ESC self-renewal in culture has been extensively characterised. Core components consist on Oct4, Sox2 and Nanog, which act in a positive feedback loop to regulate themselves, in addition to preventing expression of lineage specifier genes (Boyer, Lee et al. 2005, Loh, Wu et al. 2006). Consistent with their key role in pluripotency, ESCs

cannot be derived from *oct4-* or *sox2-* null embryos (Nichols, Zevnik et al. 1998, Avilion, Nicolis et al. 2003). Moreover, the levels of Oct4 are tightly controlled in self-renewing ESCs, in such a way that its downregulation promotes TSC differentiation whereas its upregulation induces PrE and mesoderm differentiation (Niwa, Miyazaki et al. 2000). *Nanog*, identified for its role promoting pluripotency in the absence of LIF, enhances ESC self-renewal but is dispensable to maintain this property (Mitsui, Tokuzawa et al. 2003, Chambers, Silva et al. 2007). Similar roles were reported for Klf4 and Esrrb (Niwa, Ogawa et al. 2009, Martello, Sugimoto et al. 2012, Martello, Bertone et al. 2013).

In summary, the regulatory network that sustains ESC self-renewal integrates both exogenous and intrinsic signals that maintain the balance between the pluripotent state and differentiation (recently reviewed by Li et al (Li and Belmonte 2017)).

1.2.2 Trophoblast Stem Cells (TSCs)

Early transplantation and chimerisation experiments have suggested the presence of a stem cell population within the E5.5 – E6.5 trophoblast compartment (Rossant, Gardner et al. 1978). Reciprocal expression of FgfR2 and its ligand Fgf4 suggested the trophoblast could be dependent on embryonic FGF signalling (Orr-Urtreger, Bedford et al. 1993, Feldman, Poueymirou et al. 1995). Consistently, TSC lines could be established from the extraembryonic ectoderm in E6.5 embryos, but also from blastocysts, when cultured on inactivated MEF feeder layers and serum-medium supplemented with Fgf4 and Heparin (Tanaka, Kunath et al. 1998). Unlike ESCs, TSCs could be derived from embryos up to E8.5, from particular niches within the extraembryonic compartment (Tanaka, Kunath et al. 1998, Uy, Downs et al. 2002). Importantly, all these different origins allow establishment of TSC cultures with similar developmental plasticity, equivalent to the blastocyst TE.

Under appropriate culture conditions, TSCs can self-renew maintaining an undifferentiated state. Withdrawal of MEFs, Fgf4 and Heparin results in terminal differentiation to different trophoblast types (Tanaka, Kunath et al. 1998). Furthermore, TSCs contribute exclusively to placenta in chimeras, being able to support normal development to term (Tanaka, Kunath et al. 1998).

Continued studies on TSC culture conditions found that members of the Tgfβ superfamily Activin and Tgfβ, but not Nodal, can replace feeders, demonstrating the role for this signalling in TSC self-renewal (Erlebacher, Price et al. 2004, Natale, Hemberger et al. 2009). The presence of unknown factors in serum stimulated the development of

chemically defined media for the precise study of TSC biology. Kubaczka *et al* developed TX media, a formulation based on DMEM/F12 supplemented with ten components including Fgf4, Heparin, insulin and Tgfβ1 (Kubaczka, Senner et al. 2014). Culturing TSCs in TX media requires coating the plates with extracellular basement membrane components such as Matrigel (Kubaczka, Senner et al. 2014). Ohinata *et al* developed another chemically defined medium, CDM/FAXY, with a formulation that includes Fgf2, activin A and small molecules inhibitors XAV939 (canonical Wnt inhibitor) and Y27632 (Rho-associated kinase p160ROCK inhibitor) (Ohinata and Tsukiyama 2014). In this case, TSCs need to be maintained in fibronectin-coated dishes. Both TX and CDM/FAXY media allow derivation and maintenance of TSC cultures with full differentiation capacity and *in vivo* potency to colonise the placenta in chimeras (Kubaczka, Senner et al. 2014, Ohinata and Tsukiyama 2014).

FGF signalling is essential for maintenance of TSCs. Consistently, deletion of its components Fgf4, FgfR2 or Erk2 results in preimplantation (Fgf4) or early postimplantation (FgfR2 and Erk2) developmental arrest, affecting extraembryonic layers (Feldman, Poueymirou et al. 1995, Arman, Haffner-Krausz et al. 1998, Saba-El-Leil, Vella et al. 2003). The transcriptional network governing TSC self-renewal revolves around Cdx2 expression, as reviewed previously for TE specification, and will be discussed in more detail in section 1.3.2.

1.2.3 Extraembryonic Endoderm Stem Cells (XEN cells)

XEN cell lines can be established from the blastocyst stage embryos (Kunath, Arnaud et al. 2005). These cells were derived in different culture conditions (Kunath, Arnaud et al. 2005, Niakan, Schrode et al. 2013). Surprisingly, they do not always require Fgf4 signalling for derivation and maintenance (Kunath, Arnaud et al. 2005). XEN cells have also been derived using ESC conditions, with feeders and serum and LIF supplementation. *In vitro*, XEN cells have gene expression profile similar to specified PrE cells, with expression of Gata6, Gata4, Sox7 and Sox17. In fact, these factors have been implicated in XEN cell self-renewal mechanism (Lim, Tam et al. 2008, Niakan, Ji et al. 2010). *In vivo*, XEN cells maintain the lineage potency of PrE cells and contribute to the extraembryonic visceral and parietal endoderm in chimeras (Kunath, Arnaud et al. 2005). The signalling pathways required to maintain XEN cells are largely unknown, but Pdgfα is required for the establishment of XEN cultures (Artus, Panthier et al. 2010).

1.3 The trophoblast lineage and early placenta development

The placenta is a complex organ that accomplishes remarkable functional diversity. The trophoblast lineage forms the main structural and functional elements needed for the placenta to establish the maternal-fetal interface (Rossant and Cross 2001). Early on, trophoblast cells are responsible for attachment and invasion of the uterine walls during implantation and dictate vascular remodelling and cell fusion that bring maternal and fetal blood systems in close proximity (Cross, Hemberger et al. 2002), establishing the maternal-fetal interface. Here, the placenta mediates exchange of gases, nutrients and waste products between mother and fetus (Maltepe, Bakardjiev et al. 2010). Additionally, the placenta is involved in immune protection of the fetus (PrabhuDas, Bonney et al. 2015), as well as secretion of hormones and growth factors essential for pregnancy progress (Newbern and Freemark 2011).

1.3.1 Early placenta development and trophoblast differentiation

The trophoblast layer of the placenta develops from the very first differentiated cell lineage in the mammalian embryo, the trophectoderm (TE). In the mouse embryo, trophoblast differentiation starts around E4.5, at the onset of implantation. In the blastocyst, TE cells in direct contact with the epiblast are denoted polar TE, whereas TE cells in contact with the blastocoel cavity form the mural TE (Figure 1.6A). Around E4.5, the mural TE cells stop dividing and undergo endoreplication – that is, several rounds of DNA replication without cell mitosis –, yielding the first layer of polyploid trophoblast giant cells (TGCs), the primary TGCs (Hu and Cross 2010). These are terminally differentiated and highly specialised cells. The polar TE on the other hand, continues to proliferate and shortly after implantation generates the diploid extraembryonic ectoderm (ExE) (Copp 1979). The ExE further expands and is the precursor of the ectoplacental cone (EPC), formed by E6.0, as well as primary TGCs. As development progresses, by E7.5 ExE gives rise to the chorionic epithelium, which is in contact with a layer of mesothelium. The allantois is also formed at this stage, and is derived from embryonic mesoderm. By E8.5, the allantois attaches to the chorion, in a process termed chorioallantoic fusion. It results in morphogenesis in this interface, opening space for fetal blood vessels growth from the allantois and in this way establishing the fetal side of the placental vascular network. This dual structure undergoes villous branching, forming a highly packed structure termed the labyrinth, which constitutes the area for direct

product exchange between fetal and maternal blood supply (Hemberger, Kurz et al. 2001, Cross, Hemberger et al. 2002, Maltepe, Bakardjiev et al. 2010). Around the chorioallantoic fusion, trophoblast from the chrorionic plate differentiate into different layers of labyrinth trophoblast, the syncytiotrophoblast (SnT), type I and type II. Additionally, chorion also forms sinusoidal TGCs, located at the blood vessels interface (Simmons, Fortier et al. 2007).

