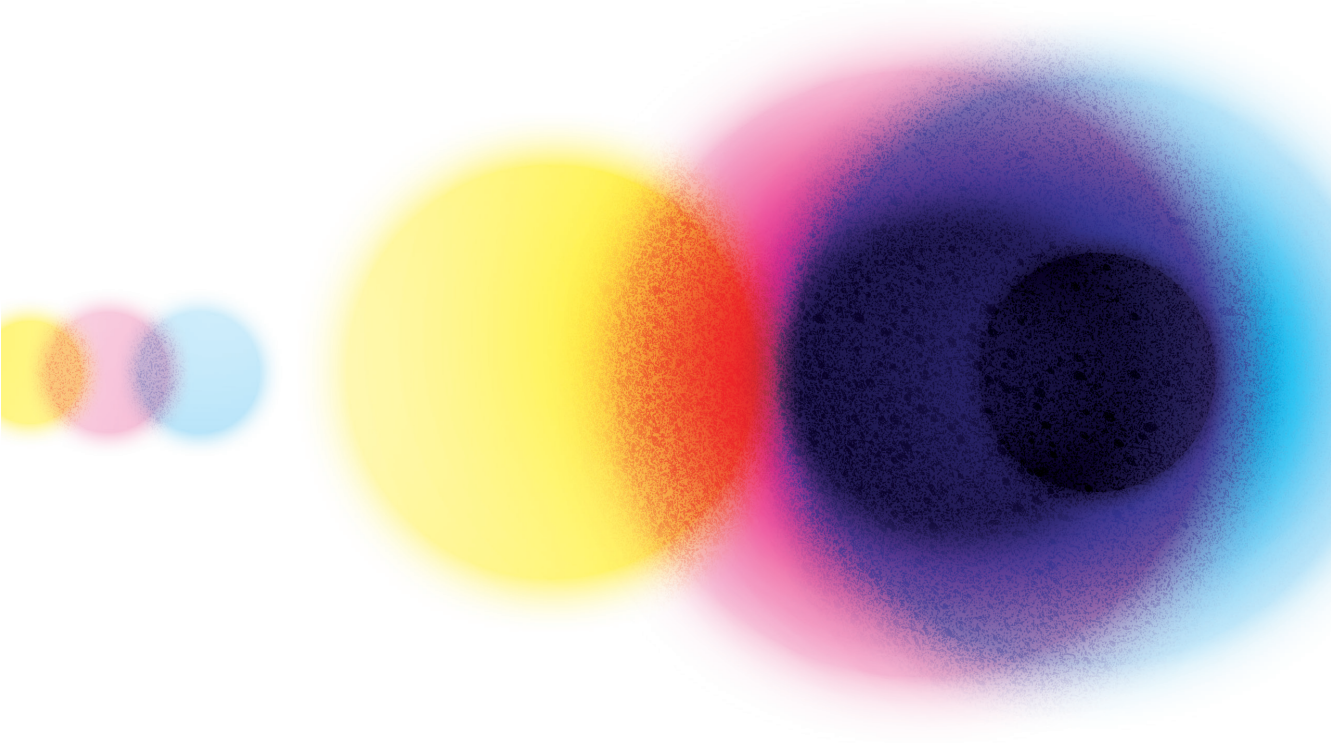
UNIVERSITAT DE BARCELONA

Alternative splicing control over naïve and primed pluripotency

Victoria Eugenia Rodríguez Vaello



Alternative splicing control over naïve & primed pluripotency

Victoria Eugenia Rodríguez Vaello

Alternative splicing control

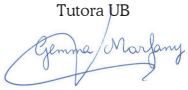
OVER NAÏVE & PRIMED PLURIPOTENCY.

Memoria de **Victoria Eugenia Rodríguez Vaello**, optando al grado de doctora por la Universidad de Barcelona en 2020.

Dirigido por Dr. Manuel Irimia Martinez
y tutorizado por Dra. Gemma Marfany.

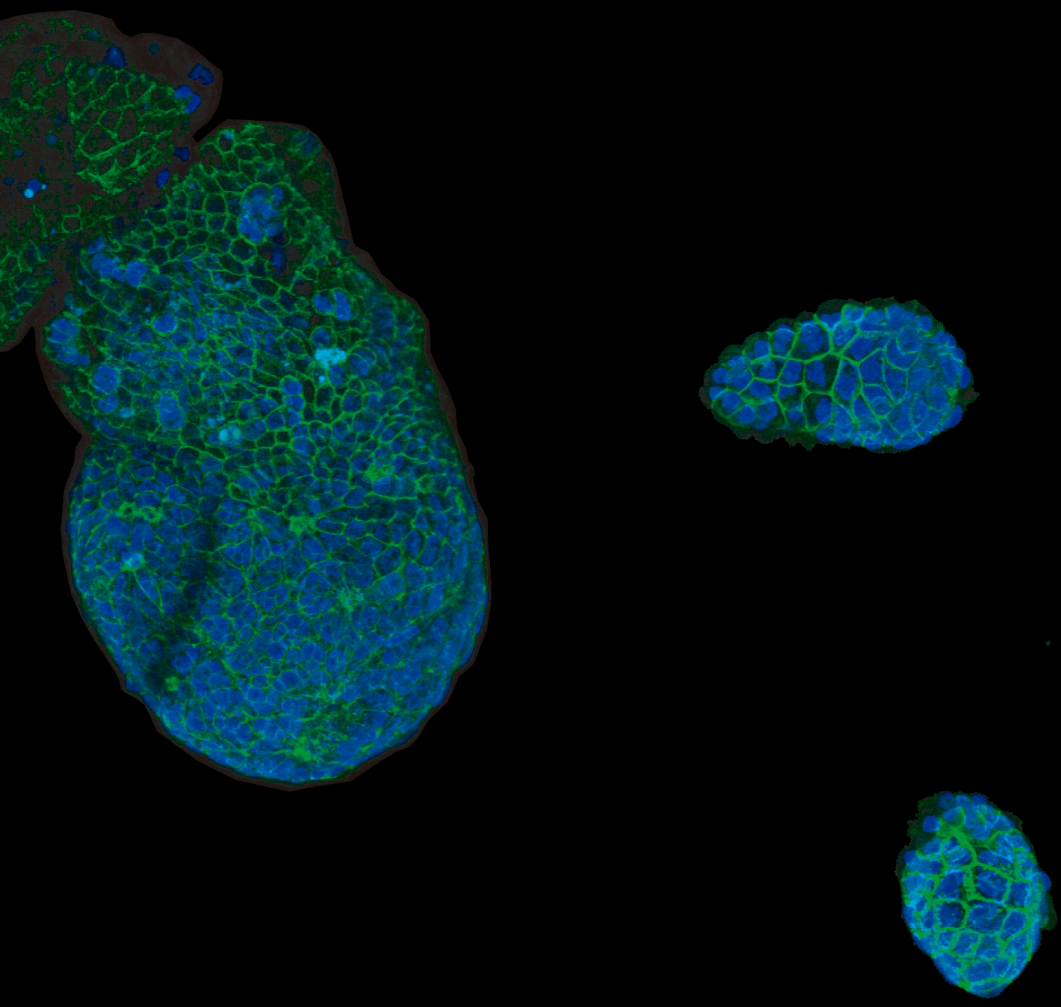
Programa de Doctorado en Biomedicina.
Universidad de Barcelona.
Centro de Regulación Genómica (CRG)

Estudiante | Tutora UB | Director


CRG Centre for Genomic Regulation

Universitat de Barcelona

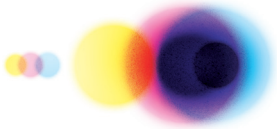


Above: E6.5 embryo and two E4.5 embryos immunostained for E-cadherin (green) and hnRNPK (blue)
Back cover: E6.5 embryo immunostained for OCT4 (red) and alpha-Tubulin (yellow) and nuclear-stained with Hoechst
Cover & template designed by Raquel Ramirez

Suppose it were perfectly certain that the life and fortune of every one of us would, one day or other, depend upon his winning or losing a game of chess. Don't you think that we should all consider it to be a primary duty to learn at least the names and the moves of the pieces; to have a notion of a gambit, and a keen eye for all the means of giving and getting out of check?

Thomas H. Huxley

Abstract



A fundamental step in embryo development is the transition through pluripotency. There are two known sequential states of pluripotency termed naïve and primed pluripotency. Considerable efforts have been made to characterize these two cell identities using gene expression, protein, metabolic and epigenetic profiling; here we explore an additional regulatory layer, at the level of alternative pre-mRNA splicing (AS). We performed deep coverage RNA sequencing on embryo-derived stem cell lines and identified a regulated AS landscape between the two states involving over 900 splicing events. These events show distinct *cis*-regulatory features and can also be observed *in vivo* and in the equivalent human pluripotency states. Differential gene expression analyses helped to identify two regulatory axes that contribute to shape these AS landscapes: the Rbmx/RbmxL2 and the Rbm47/Esrp axes. We further performed a high-throughput screen to test the effect of 1,500 single knockdowns of RNA binding proteins and chromatin modifiers on naïve-to-primed AS switches. This led us to the identification of Qki as an additional master naïve-to-primed AS regulator. Functional assays show that depletion of these splicing regulators affect the differentiation dynamics of embryonic stem cells. Overall, our results provide new insights into the regulatory programs governing naïve and primed pluripotency and the gain of differentiation potential.

Index

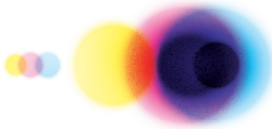


Table of Contents

Introduction

Results

Discussion

Abbreviations

Abbreviations

2iL – 2 inhibitors (MEK inhibitor and GSK3 inhibitor) and LIF
AJs – Adherens junctions
AP – Alkaline Phosphatase
AS – Alternative Splicing
ASSEA – Alternative splicing enrichment analysis
bp – base pair
C-to-U RNA editing – Cytidine-to-Uridine RNA editing
CLIP – Crosslinking and immunoprecipitation
cRPKMs – corrected-for-mappability Reads per Kilobase pair and Million mapped reads
DKO – Double knockout
EBs – Embryoid Bodies
EMT – Epithelial to mesenchymal transition
EpiLC – Epiblast-like stem cells
EpiSC – Epiblast stem cell
EPS – Extended pluripotent stem cell
EPSC – Expanded potential stem cell
ESC – Embryonic stem cell
gRNA – CRISPR/Cas9 guide RNAs
GSEA – Gene Set Enrichment Analysis
ICM – Inner Cell Mass
KD – Knockdown
KO – Knockout
m6A – N6-Methyladenosine
miRNAs – MicroRNAs
mRNA – messenger RNA
NLS – Nuclear localization signal
nt – nucleotide
NTD – Nascent transcript targeting domain
PGCs – Primordial germ cells
PIR – percentage of transcript with intron retained
PrE – Primitive Endoderm
PSI – Percentage of Spliced in
RBP – RNA binding protein
RGG – Arginine-glycine-glycine
RNA-seq – RNA sequencing
RRM – RNA recognition motif

RT-PCR – Reverse transcription polymerase chain reaction
RTqPCR – Reverse transcription quantitative polymerase chain reaction
SL – Serum and LIF
SPAR-seq – Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing
TE – Trophectoderm
TKO – Triple knockout
TSCs – Trophoblast stem cells
UTR – Untranslated region
XEN – Extraembryonic endoderm

Introduction

The rise and fall of pluripotency: embryogenesis at the peri-implantation stage

Throughout embryogenesis, a meticulous program of coordinated events enables a single cell to develop into a multicellular organism. A critical step in this process is the establishment and exit from pluripotency. Pluripotency is defined as the potential to differentiate into any somatic and germ cell lineage but not into extraembryonic tissues. In placental mammals, this state arises in the epiblast cells of the early embryo over the course of implantation. Shortly after implantation, these epiblast cells will arrange into a unilaminar epithelium that will start the differentiation process into all tissues in the embryo proper through gastrulation and subsequent organogenesis. Consequently, pluripotency is lost.

The pluripotent epithelial structure of the epiblast is conserved among all amniotes, however its morphogenesis can follow different strategies[1]. Within amniotes, placental mammals are unique in that their embryos are nourished by interfacing with the maternal reproductive tract. For this purpose, during embryogenesis, embryos must synchronize epiblast formation with the formation of extra-embryonic tissues (trophoblast and hypoblast) specialized in mediating their implantation and securing their nutritional supply. Despite the differences that this specialization entails (mainly to do with yolk reduction and placentation), the specification of the epiblast as a pluripotential cell lineage is also synchronized with the appearance of extraembryonic tissues in all other amniotes[1,2]. It is therefore considered that the earliest phylotypic stage in amniote embryogenesis shares two characteristics: the epithelialization of the pluripotent epiblast and the specification of the embryonic versus extraembryonic lineages[1].

In the coming pages, I will describe the appearance, progress and loss of pluripotency during mouse embryogenesis. This period represents approximately one quarter of the total murine gestation period, starting after two waves of lineage specification that segregate pluripotent epiblast cells from extraembryonic lineages, and finishing upon gastrulation. Bear in mind that although several of the features described here are conserved among other placental mammals, many others are not, thus caution must be taken when extrapolating species-specific features.

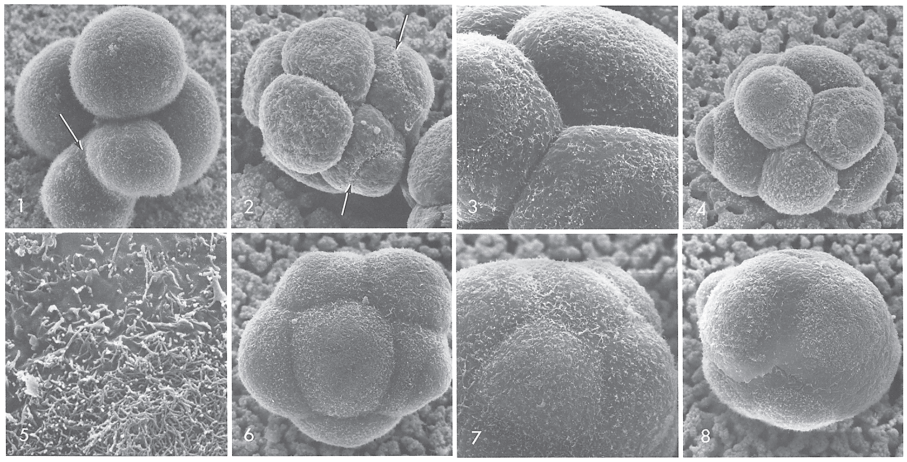
Preimplantation: the dawning of lineage restriction

Lineage restriction is a complex phenomenon that dubiously occurs in a single-step. Instead, it is considered to have three phases: an initiation phase, a commitment phase and a maintenance phase, reviewed in[3]. At the initiation phase, differences in cell states will emerge in a homogenous cell population driven by external cues or by stochastic fate choices[4]. Here, cells will remain receptive and interconverting to different states in response to inducing signals. In the commitment phase, specific fate programs are activated, cells are no longer as malleable and do not respond to the initial inducing stimuli. These fate programs mostly become dependent on cell-intrinsic cues through the activity of early response transcription factors and epigenetic regulators[3,5]. These act primarily by strengthening their expression through feedback-loops or by activating downstream late response transcription factors that will later become the hallmarks of that cell lineage, reviewed in[3,6]. Finally, in the maintenance phase, cells must lock the expression of these hallmark genes and avoid deviating into other cell states. This, however, does not necessarily mean that during the maintenance phase cells remain refractory to differentiation cues. In fact, in many cases during development, the transcription of many differentiation genes remains poised, waiting for an activating cue. Thus, being in a maintenance phase of lineage restriction does not necessarily mean being non-responsive and we will see examples of this in the following sections.

First cleavage divisions

After fertilization of an oocyte by a sperm, a 1-cell embryo termed zygote is generated. This zygote is totipotent by definition because it can produce all embryonic as well as extraembryonic lineages and can organize them into a coherent body plan[7]. The zygote will then travel through the oviduct as it undergoes the first five cell cycles, known as cleavage divisions. Cleavage takes place with no interphasic cell mass growth[8] and relies on the vast cytosolic compartment that the oocyte had. Therefore, cells at these stages, termed blastomeres, become progressively smaller with each division. Undeniably, individual 2-cell blastomeres can form a full conceptus as demonstrated by the existence of dichorionic identical twins and confirmed elegantly in mouse embryogenesis by Tarkowski who in 1959, inspired by seminal work in mammals using rabbit[9] and rat[10] blastomeres, obtained viable embryos after destruction of one of the two blastomeres at the 2-cell stage[11].