Development of the labyrinth is structurally supported by the spongiotrophoblast (SpT) derived from the EPC, which form a compact layer between the labyrinth and the external TGC layer. EPC further differentiates into glycogen cells, with unknown function (Prudhomme and Morey 2016), and secondary TGCs which are located at the interface between SpT and the decidua (John and Hemberger 2012).

SpT and sinusoidal TGCs produce hormones and cytokines that lead to remodelling of the maternal vasculature, contribute to maternal and fetal nutrient exchange as well as immunity factors (Cross, Hemberger et al. 2002, John and Hemberger 2012).

As described, the trophoblast lineage has a high degree of cellular plasticity, producing different cell types with a wide variety of functions in the placenta. However, the transcriptional network regulating the trophoblast lineage tree is still largely unknown, despite the identification of some subtype markers (Figure 1.6B). TSCs are at the top of the lineage tree and express the markers *Cdx2*, *Eomes* and *Elf5*. Despite *in vivo* exhaustion of trophoblast stem cell potential by E8.5, the rodent placenta continues to grow nearly until term (Adamson, Lu et al. 2002, Uy, Downs et al. 2002, Simmons and Cross 2005). This raised the possibility that thereafter, placenta development depends on multipotent progenitor cells. In fact, two different progenitor populations have been identified. Ueno *et al* isolated progenitors of the labyrinth layer, identified by the high expression of Epcam, which can differentiate into Syn-I, Syn-II and sinusoidal TGCs, and are dependent on the hepatocyte growth factor c-Met signalling (Ueno, Lee et al. 2013). In addition, Mould *et al* identified a population of proliferative precursors within the spongiotrophoblast layer, characterised by expression of Blimp1, which can differentiate to particular TGC types and glycogen cells (Mould, Morgan et al. 2012). These progenitors could not be maintained in culture so far, but their identification opens new avenues for the discovery of new trophoblast subtypes.

Much has been described about the placenta (reviewed recently by Malpete *et al* (Maltepe and Fisher 2015)). However, precise functions for different cell types, as well as the processes that lead to differentiation and subtype specification remain largely unknown and will provide valuable insights for both developmental biology and reproductive medicine.

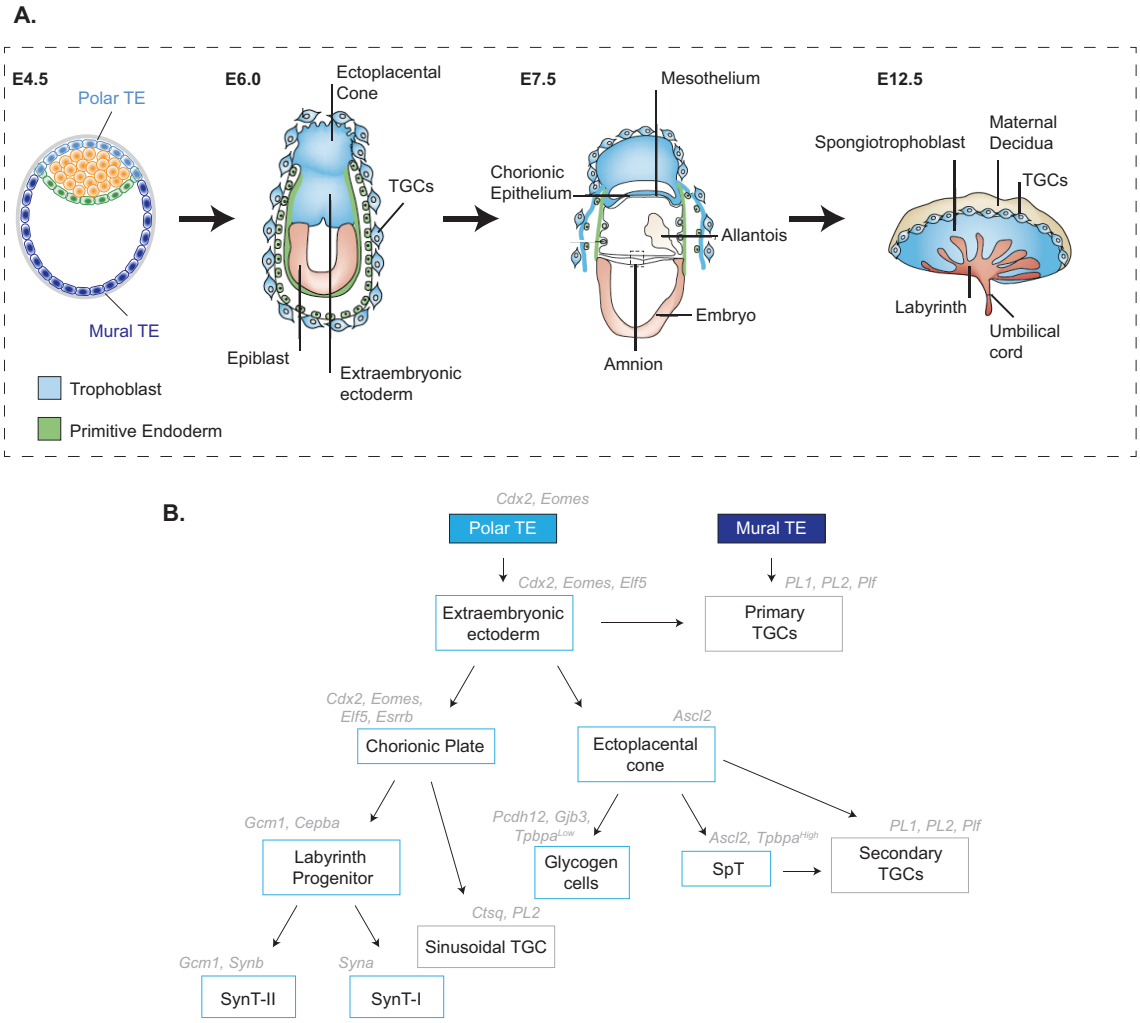

Figure 1.6 | Early placenta development and trophoblast differentiation. A | Early mouse embryo development, depicting the organization of extraembryonic lineages from embryonic day E4.5, leading to an established functional placenta by E12.5. TGCs, trophoblast giant cells. Image adapted from Rossant *et al* (Rossant and Cross 2001). **B |** Differentiation of trophoblast stem cells into the different trophoblast lineages that compose a mature placenta. The mural TE gives rise to primary TGCs. The polar TE forms the extraembryonic ectoderm which then divides into two axes of differentiation: the chorionic plate and the ectoplacental cone. The chorionic plate will form sinusoidal TGCs and labyrinth progenitors, which then generate syncytiotrophoblast (SynT). The ectoplacental cone forms glycogen cells and spongiotrophoblast (SpT), which can also produce secondary TGCs (also directly generated from ectoplacental cone progenitors). Highlighted in grey are gene markers expressed by the different populations. Scheme adapted from Latos *et al* (Latos and Hemberger 2016).