There is clear evidence that the blastomeres generated after the two subsequent cell divisions (4-cell and 8-cell stages) show molecular differences between them[12–18], yet cell fates remain plastic and thus, blastomeres remain totipotent[19]. In fact, the 8-cell embryo forms at around embryonic day 2.5 after fertilization (E2.5) and individual blastomeres will remain totipotent until the first lineage decision.



Introduction Figure 1: Electron microscopy showing changes in cell surface during compaction of mouse embryos, the first manifestation of cell polarity in embryogenesis. (1) Four-cell embryo with a dividing blastomere with a clear cleavage furrow (arrow) showing all blastomere surfaces uniformly coated with microvilli. x1500 magnification **(2)** Eight-cell embryo commencing compaction showing blastomeres flattening with each other and some apical microvilli (arrow) located in the grooves of their membrane. x1700 magnification **(3)** Blastomeres of a compacting embryo showing microvilli in the apices between two blastomeres. x3200 magnification **(4)** Apical localization of microvilli in 4 blastomeres of a compacting eight-cell embryo. x1500 magnification **(5)** Surface of a compacting blastomere at the border region between the smooth area (top left) and the apical localization of microvilli (bottom right). x8000 magnification. **(6)** Embryo close to compaction completion with few areas of smooth membrane remaining exposed. x1600 magnification **(7)** Compacted embryo with no visible smooth surfaces. x2200 magnification **(8)** Compacted embryo with smooth areas between cell-cell contacts and apical localizations of microvilli. x1500 magnification. All images are taken from Ducibella et al. 1977[21].

First lineage decision: inner cell mass & trophectoderm

This begins with compaction during the morula stage at E2.75, followed by segregation of the inner cell mass (ICM) and the trophectoderm (TE) in the blastocyst at E3.5. The outer cells of the morula are biased towards becoming the first extraembryonic lineage, the TE, which is then pivotal for the implantation process. The TE will develop into the foetal part of the placenta, key for nourishment and detoxification but also key in the induction of differentiation of certain lineages in the embryo[20]. On the other hand, the inner cells of the morula are biased towards becoming the ICM.

First, during compaction blastomeres that were once loosely adhered to one another, broaden their cell-cell contacts through E-CADHERIN mediated adherens junctions (AJs) resulting in close intercellular apposition and triggering blastomere apico-basal polarization. The apical domain (outward domain) will increase its contents on F-actin, myosin, clathrin, cell-surface glycoproteins and microvilli[21–25] (Introduction figure 1).

After compaction and polarization there are two rounds of cell divisions. Depending on their relative cleavage position to the apico-basal plane, these divisions can either be symmetrical or asymmetrical and daughter cells will be polarized or not accordingly. Consequently, polarized cells will remain outside and maintain an apical domain while apolar cells will remain inside and have a higher degree of cell contacts[26]. Simultaneously, the differential cellular localization of the transcriptional co-activator YAP1 among blastomeres, will tune their positional information into lineage specification[27]. The absence of an apical domain triggers cell-cell contact-dependent YAP1 cytoplasmic localization and degradation through E-CADHERIN mediated adherens junctions[28]. Conversely when an apical domain is present, YAP1 is nuclear and can upregulate the expression of the TE-specific master transcription factor CDX2 in outside blastomeres[27,29]. This way key positional information (inside versus outside) is translated into ICM or TE- specific transcriptional programs.

Single-cell profiling of transcription factors at 16- and 32-cell stage embryos first showed that SOX2 and ID2 were the earliest markers of inner and outer cells respectively[30]. Additionally, as segregation of the ICM and the TE occurs, the expression of CDX2 is gradually lost in inside cells and becomes upregulated in outside cells. Although CDX2 expression is not essential in the initial specification of the TE lineage, it is required for its later commitment[29,31,32].

Antagonistically, ICM cells maintain the expression of the transcription factor OCT4 while it is lost in TE. OCT4 is essential for the pluripotential identity of ICM cells as *Oct4*-null ICM cells fail to maintain that identity and differentiate into TE cells[33]. Importantly, although OCT4 together with SOX2 coordinate pluripotency gene expression, at this stage, the full pluripotency network has yet to be stabilized and these cells will keep a certain degree of developmental plasticity until the second lineage decision[34,35].

At around E3.25, the blastocoel cavity appears. This constrains the ICM to one pole and the embryo becomes an early blastocyst. At this stage, the 32-cell stage, the transcriptomes of individual ICM cells are indistinguishable, but modest transcriptional changes arise within the following hours[30,36]. At that moment, the second lineage specification starts.

Second lineage decision: Epiblast & primitive endoderm

During the second lineage decision, the ICM segregates into the pluripotent epiblast and the primitive endoderm (PrE). The epiblast will later develop into the embryo proper whereas the PrE will be a vital constituent of the yolk sac and provide crucial signals for embryo patterning, reviewed in[37]. Whilst for the first lineage decision, positional cues were decisive; the second lineage decision relies upon stochastic cell-to-cell expression heterogeneity and gradual acquisition of transcriptional identity through signal reinforcement[36]. This is directed, but not originated, by fibroblast growth factor (FGF) signalling, reviewed in[38,39]. Interestingly, a recent study showed that blastocoel cavity expansion also seems to play a critical role in specifying epiblast and PrE cell fates and cell positioning within the blastocyst since reduced expansion impairs lineage specification and spatial lineage sorting in an FGF-signalling dependent manner[40].

Precisely, the drivers of this second lineage restriction are three FGF signalling components: on one side, the receptors FGFR1 and FGFR2, and on the other side, their ligand, FGF4[41–44]. At E3.25, before cells have made a fate choice, FGFR1 is expressed across all ICM cells whereas FGF4 already shows a bimodal distribution and from then on, its expression will be restricted to the epiblast lineage[36,43,44]. FGFR2 is expressed later, by E3.5, and only in PrE-biased cells[36]. Importantly, since all ICM cells possess FGF receptors, they are all receptive to FGF4, but unique feedback from the receptors generates differential FGF signalling activity in each ICM cell and promotes bias towards PrE or epiblast

lineages[43,44]. Additionally, this differential FGF signalling may also stem from differential exposure to FGF4 levels, which in turn depends on the FGF4 expression of surrounding cells[39]. Depletion of *Fgf4* or *Fgfr1/2* or inhibition of the MEK1/ERK pathway, a downstream signalling cascade of FGF receptors, blocks PrE differentiation, leading to an ICM entirely composed of epiblast cells[43–47]. Reversely, exposure to exogenous FGF4 biases ICM cells towards the PrE lineage[48].

Two transcription factors are the first to display the second lineage specification in ICM cells in response to FGF signalling: NANOG and GATA6[36,49,50]. They gradually become mutually exclusive starting at E3.25 (32-cell stage), and, by E3.75 (64-cell stage) NANOG is expressed in the epiblast-biased cells and GATA6 in PrE-biased cells[45,50]. Importantly, this specification is asynchronous among ICM cells[41,50,51]. The driving function of these two master regulators is evidenced by the phenotypes of their deletion mutants. While in *Nanog* mutants, all ICM cells express GATA6 and acquire a PrE fate[52], in *Gata6* mutants, all ICM cells express NANOG and acquire a epiblast fate[41,53]. PrE-biased cells will progressively trigger PrE transcription factor expression including SOX17, GATA4 and SOX7 while epiblast-biased cells will express SOX2[54]. Additionally, these two populations will first appear in a seemingly mixed "salt and pepper" fashion within the ICM but PrE cells will gradually migrate, gain apico-basal polarity and epithelialize to form a layer adjacent to the blastocoelic cavity by E4.5[36,51]. By then, epiblast and PrE cell fates become resistant to FGF4-mediated fate remodelling, indicating that the window of lineage commitment has already closed[48,55]. One interesting fact however, is that epiblast cells appear be the first to specify their lineage and also to lose their lineage plasticity compared to PrE cells[35,55]. Therefore, symmetry in ICM is not only broken through transcriptional regulators that drive lineage divergence but also through the asymmetrical restriction of plasticity itself.

Plasticity and the regulatory nature of preimplantation embryo development

The question of how transcriptional asymmetries are first generated in the embryo and how they are amplified to direct lineage divergence has gained a lot of attention over the last decades. Despite causing an intense debate in the field, it remains an open question. We do however know that transcriptomic differences between blastomeres appear as early as the first cleavage division and are mainly due to random segregation during cell partitioning[56–58].