1.3.2 Trophoblast Stem Cell Regulatory Networks

1.3.2.1 TSC self-renewal core transcriptional network

Analysis of mouse mutants exhibiting embryonic lethality at different developmental stages has identified several core components of trophoblast specification and differentiation. Tead4 was identified as one of the earliest regulators of TE specification as *tead4*-null mutants arrest during preimplantation, lacking the TE layer (Yagi, Kohn et al. 2007). *Cdx2* was also identified as essential at the blastocyst stage (Strumpf, Mao et al. 2005), whereas *Eomes* displayed a role post implantation (Russ, Wattler et al. 2000, Strumpf, Mao et al. 2005). Other genes were associated with embryonic lethality at later stages such as *Esrrb*, *Ets2*, *Tfap2c* or *Elf5* (Luo, Sladek et al. 1997, Yamamoto, Flannery et al. 1998, Auman, Nottoli et al. 2002, Werling and Schorle 2002, Donnison, Beaton et al. 2005). *Ets2*-null embryos arrested by E8.5, with failure of ectoplacental cone proliferation (Yamamoto, Flannery et al. 1998). *Tfap2c* was found to be expressed in all preimplantation lineages, but becomes restricted to the extraembryonic lineages at implantation (Auman, Nottoli et al. 2002). *Tfap2c*-null embryos however arrest around E8.5 with malformation in the extraembryonic tissues, demonstrating it is not essential for TE specification or early differentiation, but rather for later stages in the trophoblast lineage (Auman, Nottoli et al. 2002, Werling and Schorle 2002). *Elf5*-null embryos were shown to fail at the gastrulation stage, with an abnormal extraembryonic ectoderm layer; additionally, TSCs could not be derived from these embryos demonstrating the essential role for *Elf5* in the TSC transcriptional network (Donnison, Beaton et al. 2005).

Cdx2 and *Eomes* were initially found at the centre of the TSC self-renewal network (Tanaka, Kunath et al. 1998). *Elf5* was later established as necessary to boost and fully establish the TSC self-renewal programme through positive feedback loops with *Cdx2* and *Eomes* (Ng, Dean et al. 2008). *Tfap2c* was also established as essential for TSC self-renewal in a mechanism independent from *Cdx2*, but that requires cooperation of both *Tfap2c* and *Cdx2* networks for upregulation of *Elf5* and complete establishment of a self-renewing state (Kuckenberg, Buhl et al. 2010). In fact, the stoichiometric levels of *Eomes*, *Tfap2c* and *Elf5* determine TSC self-renewal and onset of differentiation (Latos, Sienerth et al. 2015). Comparison between gene expression profiles of TSCs, ESCs and XEN cells identified *Gata3* as another axis of TSC self-renewal, acting in response to *Tead4*, but in a parallel network to *Cdx2* (Ralston, Cox et al. 2010). Additionally, *Sox2* and *Esrrb* were identified as key genes for TSC self-renewal acting in response of Fgf signalling (Adachi, Nikaido et al. 2013, Latos, Goncalves et al. 2015).

Genome-wide occupancy analysis of the transcription factors stated above revealed they co-occupy a significant number of target *loci*, including themselves, this way establishing an autoregulatory positive feedback loop that reinforces TSC self-renewal network (Kidder and Palmer 2010, Adachi, Nikaido et al. 2013, Chuong, Rumi et al. 2013, Latos, Goncalves et al. 2015).

As described previously regarding the first cell lineage decision, cell positioning, polarity and Hippo signalling are major determinants of the trophoblast fate. In addition, Fgf signalling is also differentially regulated between embryonic and extraembryonic layers with Fgf4 ligand being secreted by embryonic lineages, but having its receptor FgfR2 restricted to extraembryonic TE and PrE. Indeed, the establishment of TSC cultures is dependent on Fgf4 signalling (Tanaka, Kunath et al. 1998).

Recent studies aimed at analysing the downstream effectors of Fgf signalling in TSCs (Figure 1.7). Adachi *et al* and Latos *et al* used small molecule inhibitors of Fgf/Erk signalling to dissect induced transcriptomic changes (Adachi, Nikaido et al. 2013, Latos, Goncalves et al. 2015). Both teams modulated Fgf signalling by inhibition of Fgf receptor using PD173074, and of Mek using PD0325901, concluding that TSCs downregulate self-renewal genes such as *Cdx2*, *Eomes*, *Elf5*, *Sox2* or *Esrrb*. Transcriptomic analysis at 3h and 24h of Fgf inhibition revealed *Esrrb* and *Sox2* as two of the genes with faster and stronger downregulation (Latos, Goncalves et al. 2015). In fact, it was demonstrated that co-overexpression of these two factors was sufficient to allow TSC self-renewal in Fgf4-depleted medium (Adachi, Nikaido et al. 2013).

Both *Sox2* and *Esrrb* are also pivotal for ESC self-renewal (Avilion, Nicolis et al. 2003, Martello, Sugimoto et al. 2012). Adachi *et al* performed microarray analysis and chromatin immunoprecipitation experiments to determine the *loci* occupancy of Sox2 in both ESCs and TSCs. They demonstrate Sox2 occupies a different subset of genes in ESCs and TSCs, including *Oct4* and *Cdx2*, respectively. In addition, *Sox2* knockout promoted downregulation of *Oct4*, *Nanog* or *Klf4* in ESCs, and *Cdx2* or *Elf5* in TSCs. Furthermore, it was demonstrated that in TSCs, *Sox2* interacts with *Tfap2c* to upregulate several TSC core genes (Adachi, Nikaido et al. 2013).

Similarly, mass spectrometry analysis of the *Esrrb* interactome in ESCs and TSCs showed it interacts with different networks, and chromatin immunoprecipitation revealed it binds to *Elf5* exclusively in the TSC compartment (Latos, Goncalves et al. 2015). *Esrrb* was shown to directly bind and regulate several TSC transcription factor *loci* such as *Eomes*, *Elf5* and *Cdx2*.

The dual role for Sox2 and Esrrb in TSCs and ESCs demonstrate that transcription factors display a versatile function depending on their cellular environment and can orchestrate completely different transcriptional networks even in neighbouring cell types.