These initial blastomere-to-blastomere biases may be further amplified by transcriptional noise at zygotic genome activation at the two-cell stage[56]. Noise can be generated through a burst of transcription of a given gene but also through the propagation of fluctuations upstream of that gene, reviewed by[59]. The random stochastic fluctuations of protein activity and gene expression that these generate, enable diversity among genetically identical cells while maintaining the flexibility to respond to external cues[59,60]. Accordingly, cells can explore a larger landscape of differentiation possibilities while being poised to change[61]. By introducing biological noise into the embryo early on, it allows for it to be fine-tuned by positive and negative feedback loops to tilt the balance of lineage specifiers in each blastomere. Hence, cells in the embryo would gradually acquire a specific fate but keep a narrowing window of plasticity granted by certain noise levels that allow for redirection towards another fate.

One very debated example of symmetry breaking during preimplantation development is precisely the transition in ICM cells from NANOG and GATA6 coexpression to the mutually exclusive "salt and pepper" disposition in the early blastocyst. What breaks the symmetry? What originates the divergence of epiblast and PrE lineages starting from a seemingly homogenous population of cells in the first place? There are two opposing theories that try to answer these questions: a probabilistic model and a deterministic model.

In the first, some suggest that stochastic transcriptomic fluctuations followed by signal reinforcement are enough to establish the lineage decision[36,50,62]. In fact, *Nanog* transcription is pulsatile and occurs in bursts in embryonic stem cells which hints at the source for its heterogeneous expression at the blastocyst stage[63,64]. However, this pulsatile transcription has not been seen in embryos so far. Additionally, other studies suggest that it is possible to predict cell fate already by the 4-cell stage just by looking at the dynamics of SOX2-DNA binding[65] or at the levels of SOX21[15] in each blastomere. This implies that there is a previous lineage determination event much earlier than expected, and thus constitutes hidden variables in the probabilistic model. Importantly, many seemingly probabilistic mappings turn out to be more deterministic as hidden variables get exposed[66].

Others have a more deterministic model to explain symmetry breaking whereby the timing and origin of the ICM cell influences its future fate[67,68]. This model proposes that ICM cells that internalize during the first wave of asymmetric

cell division (8-to-16 cell stage) are epiblast-biased and cells that internalize in the second and third wave of asymmetric cell division (16-to-32 and 32-to-64 cell stage) are PrE-biased[67,68]. An independent study however, had contrasting results[48]. After some apparent controversy, the discrepancies among the studies were attributed to the different mouse strains and experimental set-up used[69,70]. Indeed, embryos from different strains produce differing numbers of ICM cells during the first division[71]. In the light of following studies, these differences became obvious and relevant[30,35,52,72]. Taking this into account, a model now posits that if fewer cells are internalized in the first wave, they won't secrete enough FGF4 to inhibit *Nanog* transcription, so they would acquire an epiblast fate[72]. On the other hand, if more cells are internalized in the first wave, the FGF4 they secrete would be enough to induce PrE fate in some of the newly formed ICM cells[72]. Importantly, it is widely accepted that no matter the origin, PrE and epiblast cells are specified independently from their position within the ICM and generate a "salt-and-pepper" pattern that will later reorganize.

Another thought-provoking concept revolving around early embryonic fate specification and plasticity is that many of the factors driving lineage specification are in fact, members of the same transcription factor families. One example of such family is the Gata family. *Gata3* is involved in TE specification early on[73]. Then, *Gata6* and later, *Gata4* are sequentially activated during PrE specification[53]. The later have a key position in PrE fate determination, as expression of either of them is sufficient to drive differentiation towards PrE[74]. It might be that while blastomeres express *Gata6*, they are biased towards PrE specification but still remain plastic towards other lineages. Later, when *Gata4* is expressed, blastomeres would have already acquired the PrE fate.

Another compelling example of evolutionary relationships between master lineage regulators lies in the Sox family of transcription factors. While SOX2 is the first marker of ICM cell specification, afterwards, *Sox17* and *Sox7* are activated sequentially in PrE-biased cells[54]. Additionally, SOX2 is part of the core pluripotency network in epiblast cells (together with OCT4 and NANOG). The close similarity across these family members might point at a scenario where OCT4 is easily allowed to switch its Sox-binding partner to adjust rapidly to fate-determining cues[75]. Accordingly, an OCT4- binding partner switch from SOX2 to SOX17 would disrupt the pluripotency network and initiate specification towards the PrE fate.

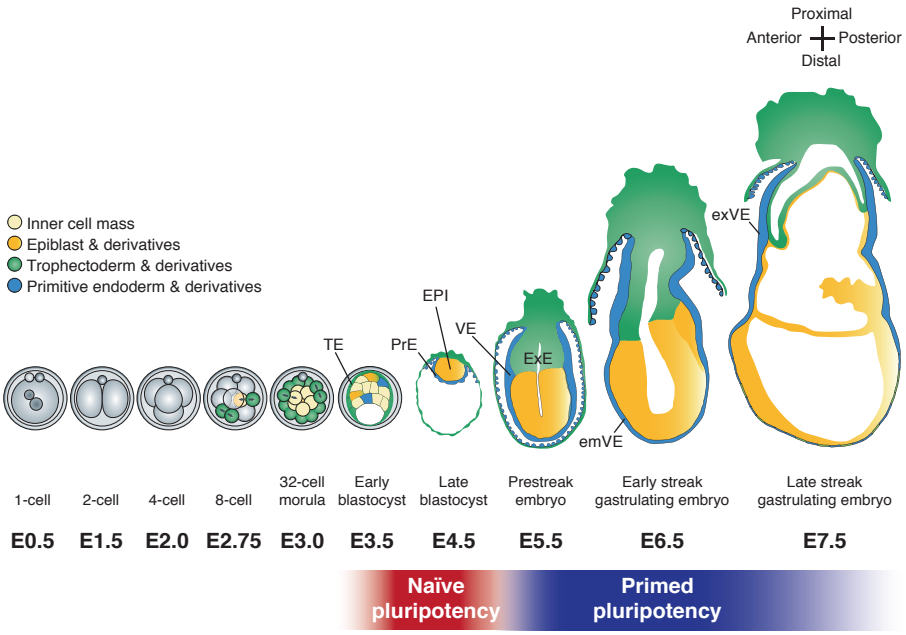
All in all, the fundamental outcome of this tight developmental regulation is that by E4.5 the blastocyst is composed of three clearly distinguishable cell populations with the trophectoderm constituting an epithelium enclosing a fluid-filled cavity (blastocoel cavity); the epiblast constituting an aggregation of cells on one side of that cavity; and the PrE constituting the epithelium separating the epiblast from the cavity. Some view the whole process of formation of this blastocyst as a "preembryonic stage", a deterministic preparatory system necessary to generate the extraembryonic tissues[76,77]. In this line, the epiblast can be viewed as a "ground state" or *tabula rasa*, a state that harbours the developmental potency and plasticity needed to make all embryonic lineages.

The continuum of pluripotency and the starting point of embryonic pattern formation

The pluripotency state exists in the epiblast of the mouse embryo from the early blastocyst stage at E3.5 and is lost progressively during gastrulation at E8.0, reviewed in[78]. Over this period, the pluripotent cell population likely progresses through a broad continuum of pluripotency states characterized mainly by differences at the level of gene expression, epigenetic signatures, metabolic status and functional properties. To this date, however, it has only been possible to stably capture *in vitro* two of these pluripotency states: the naïve and primed state, which in turn correspond to the states in the pluripotent epiblast of pre- and post- implantation embryos respectively (Introduction figure 2).

Additionally, we are now capable of deriving and stably maintaining *in vitro* stem cell populations of the three developmental lineages at the peri-implantation stage. On one side, the aforementioned epiblast lineage can be recapitulated by either embryonic stem cells (ESCs)[79–81] or by epiblast stem cells (EpiSCs)[82,83], depending on whether this lineage is in the naïve or primed state of pluripotency, respectively. On the other side, trophoblast stem cells (TSCs) from the trophoblast lineage[84] and extraembryonic endoderm (XEN) cells from the PrE lineage[85] can also be maintained in culture. Importantly, all these cell lines contribute to their corresponding lineage in chimaeras generated by their injection into developing embryos at their corresponding stages.