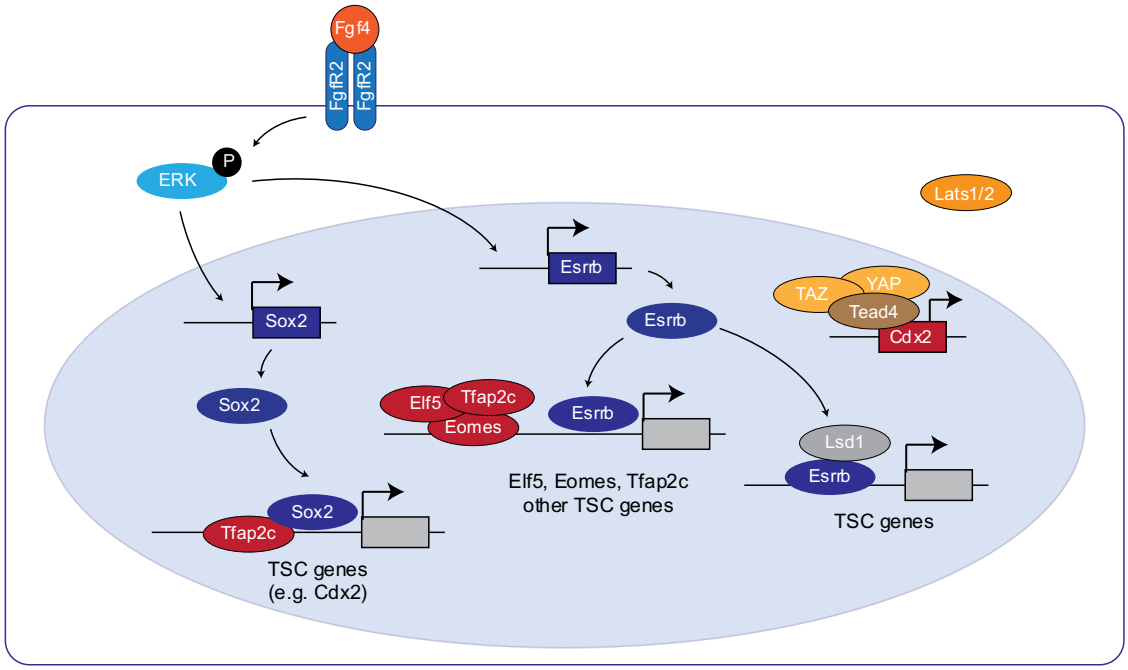

Figure 1.7 | Signalling and transcriptional regulation of self-renewal in TSCs. Fgf and Hippo signalling are the two main signalling pathways regulating TSC transcriptional network. Inactive Hippo signalling maintains YAP/TAZ unphosphorylated, which translocate to the cell nucleus and cooperate with Tead4 to activate Cdx2 expression. Fgf signalling is activated by binding of Fgf4 to its receptor FgfR2, promoting activation of the Mek/Erk pathway. Phosphorylated Erk then indices the expression of *Sox2* and *Esrrb*. Sox2 forms a complex with Tfap2c and together activate the expression of several TSC genes, such as *Cdx2*. Similarly, Esrrb also promotes expression of core TSC genes such as Elf5, Tfap2c and Eomes, as well as others. This activation is individually regulated by also in association with histone demethylase Lsd1. *Elf5*, *Tfap2c*, *Eomes* and *Cdx2* establish positive feedback loops, reinforcing each other's expression and maintaining an undifferentiated TSC state. Schematics based on Latos *et al* (Latos and Hemberger 2016).

1.3.2.2 Epigenetic regulation of TSCs

TSCs and ESCs can be readily distinguished based on their chromatin state, distribution of histone marks and global methylation status (Alder, Lavial et al. 2010, Rugg-Gunn, Cox et al. 2010, Senner, Krueger et al. 2012).

ESCs have an atypical chromatin state, which is highly decondensed and enriched for active histone marks (Meshorer and Misteli 2006, Meshorer, Yellajoshula et al. 2006).

This results in low expression of both coding and noncoding genetic elements (Efroni, Duttagupta et al. 2008). Exit from the pluripotent state is accompanied by the appearance of large condensed domains of heterochromatin, accumulation of repressive histone marks and activation of a tissue-specific transcription network (Meshorer and Misteli 2006, Efroni, Duttagupta et al. 2008). Furthermore, ESCs display bivalent chromatin domains at lineage specifier *loci*, composed by active and repressive histone marks, H3K4me2/3 and H3K27me3, respectively (Azuara, Perry et al. 2006, Bernstein, Mikkelsen et al. 2006). Upon differentiation, these bivalent domains resolve to maintain one of the marks according to expression or repression of the target gene. Bivalent domains might reflect culture heterogeneity in serum-conditions as both naïve ESCs and ICM display lower levels of these domains (Marks, Kalkan et al. 2012, Liu, Wang et al. 2016).

Evaluation of chromatin structures by electron microscopy showed that, similarly to ESCs, epiblast cells in the embryo also displayed an open chromatin (Ahmed, Dehghani et al. 2010). In contrast, cells in the extraembryonic layers and their *in vitro* counterparts, TSCs and XEN cells, display larger domains of condensed chromatin (Ahmed, Dehghani et al. 2010).

The overall levels of H3K27me3 repressive mark are much lower in TSCs compared to ESCs, suggesting gene repression might be mediated by different marks. This observation could be related to lower expression of *Eed* in TSCs, a component of the Polycomb repressive complex 2 (PRC2) which mediates H3K27 methylation (Rugg-Gunn, Cox et al. 2010, Saha, Home et al. 2013). In fact, Rugg-Gunn *et al* demonstrated that H3K9me3 functions as a repressive histone mark in extraembryonic stem cells (Rugg-Gunn, Cox et al. 2010).

TSCs and XEN cells display lower number of bivalent domains compared to ESCs, suggesting a different regulation in the extraembryonic stem cells (Alder, Lavial et al. 2010, Rugg-Gunn, Cox et al. 2010). Additionally, it has been reported that in TSCs a set of somatic lineage regulators marked with bivalent chromatin marks H3K27me3 and H3K4me2 are specifically recognised by Suv39h1 (Alder, Lavial et al. 2010). Suv39h1 deposits an extra repressive H3K9me3 mark at such *loci*, resulting in stable repression of genes associated with embryonic development, so that they are not erroneously expressed during trophoblast differentiation. The mechanism involves exclusion of the PRC1 component Ring1b (Rnf2) from bivalent domains, this way removing the poised RNA polymerase II from these sites and effectively preventing their expression (Alder, Lavial et al. 2010). Besides the differences in chromatin condensation, distribution of active and repressive histone marks and abundance and regulation of bivalent domains

(Figure 1.8A), TSCs also display distinct global methylation profiles compared to ESCs (Figure 1.8B).

Early analysis of DNA methylation status used southern-blot to compare derivatives of the epiblast and extraembryonic lineages, revealing these displayed significantly lower methylation in repetitive elements and specific protein coding genes compared to the epiblast (Chapman, Forrester et al. 1984, Rossant, Sanford et al. 1986). This early observation was later confirmed by 5-mC (5-methylcytosine) immunostaining for blastocyst and early postimplantation embryos (Santos, Hendrich et al. 2002).

In development, lineage commitment should be followed by the establishment of an epigenetic memory that creates stable lineage barriers. In the early embryo, lineage restriction is followed by asymmetric distribution of DNA methylation, such that around implantation, ICM derivatives show *de novo* methylation whereas TE derivatives do not (Santos, Hendrich et al. 2002).

Mass spectrometry and MeDIP (methylated DNA immunoprecipitation) analysis revealed the first global view of differences between ESCs, TSCs and XEN cells demonstrating these have distinct methylomes (Senner, Krueger et al. 2012). Again, results confirmed the extraembryonic stem cells were significantly hypomethylated at repetitive satellite elements, retaining this property from their *in vivo* counterparts (Chapman, Forrester et al. 1984, Oda, Yamagiwa et al. 2006). MeDIP identified that extraembryonic stem cells were hypermethylated at CpG islands (CGI) compared to ESCs, in particular at CGI and promoters from pluripotency genes such as *Nanog* and *Oct4*. Importantly, for most CGIs, TSCs displayed higher methylation than ESCs. Key TSC genes such as *Cdx2*, *Eomes* or *FgfR2*, remain hypomethylated in ESCs (Ng, Dean et al. 2008). The discovery of differentially methylated genes, which show hypermethylated CGI in ESCs, and are thus repressed, but are hypomethylated and expressed in TSCs represent potential gatekeepers of the first cell lineage decision (Ng, Dean et al. 2008, Senner, Krueger et al. 2012, Cambuli, Murray et al. 2014). In fact, *Elf5* was identified as such a gene and is shown to be part of an epigenetic memory and lineage barrier between ESCs and TSCs (Ng, Dean et al. 2008, Cambuli, Murray et al. 2014).

Collectively, these observations reveal that TSCs and ESCs use DNA methylation in different regulatory roles: TSCs repress embryonic determinant genes by methylation of their CGI, whereas ESCs use methylation to silence repetitive sequences. However, ESCs seem to methylate a small subset of key gatekeeper genes, which form an epigenetic memory of epiblast origin and prevent the effective activation of a trophoblast programme in the embryonic compartment (Cambuli, Murray et al. 2014).

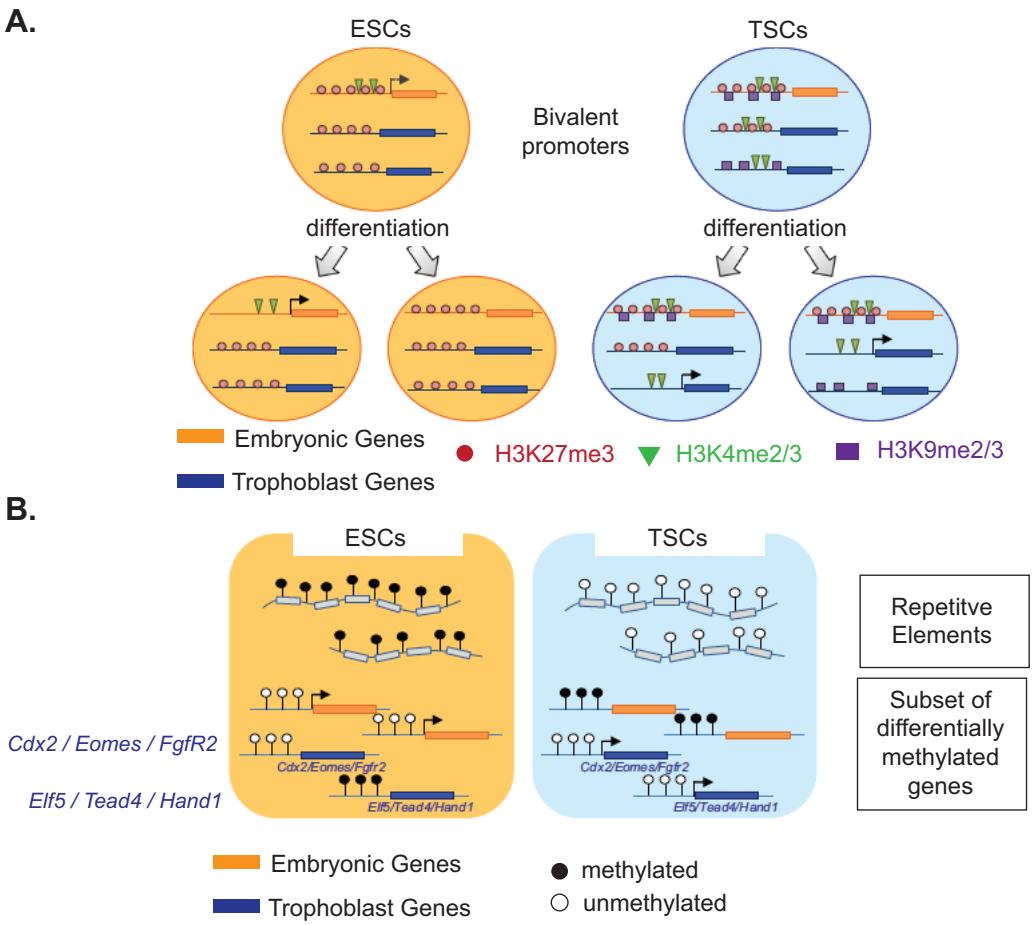

Figure 1.8 | Epigenetic regulation of TSCs. A | Characteristics of bivalent domains in ESCs and TSCs. In ESCs, bivalent domains are marked by H3K4me2/3 and H3K27me3. Upon differentiation the bivalent domain is resolved, H3K27me3 mark is maintained in genes that are repressed, whereas H3K4me2/3 marks expressed genes. In TSCs, the bivalent domains marking embryonic/somatic genes are permanently repressed with an extra layer of chromatin marks H3K9me2/3, ensuring these will not be expressed during TSC differentiation. **B |** DNA methylation in ESCs and TSCs. Highly repetitive elements are hypermethylated in ESCs and hypomethylated in TSCs. ESCs have a general hypomethylated in gene promoters. However, a small subset of trophoblast genes shows differential methylation status, being hypermethylated and repressed in ESCs and hypomethylated and expressed in TSCs. Figure adapted from Prudhomme *et al* (Prudhomme and Morey 2016).

Overall, our knowledge of regulatory networks in the extraembryonic tissues and stem cells is gradually growing. Nevertheless, in comparison with pluripotency and self-renewal, we are just starting to understand how the trophectoderm lineage is established and maintained. We have also limited knowledge on the pathways involved in the first cell lineage decision, that result in trophoblast specification. Differences in epigenetic and methylation profiles can readily distinguish TSCs and ESCs. However, the role of

these differences in the lineage decision, or how do they cooperate to maintain an epigenetic memory and cell fate commitment remain to be determined.

1.3.3 *In vitro* Trophoblast Stem Cell Lines – Identity and Potency

The evidence for the possible existence of a stem cell population within the extraembryonic compartments of preimplantation and early postimplantation mouse embryos arose through chimera assays and embryo reconstitution experiments in the early 70s-80s. Gardner *et al.* used reconstitution of blastocysts with ICM and TE cells containing different glucose phosphate isomerase (GPI) isozymes as genetic markers (Gardner, Papaioannou et al. 1973). This way, it was possible to trace the descendants of both layers, showing that TE progeny could be found at the EPC and TGCs in E7.5 embryos, but not in the embryo proper. Similar conceptual experiments were later performed using ICM and TE from *Mus caroli* or *Mus musculus*, respectively, followed by *in situ* hybridization with probes that distinguish the two genetic backgrounds at E7.5 embryos. The results consistently demonstrated that TE derivatives could be found in the EPC, ExE and TGCs, but not in the embryo nor yolk sac (Rossant and Croy 1985). In addition, isolation of ExE cells from E5.5 and E6.5 embryos followed by re-introduction into E3.5 host blastocysts, showed these could colonize the EPC and TGCs in E9.5 chimeric conceptuses (Rossant, Gardner et al. 1978). These experiments however, showed about 36% contribution towards EPC, compared to 70-90% contribution to TGCs, demonstrating a possible *in vivo* bias for differentiation of ExE cells at E5.5 and E6.5 stages.

The identification of a series of polypeptide markers which allowed distinction of ExE, EPC and TGCs, led to the discovery that ExE cells isolated from E7.5 conceptuses and cultured *in vitro* would differentiate to EPC after two days, followed by TGCs after six days. Additionally, EPC cells isolated from E7.5 would progress to TGC profile after four days of *in vitro* culture (Johnson and Rossant 1981).

Together, these early experiments suggested that TE cells from the E3.5 blastocyst could differentiate to ExE in postimplantation embryos, which would then originate the EPC and TGCs through development. It was later shown that *in vitro* co-culture of ExE isolated from E7.5 embryos, enclosed by ICM derivatives could maintain an ExE diploid population, whereas the same diploid state could not be maintained for EPC-isolated cells (Rossant and Tamura-Lis 1981). This observation led to the proposal of a stem cell model for the trophoblast lineage in which ExE diploid cells act as a stem cell pool for

all trophoblast derivatives (Rossant and Tamura-Lis 1981). Furthermore, it showed that *in vitro* maintenance of this presumed stem cell population was dependent on embryonic signalling.

Indeed, as discussed in section 1.2.2, later studies narrowed down this embryonic signalling to the FGF pathway (Orr-Urtreger, Bedford et al. 1993, Feldman, Poueymirou et al. 1995) and paved the way for the first report of the derivation of trophoblast stem cells (TSCs), seventeen years after the stem cell model proposal (Tanaka, Kunath et al. 1998). Consistent with the early chimera / embryo reconstitution experiments, TSCs could be derived from the ExE of E6.5 embryos, but also from E3.5 blastocysts (Tanaka, Kunath et al. 1998) and particular niches in embryos up to E8.5 (Uy, Downs et al. 2002). Interestingly, these different sources converge towards an *in vitro* stem cell state that is morphologically indistinguishable and with an equivalent cellular potency (Tanaka, Kunath et al. 1998). Nevertheless, to date there haven't been studies that systematically address these similarities – it would be important to actually perform transcriptomics and epigenomics analysis of TSC cell lines derived from different niches. Similarly, there are no rigorous assessments of developmental potency and the derivatives of each of these *in vitro* cell lines towards the different trophoblast cell types.