The crucial establishment of these culture techniques revolutionized the field of developmental biology and led to a growing number of studies that have resulted in the discovery of many developmental gene functions. Importantly, we have been able to identify key building blocks of stem cell identities, and



Introduction Figure 2: Early mouse embryonic development and pluripotency states. The zygote goes through three rounds of cleavage events, to become an 8-cell embryo which after compaction and polarization, will produce two populations of cells in the first lineage specification. Outside cells become trophectoderm while inside cells become the inner cell mass. The inner cell mass then undergoes the second lineage specification which separates two populations: the primitive endoderm (PrE) and the epiblast (EPI). After implantation, at around E4.75-E5.0, the epiblast turns into a columnar epithelium, the trophectoderm forms the extraembryonic ectoderm (ExE) and the primitive endoderm forms the parietal and the visceral endoderm (VE). A continuous layer of visceral endoderm covers the epiblast (EmVE) and the extraembryonic ectoderm (ExVE). The two types of pluripotency states are shown indicating the time that they are present in the embryo. Naïve pluripotency appears at the blastocyst stage and persists until implantation whereas primed pluripotency arises after implantation and is gradually lost during gastrulation. The days of embryonic development are depicted (E, embryonic day). Embryo artwork is based on schematics from Zernicka-Goetz et al. 2009[6] and Nowotschin et al. 2019[580].

continue to do so as new technologies advance. Nevertheless, we should not overlook that during embryogenesis, pluripotency is a highly dynamic and transient process, but when it is captured in vitro, it is maintained indefinitely in a self-renewal state. Self-renewal is not a defining feature of pluripotency but rather a by-product that came from the derivation of these cells, and this together with additional adaptations to the *ex vivo* environment might result in *de novo* regulative properties. Consequently, one must be careful when interpreting experiments using *in vitro* stem cells and extrapolating them to their *in vivo* cellular states as these cell lines have been artificially stabilized and lack a number of extracellular, morphological and positional cues that are present *in vivo*.

In the next section, I will thoroughly discuss the discoveries made on the two pluripotency states captured *in vitro* and describe the molecular mechanisms that shape the naïve and primed cellular identities. In this section however, I will focus on describing what happens *in vivo*. Particularly, I will describe the cross talk between embryonic and extraembryonic tissues in the embryo that promotes morphogenetic reorganizations and cell state switches in the epiblast and that sets the stage for the establishment of the body plan through embryonic pattern formation.

Implantation & axis formation in the embryo

Embryonic pattern formation is the cornerstone of embryogenesis. It arises from the establishment of the three body axes (anterior-posterior, dorsal-ventral and left-right) and the three germ layers (ectoderm, mesoderm and endoderm). These key establishments start in the epiblast embryonic stem cells upon implantation and continue through the first days of post-implantation. It is a complex process that is spatiotemporally controlled and involves numerous regulatory levels and coordinated interactions between multiple tissues. Coordinated tissue movement and growth is decisive. Ultimately, epiblast cells go through symmetry breaking and start lineage specification concomitantly with a general morphological reorganization of the whole embryo.

During implantation at E4.75-E5.0, trophoblast cells tightly anchor the embryo to the endometrial stroma. For this, the blastocyst must attach to the uterine luminal epithelium and trigger local apoptosis so that trophoblast cells can penetrate through the uterine apoptotic epithelium and reach the underlying stroma, reviewed in[86]. Importantly, for successful implantation to occur the

uterus must be receptive and the embryo must have gained full implantation competency. This depends on many signalling players that serve as paracrine, autocrine and juxtacrine factors coming from both the mother and the embryo. These include cytokines, homeobox transcription factors, adhesion molecules, growth factors, receptors, lipid mediators and morphogens, reviewed in[87]. All coordinate the essential crosstalk between blastocyst and uterus for a successful implantation.

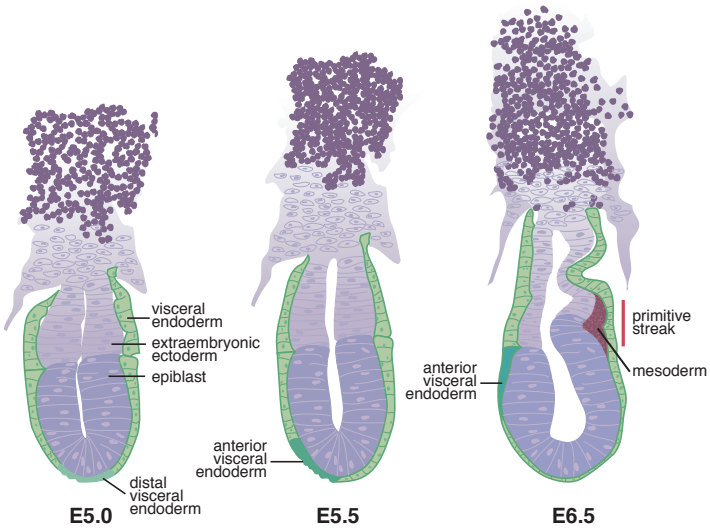
Subsequently, implantation triggers a series of morphogenetic reorganizations through which the first axial asymmetry is established. First, implantation stimulates polar trophectoderm cells to form the ectoplacental cone. This is the anchor point to the endometrial stroma. The endometrial cells then react by proliferating, decidualizing and surrounding the whole conceptus[88]. Importantly, ectoplacental cone cells will later ensure adequate maternal blood supply and, hence, nutrients and oxygen by penetrating deeply into the maternal decidua and contacting with the spiral arteries, reviewed in[89]. Polar trophectoderm cells also form the extraembryonic ectoderm at the distal part of the ectoplacental cone[90]. Specially at these early stages, extraembryonic tissues formed by the trophectoderm and the PrE, will not only have a key role in nutrition and homeostasis of the embryo but will also be critical players in its patterning and lineage segregation; as I will summarize below.

What was once a ball of epiblast cells in the blastocyst will undergo cell growth and consequently, the embryo will change both in size and shape, elongating from the anchor point (the proximal region, where the ectoplacental cone is). Cells in the epiblast will continue dividing and expanding, forcing the epiblast to cavitate and form a cup-shaped epithelium. The contact of these epiblast cells with the basement membrane will trigger their apico-basal polarization. The luminal space that is generated in the interior is the proamniotic cavity and is key for polarization and naïve pluripotency exit[91,92]. Therefore, at this point, the embryo has a clear elongated structure with a proximal and a distal region and has expanded distally into the blastocoel cavity[93]. This structure is termed egg cylinder.

Concomitantly, cells of the PrE which at the blastocyst stage were contacting only the epiblast cells, now grow and divide to cover both the epiblast cells and the extraembryonic ectoderm in a continuous single-cell layered epithelium (reviewed in[94,95]). This epithelium is termed visceral endoderm[96]. It will ultimately develop into the visceral yolk sac and will also be part of

the definitive endoderm that contributes to the internal lining of the gut[97–99]. Depending on the developmental stage and the place within the tissue, the cells constituting the visceral endoderm have a varying morphology from columnar to cuboidal or squamous.

At E5.5, a gradient of NODAL signalling will be set up through the reciprocal interactions of embryonic and extraembryonic tissues. The extraembryonic ectoderm adjacent to epiblast cells secretes FURIN and PACE4, two proprotein-convertases required for the maturation of NODAL. NODAL, in turn, is secreted by the epiblast cells[100]. Additionally, NODAL triggers BMP expression



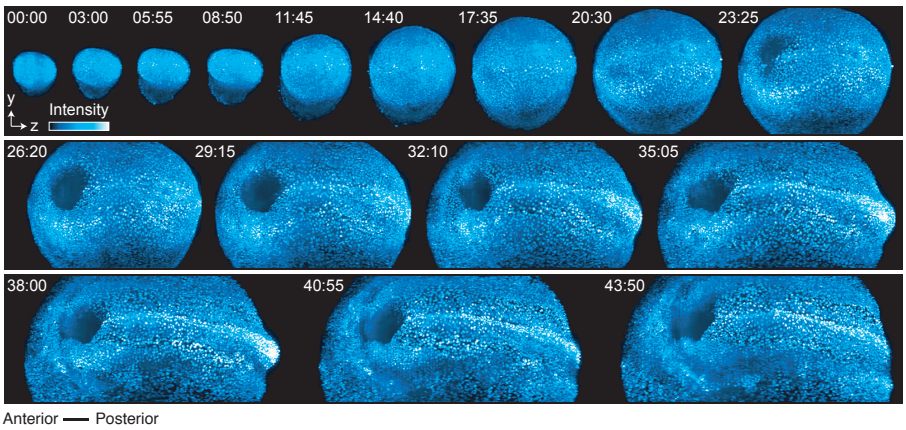
Introduction Figure 3: Egg cylinder stages upon distal visceral endoderm induction and migration showing the initiation of anterior-posterior axis formation. Mouse embryos at the E5.0 stage have a single continuous layer of visceral endoderm which covers the epiblast and the extraembryonic ectoderm. Distal visceral endoderm cells are induced at the most distal edge of the egg cylinder and switch from having a squamous epithelial morphology to having a columnar one at embryonic stage E5.5. Then, these cells migrate unidirectionally to one side of the egg cylinder, forming the anterior visceral endoderm, and by doing so they determine the anterior of the embryo. The primitive streak will form on the epiblast cells opposite to and furthest from the anterior visceral endoderm as a consequence of repressive signals from the anterior visceral endoderm. At the most proximal location, the anterior visceral endoderm cells retain their columnar morphology. Modified from Arkell and Tam 2012[581].

in extraembryonic ectoderm, which in turn, amplifies NODAL signalling by activating WNT3 in the epiblast cells[100]. These two independent feedback loops originating at the epiblast-extraembryonic ectoderm interphase will signal the visceral endoderm covering the area and will trigger the subsequent regional pattern acquisition.