Most published work shows a robust integration of TSC-descendants into the labyrinth layer of the placenta, as well as spongiotrophoblasts and giant cells (Tanaka, Kunath et al. 1998, Erlebacher, Price et al. 2004, Kubaczka, Senner et al. 2014, Ohinata and Tsukiyama 2014). This analysis is mainly performed based on the presence of GFP-labelled TSC descendants within the placenta structures. These would have to be verified by immunostaining with markers for the different trophoblast cells to truly consider cellular identity (see Figure 1.6). Alternatively, chimera assays followed by sorting of TSC-descendant cells at different developmental points and single-cell RNA sequencing would help dissect the precise differentiation potential of TSCs.

Interestingly, *in vitro* differentiation of TSCs has a bias towards EPC and TGC-fate, with very limited differentiation towards chorion and labyrinth lineage (Tanaka, Kunath et al. 1998, Kubaczka, Senner et al. 2014, Benchetrit, Herman et al. 2015, Kubaczka, Senner et al. 2015). Further, unlike pluripotent stem cells, which are capable to generate all-ESC chimeras, *in vivo* contribution of TSCs to the placenta is generally formed by small patches of cells, with few reports of extensive contribution and, so far, none of all-TSC placentas. In one hand, limitations in *in vitro* differentiation towards chorion and labyrinth layers might be a result of our limited understanding of the *in vivo* interplay between embryonic and extraembryonic signalling driving trophoblast differentiation. On the other hand, the lack of extensive contribution to placenta from TSC might imply that

the culture conditions to maintain a stem cell state with higher potency is still to be discovered.

Assuming that morphology and some evidence of contribution to different layers in the placenta (Tanaka, Kunath et al. 1998) actually reflect similar *in vitro* states for TSCs derived from E3.5 or E6.5, then it is worth asking if FGF signalling selects for a particular cell state rather than recapitulating the original embryonic populations.

Recent studies demonstrated that TSC cultures are heterogeneous both in morphology (Motomura, Oikawa et al. 2016) and expression of core transcription factors *Cdx2*, *Eomes* and *Elf5* (Kuales, Weiss et al. 2015). These seem to reflect the co-existence of undifferentiated TSC colonies, together with colonies which are partially differentiated. With single-cell RNA-sequencing, it is now possible to precisely assess *in vitro* models and evaluate the resemblance to their embryonic counterparts. A new study has finally addressed TSC culture heterogeneity and compared it with the expression profile of single mural or polar TE cells (Javier Frias-Aldeguer 2019). The authors confirm culture heterogeneity and classify subpopulations according to Cdx2 expression levels, showing that $Cdx2^{High}$ cells have strong expression of self-renewal transcription factors and reflect a transcriptome that resembles polar TE. On the other hand, $Cdx2^{Low}$ cells express higher levels of trophoblast differentiation markers and their transcriptome shows they are differentiating cells. Comparison of TSC subpopulations with differentiating TSCs (through withdrawal of Fgf4 signalling) revealed the TSC cultures possess cells that reflect both pre- and postimplantation trophoblasts. With these differences in mind, the authors then screened for media formulations that would maintain a homogeneous $Cdx2^{High}$ population. Upon optimisation they were able to derive polar-like TSCs (pTSCs) that resemble preimplantation TE with clonogenicity equivalent to naïve ESCs (Javier Frias-Aldeguer 2019). Importantly, *in vivo* assays demonstrated these could contribute extensively to the placenta, although further molecular analysis will be necessary to determine exact cells types and comparison to "conventional" TSCs, as well as reproducibility.

Comparison of TSC with postimplantation cell types hasn't been done yet, so the question remains if these cell lines simply lack signalling to select for undifferentiated stem cell populations, or if they reflect different *in vivo* counterparts. The discovery of pTSCs might resemble a ground-state TSC and opens the possibility to fine-tune culture conditions that would capture homogeneous TSC populations. Like pluripotency, we might be on the path to discover cell populations which recapitulate different developmental stages for the trophoblast lineage and represent preimplantation and postimplantation trophoblast states. It is now time to re-visit the hierarchy of the TSC

lineage tree and evaluate if pTSC could represent a polar TE cell state that progresses in development through a postimplantation TSC state. Further, it is pertinent to ask whether it is possible to derive pTSC lines from later developmental stages, as seen for TSCs. Can this be a stem cell population preserved through development, or as seen for ground-state pluripotency, does it dissolve and acquire different states at the onset of differentiation?

Technological advances bring new tools for the dissection of the trophoblast lineage. It is an exciting time to reassess our knowledge of this key lineage and *in vitro* stem cell models.

1.4 Mammalian Genome Engineering

1.4.1 Gene targeting and Homologous Recombination

Classical approaches for specific disruption (commonly referred as knockout) of a gene used gene targeting inactivation via homologous recombination to assess how the specific genotype influences a phenotype (Capecchi 2005). The experimental demonstration that ESCs could be maintained in culture preserving pluripotency and additionally contribute to germline transmission in mice (Evans and Kaufman 1981, Martin 1981, Bradley, Evans et al. 1984, Robertson, Bradley et al. 1986) prompted the rapid development of gene targeting in mammalian ESCs. ESCs provided a unique means to generate and select specific mutations *in vitro*, that could then be transferred to mice for *in vivo* evaluation of gene function.

Early experiments in yeast demonstrated that exogenous sequences with homology to chromosomal regions could be specifically introduced in the host genome by homologous recombination (Hinnen, Hicks et al. 1978, Orr-Weaver, Szostak et al. 1981). The observation that microinjection of DNA in mammalian cells resulted in integration of multiple copies of the transforming gene at very few *loci*, and in highly ordered head-to-tail concatemers provided the first evidence for homologous recombination in this cell type (Folger, Wong et al. 1982). The first gene targeting experiment in cultured mammalian cells was done in 1984 by Lin and Sternberg, using the herpes simplex virus thymidine kinase (*tk*) gene (Lin and Sternberg 1984). In 1985 the first targeting of an endogenous locus was reported and consisted of the introduction of a neo^R (neomycin resistance gene) selection cassette into the human β-globin locus, using a 11.1 kb (kilobase pairs) homology cassette (Smithies, Gregg et al. 1985). Using similar

strategies, the *Hrpt1* locus (hypoxanthine phosphoribosyl transferase), located in the X-chromosome, was also readily targeted in male ESCs (Doetschman, Gregg et al. 1987, Thomas and Capecchi 1987). This process occurred with an efficiency of 1 to 1000, comparing homologous recombination-mediated insertion to random integration, which highlights one of the key limitations for gene targeting. Importantly, genetic manipulation in ESCs followed by drug selection, did not affect their capacity to confer germline transmission in chimeras (Koller, Hagemann et al. 1989, Schwartzberg, Goff et al. 1989, Thompson, Clarke et al. 1989).

The disruptive capacity to modify the mammalian genome by homologous recombination in ESCs, pioneered by the laboratories of Mario Capecchi, Oliver Smithies and Martin Evans, was awarded a Nobel Prize in Physiology or Medicine in 2007. Since their discovery, the application of gene targeting technologies has exploded, both for *in vitro* assessment of gene function, but also *in vivo* with large projects such as the international mouse knockout consortium (International Mouse Knockout, Collins et al. 2007). This consortium aimed to disrupt every gene in the mouse genome and has already produced about 7400 different knockout mouse strains, many of which have already been phenotypically analysed.