At the same time, the visceral endodermal cells at the most distal tip of the egg cylinder change from having a cuboidal morphology to having a columnar morphology[101–103]. This small change has crucial consequences as it is the first step towards transforming the proximal-distal asymmetry that the embryo has at this point, into the anterior-posterior asymmetry that is fundamental for later development[104] (Introduction figure 3). This group of cells, termed the distal visceral endoderm, will move across from the distal position where they were to the interphase dividing epiblast and extraembryonic ectoderm, but restricted to one side of the egg cylinder[103]. NODAL signalling provides the driving source of this movement by stimulating the proliferation of visceral endodermal cells whereas its inhibitors LEFTY1 and CER1 determine the direction of migration by asymmetrically inhibiting its activity[105]. This will signal the underlying epiblast and will mark the anterior position of the embryo, reviewed by[95]. The anterior visceral endoderm will secrete NODAL and WNT antagonists, which will produce an anterior-posterior gradient of NODAL and WNT activity signalling across the epiblast [105]. At this time, these cells become known as the anterior visceral endoderm. The anterior visceral endoderm becomes a major signalling centre that by E5.75 is clearly distinguishable on one side of the egg cylinder[104]. The underlying epiblast however, continues to grow steadily and remains morphologically uniform until E6.25[106].

At E6.25, a thickening of the epiblast starts to be visible on the most proximal-posterior side of the embryo, opposite to where the anterior visceral endoderm is (Introduction figure 3). In this structure, termed the primitive streak, cells undergo epithelial-to-mesenchymal transition (EMT) by dissociating from the epithelial epiblast, ingressing and migrating to form cell layers constituting the mesoderm and endodermal lineages of the embryo, reviewed in[107]. This process is called gastrulation and as the primitive streak progressively advances towards the distal tip of the embryo, different cell identities will be positioned along the way, generating the basic body plan (Introduction figure 4). Therefore, by E6.5, the embryo is clearly asymmetrical with the anterior side still composed of epithelial epiblast cells but with the posterior already composed of progenitor cells of the future definitive endoderm and embryonic

and extraembryonic mesoderm. Later, the anterior epiblast cells will mainly form the future neuroectoderm. Concomitantly, BMP signals emerging from the extraembryonic ectoderm will induce around four to eight primordial germ cells (PGCs) in the posterior pre-streak epiblast cells[108,109]. These are the embryonic precursors of the gametes and will be the founder cells of the future germline.



Introduction Figure 4: Light sheet microscopy capturing mouse gastrulation at high spatiotemporal resolution. Ventral view projections of an H2B-eGFP expressing mouse embryo over a 44-hour period, shows the localization of single cells starting from early streak E6.5 stage until the E8.5 somite stage. Time (hh.mm). Image from McDole et al. 2018[582].

Regionalization of pluripotency in the post-implantation embryo.

The transcriptional regulatory network of naïve state pluripotency collapses after implantation[110]. At this stage, the epiblast is unpatterned[110,111] and expresses core pluripotency factors such as SOX2, OCT4 and NANOG[112–115]. NODAL[116] and BMP[117] signalling safeguard this pluripotent identity but as development advances, the extraembryonic ectoderm and visceral endoderm will trigger the regional patterning of these cells and therefore the epiblast will irrevocably become an heterogeneous population of cells with different cell identities. Importantly, during post-implantation development, epiblast cells must transit from the naïve to the primed pluripotency state. It is thought that

they do this in an orderly sequence of events passing through a continuum of intermediate pluripotency phases. These must be spatially regulated within the embryo. However, we are only starting to understand how, when and where do epiblast cells transit through these states after implantation. Additionally, in the recent years, research in the field is starting focus on characterizing the intermediate steps in this continuum as it has been proposed that an intermediate phase, termed the formative state, is required for epiblast cells to acquire the competence they need for multi-lineage induction, previous to acquiring the primed state[118].

Characterizing this phase and other phases in the pluripotency continuum of the embryo is key to understanding multi-lineage decision-making processes and pluripotency states. A key step towards it was accomplished recently in an inspiring study that generated the spatiotemporal transcriptome of epiblast cell populations from the peri-implantation to the gastrulating mouse embryo, thus exactly covering the time window when pluripotency arises and is lost in the embryo[110]. Previous studies such as[111,119] also provided high resolution transcriptomes of the pluripotency continuum of the mouse embryo but these lacked the spatial information that is so crucial when interpreting the role of pluripotency loss in body plan formation. Peng et al. describe that after the naïve pluripotency network collapses upon implantation, the continuum of pluripotency advances and by E5.5, the entire epiblast expresses what they defined as the formative pluripotency transcriptional network[110]. The expression of this regulatory network will regionalize to the anterior epiblast by E6.5 and then be retained there by E7.5. On the other hand, they posit that the primed pluripotency regulatory network appears in the posterior epiblast by E6.5, concomitantly with EMT markers. The expression of this primed pluripotency state markers is still present at E7.0 but starts diminishing by E7.5.

This finding clashes with what was previously described by Kojima et al.[120], in which by carefully analysing the transcriptional and functional properties of EpiSCs and post-implantation epiblasts, they proposed that primed pluripotency largely occurs in the anterior epiblast cells of E7.5 embryos. At this stage, the anterior ectoderm is poised for neuroectoderm differentiation which is in line with primed pluripotency propensities[78], as we will discuss in the upcoming section. Additionally to this, the fact that Kojima et al. detected primed pluripotency at the anterior epiblast instead of at the posterior epiblast, goes in line with the inefficiency that EpiSCs show when being induced to

form PGCs [121]. Most probably, the two studies clash because Peng et al. used an unconventional set of markers to define primed pluripotency. In fact, most of the widely accepted primed markers were actually included in their formative pluripotency marker list.

All in all, one could speculate that primed pluripotency could emerge at the post-implantation epiblast, then be restricted to the anterior epiblast by the late gastrula stage to then be extinguished upon differentiation.

Capturing Naïve and Primed pluripotent stem cell identities

The awakening of pluripotency research

Pluripotency research started with the finding of a mouse strain that spontaneously developed testicular teratocarcinomas[122]. These tumours were composed of both somatic tissues and undifferentiated malignant proliferative cells. These cells, termed embryonal carcinoma cells, could be maintained *in vitro* with their full pluripotential capabilities but were karyotypically abnormal[123]. Fortunately, the culture conditions that were optimized for growing embryonal carcinoma cells *in vitro* also allowed the derivation of pluripotent stem cells directly from mouse blastocysts[80,81]. These stem cells were karyotypically normal and had the remarkable ability to contribute to chimaeric mice without forming tumours[124], this included germline colonization, which announced the era of targeted mouse genome editing[125].

Originally, pluripotency was viewed as a unique cell identity yet it was clear quite early on that pre- and post- implantation epiblast cells were very different in morphology and function. While pluripotent cells taken from pre-implantation embryos could incorporate back into host blastocysts and contribute to chimaeras, cells taken from post-implantation embryos could not[126]. Looking back, this finding was the first reflection that more than one class of pluripotency was attainable.

In mouse, the two stable pluripotent cell types derived from the peri-implantation embryo were named according to the tissue they were derived from in the embryo. Thus, while ESCs were isolated from the ICM of preimplantation embryos[80,81], EpiSCs were isolated from epiblast of post-implantation embryos[82,83]. Importantly, other pluripotent stem cell types have been isolated, including: extended pluripotent stem cells, derived from blastocysts but with an additional extraembryonic developmental potential[127]; expanded-potential stem cells, derived from mouse cleavage-stage embryos[128]; embryonic germ cells, derived from early migrating PGCs[129]; and germ stem cells, derived from neonatal[130] or adult spermatogonial[131] stem cells. Additionally, an alternative road for pluripotent stem cell isolation is somatic cell reprogramming by ectopic expression of a set of defined transcription

factors that yield induced pluripotent stem cells[132], or by nuclear transfer, yielding nuclear transfer stem cells[133].

Irrespective of the stem cell origin, the growth conditions in which stem cells are expanded will define their pluripotent state. For example, if stem cells are derived from the ICM they can either be naïve or primed depending on whether they are grown in ESC or EpiSC conditions, respectively[80,81,134]. This puts forward the interest in unravelling the molecular determinants that are involved in the definition of these states to capture them properly *in vitro*.

A good example to illustrate the widespread of this rationale are the extensive attempts undertaken to capture the naïve state in human pluripotent stem cells, and the large number of culture conditions they have led to, reviewed in[135]. Human and rhesus macaque stem cells resemble EpiSCs[136–139] and for some time it was believed that the naïve pluripotency state could simply not be stably isolated from certain species either because it was not a stable state in those species or because it was not a universal pluripotency state to begin with. This included human stem cells[140] and non-permissive mouse strains[141], among others. However, a seminal study from Hanna et al found that naïve conditions of a non-permissive mouse strain could, in fact, be stabilized by replacing certain factors of the *in vitro* reprogramming cocktail[141]. This hinted at the idea that stem cells from diverse genetic backgrounds could stably acquire other pluripotency states if the endogenous genetic determinants were regulated with the correct exogenous factors. Hence, the quest for human naïve pluripotency started and brought about many studies that found alternative ways to regulate the endogenous genetic determinants towards that state. These include overexpression of transcription factors as well as transgene-free inducing or inhibiting chemical protocols, that led to the derivation of human naïve stem cells from either established primed human stem cells, human induced pluripotent stem cells, *de novo* reprogrammed human induced pluripotent stem cells or directly from the human blastocyst[142,143,152–154,144–151].

Naïve and primed stem cells differ in morphology, functional potential, culture conditions and transcriptional, epigenetic and metabolic composition. In the upcoming pages, I will compare the features of these two known stable attractor states of pluripotency.

Naïve embryonic stem cells

Derivation & maintenance of ESCs

ESCs retain a certain degree of developmental flexibility and can diverge or reverse from the usual developmental path given the proper permissive conditions to do so. Although ESCs are functionally, epigenetically and transcriptionally similar to the epiblasts of E3.5-E4.5 embryos from which they are regularly derived[79–81,155], they can also be derived from embryos as early as E0.5[155–157] and as late as E7.5[141,158]. Importantly, early embryos will develop *in vitro* as far as a late blastocyst stage before ESCs can appear whereas gastrulating embryos must be dissociated for ESCs to be derived, hinting at the idea that a refractory niche is present at this stage. In fact, most attempts to derive ESCs from whole explants of post-implantation embryos largely failed, consistent with the notion that pre- and post- implantation epiblast cells have differing identities. The ESCs that are derived from embryos at that stage are termed reprogrammed epiblast ESC-like-cells (rESCs) as they are reprogrammed to an earlier state through a series of transcriptional and epigenetic reversals during a derivation procedure where the disruption of existing cell-cell interactions is key[158]. An analogous reprogramming can also be done to revert EpiSCs to ESCs[141,158–162]. This reversal is highly inefficient but can be enhanced by genetic or chemical manipulation[162–171].

ESC cultures were first obtained by coculture of ICM cells with a layer of mitotically inactivated fibroblasts, the so-called feeder layer, grown in a calf serum containing medium[80,81]. Co-culture with feeders was a standard technique at the time to enhance cell culture viability and proliferation, yet, these conditions where far from the physiological setting of pluripotent cells in preimplantation embryos and deconvolution of serum components and feeder layer interactions soon became a priority in the field. We now know that the key factors provided by serum and the feeder layer are activators of bone morphogenetic proteins (BMPs)[172] and the leukaemia inhibitory factor (LIF)[173–175] pathways, respectively. Together, BMP and LIF stimulate pluripotency and block differentiation[172,176,177]. More specifically, BMP inhibits neural differentiation[117] through the inhibition of mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) signals[178,179] and through the induction of Id genes[172] while LIF maintains self-renewal through the activation of signal transducer and activator of transcription 3 (STAT3)[180–184]. All in all, despite the fact that BMP can replace serum for ESC maintenance,

the combination of serum and LIF (SL) is preferred for ESC culture due to its enhanced plating efficiency[172] and lower price.

Molecular signatures of ESCs

ESCs keep an open chromatin structure[185], analogous to preimplantation embryos[186], with global transcriptional hyperactivity[187] and expression heterogeneity[188]. This open and accessible chromatin state of ESCs is characterized by having extensive regions of DNA hypomethylation[189], histone acetylation and H3K4me3 marks[185,190] and is partly attributed to citrullination of histone H1[191,192]. Additionally, ESCs derived from female embryos maintain both X chromosomes active, in the same way as their *in vivo* counterparts[165,193,194]. Curiously, upon exit of naïve pluripotency, X chromosome inactivation is gender independent and is initiated in both male and female ESCs [195–197]. In males, this X chromosome inactivation-like state is rapid and transient, while in females it results in the inactivation of one X chromosome[196]. Then, as differentiation commences, the ESC genome undergoes large-scale silencing[187,198].

ESCs maintain the expression of a naïve transcription factor network that is also present in preimplantation embryos and that includes OCT4 (regulated by a distal enhancer element[199]), SOX2, NANOG, KLF4, STELLA and REX1, among others. These transcription factors are essential for sustaining naïve pluripotency gene expression, reviewed in[200]. These transcription factors also regulate the expression of key microRNAs (miRNAs) which, in turn, play crucial roles in cell cycle progression and self-renewal[201–205]. Furthermore, the ectopic expression of a set of these transcription factors or these miRNAs is sufficient to promote reprogramming of somatic cells to naïve pluripotency[132,206–208].

Importantly, the naïve pluripotency transcriptional program is highly dynamic in ESCs. This was widely accepted after two discoveries were made: first, showing fluctuations over time in the expression levels of *Nanog* and other pluripotency regulators[209–213]; and secondly, showing that different subpopulations of ESCs have different capacity to differentiate and self-renew[214]. It has since been postulated that ESCs fluctuate in multiple interconvertible states and this metastability and plasticity might be fundamental for their pluripotent identity[215–221]. In line with this, single-cell RNA-seq analysis has revealed that ESCs are highly heterogeneous in their gene expression when grown in standard SL medium conditions[220]. Particularly, two classes of transcriptional

heterogeneity have been described according to the gene expression distribution amongst ESCs: bimodal expression, with transcripts present in some cells and not in others; and sporadic expression, with transcripts in a reduced number of cells but at high levels[220]. This transcriptional heterogeneity in ESCs has been proposed to arise from stochastic fluctuations in gene expression, cell cycle phase asynchrony, transcriptional bursting, differential epigenetic regulation, partial differentiation and differences in colony size and morphology[188,214,220,222]. As a result of this heterogeneity, ESCs in culture do not simply exhibit transcriptional similarities with the preimplantation epiblast but also contain subpopulations of cells sharing transcriptional similarities with the post-implantation epiblast[214,223], the primitive endoderm[224] and the 2-cell embryo[225,226]. While in ESC culture these subpopulations interconvert, there is no evidence showing that this also happens in the peri-implantation embryo. What we do know, however, is that preimplantation E3.5 embryos and the uncommitted epiblast of E6.5 embryos also show an elevated transcriptional heterogeneity[111,119]. This transcriptional heterogeneity has been proposed to aid in symmetry breaking and cell-fate decision making as it may create competence for subpopulations of cells to respond to signalling cues[111].

In vivo versus *in vitro*: locking the naïve pluripotency state

ESCs are sustained at a naïve pluripotency state. However, this is a transitory state in the epiblast cells of the embryo. Consequently, ESCs must make use of distinct mechanisms, not found in the early embryo, to sustain infinite self-renewal and to retain naïve pluripotency. One example is the expression of genes such as *Esrrb* or *Tbx3*, essential for ESC pluripotency but not for pluripotency in the early embryo[227,228]. Another example is the distribution of bivalent chromatin domains, which differs in ESCs and preimplantation epiblast cells[229]. Bivalent chromatin are histone-bound DNA regions that have both permissive and repressing histone marks. In the case of ESCs and preimplantation epiblast cells, there is co-enrichment of H3K4me3, a permissive histone modification and of H3K27me3, a repressing histone modification, in promoters linked to developmental signalling functions[229]. Despite differing in localization between *in vivo* and *in vitro* pluripotency, bivalent chromatin has been interpreted in both systems as a mechanism by which pluripotent cells silence their permissive chromatin enriched genome while ensuring a mechanism for rapid activation or repression of developmental genes upon differentiation[230,231].