1.4.2 Genome Editing and the use of engineered chimeric nucleases

Despite the impact of gene targeting technologies in genetics, the absolute rates for homology-directed repair (HDR) efficiency remain low, and random integration is more frequently the outcome following transfection of DNA. Although specific versus random insertions can be selected and screened for in ESCs (Mansour, Thomas et al. 1988), the low efficiency represented a limitation for other cell types, such as human induced pluripotent stem cells (hIPSCs) and hESCs (Tenzen, Zembowicz et al. 2010), other organisms and for therapeutic applications.

The discovery that a rare-cutting endonuclease, I-SceI, could introduce double-strand breaks (DSB) in the DNA and stimulate gene targeting by activation of the endogenous DNA repair machinery (Rouet, Smih et al. 1994) paved the way for a new generation of genome editing technologies. Repair of the DSB happened either via HDR-mediated homologous recombination, or via non-homologous end-joining (NHEJ) pathway, resulting in junctions with insertions or deletions (indels) of nucleotides with different sizes. NHEJ-induced indels can efficiently disrupt genes via frame-shifting mutations that alter protein sequence and often introduce early stop codons. On the other hand,

HDR can be used for precise gene targeting mutations when combined with a custom homologous “donor” template, allowing a panoply of different modifications from large insertions and deletions to point mutations or gene replacement.

The next challenge was to confer specificity to control the target site of DSB and allow gene targeting at precise locations throughout the genome. This bottleneck was addressed by the development of engineered nucleases, composed of a DNA-binding domain that confers sequence specificity, fused to a nonspecific nuclease which introduces a DSB at the specific target. Three major classes for these chimeric programmable nucleases have been developed to date: meganucleases (Smith, Grizot et al. 2006, Silva, Poirot et al. 2011), zinc finger nucleases (ZFNs) (Urnov, Rebar et al. 2010, Carroll 2011) and transcription-activator-like effector nucleases (TALENs) (Miller, Tan et al. 2011, Gaj, Gersbach et al. 2013).

Meganucleases are derived from microbial mobile genetic elements and represent a family of naturally occurring rare-cutting endonucleases, that can recognise DNA sequences of 14-40 bp through non-modular protein-DNA interactions (Smith, Grizot et al. 2006). Meganucleases have successfully been used for targeted genome editing, but the lack of clear correlation between protein sequence and DNA specificity posed challenges in their engineering and thus they were not widely adopted by the scientific community (Smith, Grizot et al. 2006, Silva, Poirot et al. 2011).

ZFN combine the use of the zinc finger domain found in eukaryotic transcription factors, which confers DNA sequence specificity, with the FokI nuclease catalytic domain, which induces a DSB (Kim, Cha et al. 1996, Bibikova, Carroll et al. 2001). This technology was first used in *Drosophila melanogaster* (Bibikova, Golic et al. 2002, Bibikova, Beumer et al. 2003). An individual zinc finger consists of approximately 30 amino acids in a $\beta\beta\alpha$ configuration (Beerli and Barbas 2002), with some amino acids in the α-helix recognising specific sequences of 3 bp in DNA. Early studies used three fingers to bind 9-bp targets, which enabled ZFN dimers to recognise 18-bp per cleavage site. ZF can be assembled in a modular fashion, combining individual zinc fingers recognizing triplets of DNA and linking them into a multifinger peptide to recognise specific DNA regions (Segal, Beerli et al. 2003). ZFN are active as dimers and are engineered so that each monomer confers sequence specificity to either left or right side of the desired target site, separated by a 5 – 7 bp spacer that is then recognized by the *FokI* cleavage domain.

ZFNs are widely used in research and have been used for targeting in several organisms including human (Li, Haurigot et al. 2011, Gaj, Guo et al. 2012), zebrafish

(Doyon, McCammon et al. 2008, Meng, Noyes et al. 2008), rat (Geurts, Cost et al. 2009), *Arabidopsis* (Zhang, Maeder et al. 2010) or *C. elegans* (Wood, Lo et al. 2011), among many others.

TALE proteins were found in plant pathogenic bacteria from the genus *Xanthomonas* and are characterised by DNA binding domains composed of a series of 33-35 amino acid repeat domains, each recognising a single DNA base pair (Boch, Scholze et al. 2009, Moscou and Bogdanove 2009). TALE specificity is determined by two hypervariable amino acids, also known as the repeat-variable di-residues (RVDs), requiring a code of 4 different repeat domains to recognize the 4 DNA nucleotides (Deng, Yan et al. 2012, Mak, Bradley et al. 2012). The single base recognition of TALE domains made them an attractive tool for researchers, due to easier protein engineering as well as higher targeting flexibility compared to the triplet-confined zinc finger domains. TALE domains were quickly developed for different genetic modifications including fusion to nucleases (Christian, Cermak et al. 2010, Mussolino, Morbitzer et al. 2011), transcription activators and repressors (Maeder, Linder et al. 2013, Gao, Tsang et al. 2014, Thakore and Gersbach 2016), and site-specific recombinases (Mercer, Gaj et al. 2012).

1.4.3 CRISPR/Cas9 Genome editing

Both ZFNs and TALENs have revolutionised genome engineering and allowed precise gene targeting in countless organisms and applications from basic biology to gene therapy. In fact, the use of ZFNs to disable the HIV receptor CCR5 in primary T cells (Perez, Wang et al. 2008) and hematopoietic stem/progenitor cells (Holt, Wang et al. 2010) is the basis of three clinical trials (NCT01252641, NCT00842634 and NTC01044654). One of the major limitations for TALENs and ZFNs, however, is their DNA binding mechanism which is based on protein-DNA recognition motifs, making their engineering for target specificity very labour intensive, costly and time-consuming.

The discovery of CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR associated) systems provided a new DNA recognition mechanism, based on a short guide RNA sequence that leads the Cas9 nuclease to its target DNA. The simplicity of DNA recognition makes it a fast, inexpensive and extremely versatile system, which has propelled CRISPR/Cas9 to the centre of a new generation of genome engineering tools.

Native CRISPR/Cas systems are adaptive immunity mechanisms developed by prokaryotes and archaea to shield themselves from foreign genetic elements such as viruses and plasmids (Mojica, Diez-Villasenor et al. 2005, Barrangou, Fremaux et al. 2007). Adaptive immunity conferred by CRISPR/Cas systems occurs in three stages: i) small fragments of foreign DNA are inserted as a spacer sequence into the CRISPR array locus; ii) transcription of precursor crRNA (CRISPR RNA) that produces individual crRNAs upon maturation, each composed of a repeat sequence and the foreign-DNA targeting spacer; iii) crRNA drives Cas9 proteins to foreign DNA molecules with sequence complementary to the targeting spacer, where it promotes their degradation.

The CRISPR/Cas system can be classified into two main classes, each with three different types (type I – IV) (Mohanraju, Makarova et al. 2016). Types I, II and III use different molecular machinery to promote DNA recognition and cleavage (Makarova, Aravind et al. 2011, Makarova, Haft et al. 2011), with types I and III using large complexes of Cas proteins to mediate DNA cleavage (Brouns, Jore et al. 2008, Haurwitz, Jinek et al. 2010), whereas type II requires a single Cas protein for its action (Gasiunas, Barrangou et al. 2012, Jinek, Chylinski et al. 2012).

The CRISPR/Cas9 technology is derived from type II CRISPR/Cas system. It uses three components: i) a crRNA that confers target specificity; ii) a trans-activating crRNA (tracrRNA) which contains short sequences complementary to the crRNA and is also required for Cas9 catalytic activity; and iii) a single multi-domain effector protein, Cas9 which contains nuclease activity for cleavage at the target site. In this system, the Cas9 endonuclease assembles a ribonucleoprotein (RNP) complex with an RNA duplex formed by tracrRNA:crRNA, which is then directed to the target site for induction of a DSB mediated by Cas9. Importantly, target recognition by Cas9 requires a protospacer adjacent motif (PAM), which is a short sequence of 2 – 6 nucleotides, immediately following the crRNA-specified target site. Different species require different PAM sequences (reviewed in (Doudna and Charpentier 2014, Hsu, Lander et al. 2014)). Most of the current applications of CRISPR/Cas9 systems employ the Cas9 protein from *Streptococcus pyogenes* (SpCas9), which requires a 5'-NGG or, at lower editing efficiency, 5'-NAG as PAM sequences.

A major discovery that simplified the adoption of CRISPR/Cas9 for mammalian genome editing was that the crRNA:tracrRNA duplex could be engineered as a single molecule, known as a guide RNA (gRNA) (Jinek, Chylinski et al. 2012) (Figure 1.9). After this discovery, CRISPR/Cas9 was adapted for genome engineering in human and mouse cells by modulating the 20 nucleotide guide sequence in the gRNA to confer specificity

to the desired target and efficiently inducing DSB in mammalian genomes with sequence specificity (Cong, Ran et al. 2013, Jinek, East et al. 2013, Mali, Yang et al. 2013).

Since its advent as a genome engineering tool, CRISPR/Cas9 has been used to edit in genomes from multiple organisms such as zebrafish (Shah, Davey et al. 2015), plants (Feng, Zhang et al. 2013) and *C. elegans* (Friedland, Tzur et al. 2013), as well as a plethora of different cell types. Additionally, similarly to TALEs, it has also been adapted for gene activation and repression using a catalytically inactive version of Cas9 fused with activation or repression domains (Kiani, Chavez et al. 2015, Konermann, Brigham et al. 2015, Thakore, D'Ippolito et al. 2015).

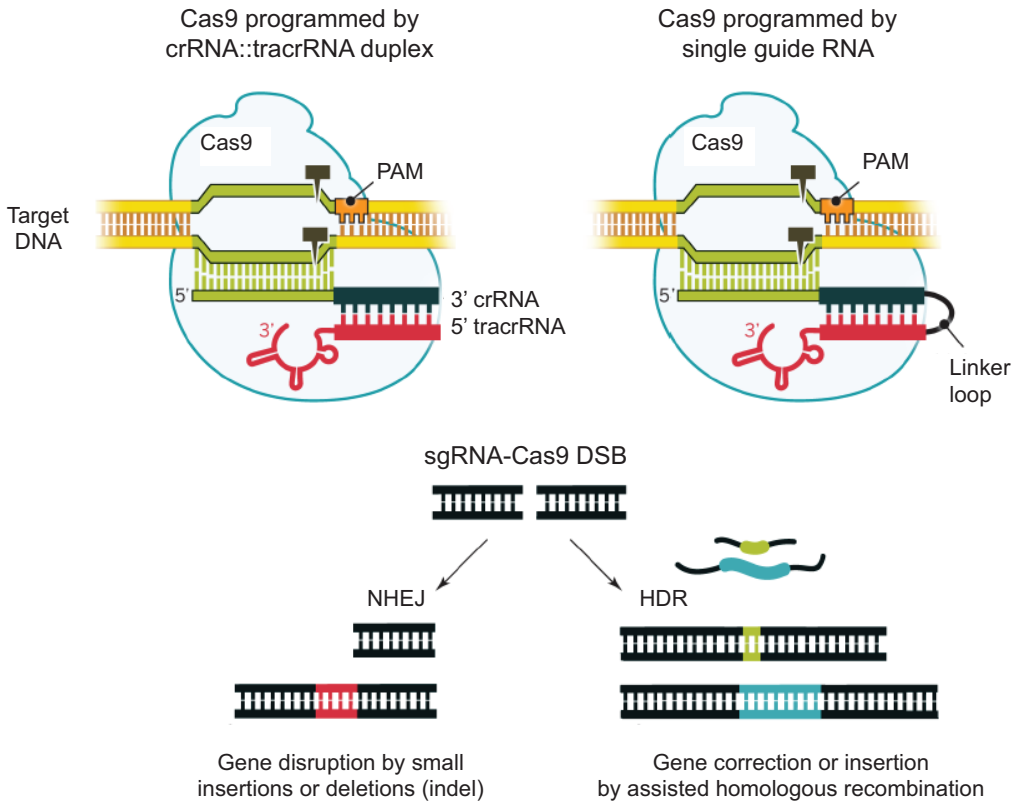

Figure 1.9 | Schematic representation of type II CRISPR/Cas9 mediated genome editing. Cas9 nuclease can be programmed with either an RNA duplex generated by sequence complementarity in crRNA and tracrRNA or a single chimeric guide RNA. The ribonucleoprotein complex is then targeted to the region defined by the 20nt sequence in the crRNA or guide RNA, where it mediates DNA double strand break (DSB). Upon DSB, the cellular DNA repair machinery corrects it via NHEJ pathway, generating small indels, or via HDR pathway, leading to precise mutations in the presence of a donor template. Figure adapted from (Doudna and Charpentier 2014).

1.5 Thesis Aims

The totipotent zygote has the capacity to generate all tissues that make an adult organism, including all embryonic lineages that constitute the embryo proper, but also all extra-embryonic tissues that support embryo development. As development proceeds, the cells that compose the embryo progressively lose potency (the ability to differentiate towards different cell types) and acquire fate specification towards the different organs/tissues that constitute a developed organism. In this context, cell lineage denotes the developmental trajectory of a specific mature tissue from its origin in the zygote, compiling a lineage tree that describes all the intermediate cellular stages that led to the final cell type.

Ultimately, all cells in a mammalian organism possess the same genetic information (DNA) and descend from a single cell. It is the series of lineage decisions during development that dictate their mature cell type and specific function. Lineage decision is then defined as a process in which a cell with higher potency irreversibly commits to one cellular lineage over another, resulting in a state with lower potency and higher fate/functional specification. It is accomplished through an initial stage of lineage/fate commitment, whereby the cell activates key determinants of a specific lineage identity and blocks alternative fates. The decision is then complete through the maintenance stage, in which different regulatory mechanisms maintain the cellular state acquired and efficiently restrict its potency, in an irreversible manner under physiological conditions. This way, once a cell commits to one lineage, its descendants lose the capacity to adopt a different cell fate (trans-differentiation) or the initial stage with higher potency (reprogramming). The molecular mechanisms in place to ensure this restriction are referred to as “lineage barrier” or “lineage restriction” in this thesis.

The first cell lineage decision in the mouse embryo specifies the pluripotent ICM and the extraembryonic TE. The cell commitment established at this stage is preserved in the two stem cell populations derived *in vitro* from these layers, ESCs and TSCs, which retain the potency and lineage restrictions of their embryonic counterparts. Hence, once injected into carrier host embryos, ESCs, like pluripotent ICM cells, contribute to all the lineages in the embryo proper, but fail to integrate the extraembryonic layers such as the placenta. Conversely, TSCs, similarly to TE cells, are restricted to the extraembryonic lineage and cannot integrate the embryo proper. Consistent with *in vivo* experiments, ESCs also resist *in vitro* differentiation towards a TSC fate and to date a limited number of genes have been identified as suppressors of this conversion. These *in vivo* and *in*

vitro experiments exemplify the strong lineage restriction established upon ICM/TE segregation in the embryo, and recapitulated *in vitro* in ESCs and TSCs.

The molecular mechanisms maintaining the extraembryonic restriction in the pluripotent compartment remain largely unknown and are the primary focus of this thesis. Here, I aim to explore the ESC-TSC lineage barrier in an unbiased approach, using a CRISPR/Cas9 genome-wide loss-of-function screen to identify genes supressing ESC differentiation to TSC. Such a screen will identify a network of genes regulating the maintenance of the first cell lineage decision, and potentially yield a refreshed perspective of the molecular players in this early developmental process.