Another reported difference between *in vivo* and *in vitro* naïve pluripotency is a higher expression of repressive epigenetic factors in ESCs compared to preimplantation embryos[232]. These include DNA methyltransferases, histone deacetylases, H3K9 methyltransferases and H4K20 methyltransferases[232]. One hypothesis is that they function to firmly block the differentiation program. Upon ESC derivation from the ICM, this increase in repressive epigenetic factors is accompanied by a decrease in activating epigenetic factors including histone acetyltransferases, H3K9 demethylases, H3K27 demethylases and H3K4 methyltransferases[232]. In line with this, thousands of genes are differentially expressed. These differentially expressed genes are strongly enriched in transcriptional related processes such as transcription factor activity and DNA-dependent regulation of transcription[232]. Much more research is needed to truly understand the effects of these epigenetic and transcriptional differences, but overall, they suggest that preimplantation embryos may require greater epigenetic flexibility due to their transient identity.

The ground state model of pluripotency

ESCs exist in a constant equilibrium between pluripotency and lineage commitment driven by the endogenous production of two factors with antagonistic roles: on one hand, LIF supports self-renewal[233] and on the other, FGF4 promotes differentiation[234]. Inhibiting one of the factors drives cells towards the opposing direction. Given that ESCs are routinely grown with exogenous LIF, this equilibrium is largely tilted towards self-renewal. In line with this, inhibiting FGF4 signalling confines ESCs in a cell state with reduced lineage commitment, termed "the ground state"[235]. In fact, the confinement of cells in this ground state might be one of the main reasons why the efficiency of ESC derivation from ICM cells is notably enhanced with genetic or chemical inhibition of FGF4[236].

The ground state model of pluripotency coins that ESCs have an innate programme for self-renewal that is reached when cells are insulated from their endogenous differentiation stimuli[235]. To insulate cells from these stimuli they are grown in serum-free (N2B27) medium containing two small molecule inhibitors (2i). These two inhibitors are: PD0325901, a MEK inhibitor acting downstream of FGF signalling; and CHIR99021, a GSK3 inhibitor that activates WNT signalling and frees pluripotency-associated genes from TCF3-mediated repression[237]. Although LIF is not necessary for ESC growth in 2i medium, it does enhance clonogenicity[237] and thus 2i plus LIF (2iL) medium is regularly

used. Importantly, the use of 2iL provides highly efficient ESC derivation which validates the ground state model of pluripotency. Additionally, it has granted the derivation of ESCs from the ICM of refractory and non-permissive mouse strains and from rat embryos[235,238].

2iL ESCs show minimal lineage commitment marker expression and minimal spontaneous differentiation but a similar differentiation potential compared to SL ESCs[239]. Furthermore, 2iL ESCs have a high degree of homogeneity in terms of pluripotency transcription factor expression which is not the case for SL ESCs[222,240]. Instead, 2iL ESCs show heterogeneity in cell cycle gene expression and contain small subpopulations sharing some transcriptional similarities with the 2-cell stage (2C-like cells) and extraembryonic endoderm[222,241].

2iL ESCs have extensive global DNA demethylation, even at imprinting control regions. Consequently, several groups reported that prolonged culture of 2iL ESC leads to genetic instability and impairment of their developmental potential[242,243]. Since then, other ESC culture conditions with modifications to the 2iL cocktail supporting higher genomic integrity in the ground state have been reported. These include: alternative 2i (a2i) where PD0325901 is replaced by an SRC kinase inhibitor[244]; titrated 2i (t2i) where PD0325901 dosage is reduced[242,243]; or R2i where both TGFb and FGF pathways are inhibited[245]. Furthermore, another modification of the 2iL chemical cocktail, named LCDM, substituted PD0325901 for (S)-(+)-dimethindene maleate and minocycline hydrochloride and enabled the derivation of a novel type of stem cells named extended pluripotent stem (EPS) cells[127]. EPS cells not only show genomic stability but also have both embryonic and extraembryonic potential at the single-cell level[127,246]. Finally, a similar type of stem cells can be derived by using the 2iL cocktail with addition of SB203580, a MAPK inhibitor; JNK inhibitor VIII; XAV939, an axin stabilizer; and A-419259, an inhibitor of SRC family kinases[128]. Derivation of stem cells with this cocktail results in the generation of expanded potential stem cells (EPSCs), which share features with EPS cells (both can contribute to embryonic and extra-embryonic lineages), but EPSCs can be readily differentiated to trophoblast stem cells by the standard *in vitro* protocols[128] while EPS cells need to be converted into self-organized three-dimensional blastocyst-like structures (termed blastoids) for TSCs to be derived[246]. Remarkably, these blastoids are able to undergo implantation, induce decidualization and generate live, yet disorganized, tissues in utero[246]. All in all, the SL and 2iL continue to be the prevailing ESC growth conditions in the field, while other conditions are still emerging.

Identity & developmental potential of ESCs

Comparative transcriptional profiling has shown that the embryonic counterpart of ESCs can vary depending on the culture conditions that the ESCs are grown in. While SL ESCs are transcriptionally similar to the epiblast of E4.5-E5.5 embryos and extraembryonic lineages, 2iL ESCs are more similar to the epiblast of E4.5 embryos[155,222,232,241,247,248]. Although this could be regarded as different culture conditions capturing specific stages of the pluripotent embryo, care must be taken when interpreting them given that ESC cultures contain underlying subpopulations among them (as we have commented on earlier sections). In line with this, single-cell sequencing of SL ESCs showed that cells exist in three distinct states within the whole population[222]. The most numerous subpopulation exhibits high levels of pluripotency factors, another exhibits a differentiation permissive profile while the smallest subpopulation is already in the differentiation path[222]. Another study found two subpopulations instead: one with a transcriptional signature similar to E2.5-3.5 embryos and another, similar to the E5.5 epiblast[249]. They proposed that the transcriptional similarity to the E4.5 epiblast was to be interpreted as the combination of these two subpopulations[249].

Although transcriptional profiling sheds light on ESCs identity, what truly measure pluripotency are functional assays. The most established and fast way of assessing functional potential directed to a specific lineage is through low-density monolayer differentiation assays. In their simplest versions, SL ESCs differentiate into endodermal and mesodermal lineages upon LIF removal but require N2B27 medium without serum to differentiate into neural lineages[250]. Other *in vitro* differentiation assays have also been used to test ESC differentiation into PrE, but the fact that many of the markers used to determine PrE differentiation overlap with definitive endoderm is a great disadvantage for their interpretation[251,252]. Throughout the years, many other directed differentiation protocols have been reported. These mostly employ small molecule inhibitors and cytokines to force ESC differentiation towards a precise developmental path. Although they usually result in a close approximation of the desired cell type, on many occasions, they do not result in fully functional cells.

Another widely used protocol to investigate the differentiation potential of stem cells is through the formation of three-dimensional aggregates named embryoid bodies (EBs)[253]. This protocol mimics certain processes

of embryogenesis and can inform us of differentiation biases or defects in ESCs. More precisely, in this protocol, stem cells are allowed to aggregate and differentiate into various developmental lineages while self-organizing morphogen gradients are established[254]. In the recent years, EBs have been used to produce organ-like formations, named organoids[255,256], and embryo-like formations, named embryoids[257–260]. While these two are very valuable tools to study lineage specification and self-organization, they are still in their early days of development and more research has to be done for them to fully assess the functional potential of ESCs.

Doubtlessly, the gold standard to fully assess the functional potential of stem cells is to test the ability of cells to integrate into an embryo and resume normal embryonic development[78,261]. On that account, cells are injected in groups or as single cells into host embryos and resulting chimaeras are analysed[261]. ESCs integrate into preimplantation embryos at many different stages[155,247,261–263] but cannot integrate into post-implantation embryos[264]. This indicates that naïve pluripotency is not compatible with the post-implantation niche and demonstrates functional equivalence between ESCs and the preimplantation embryo. Specifically, ESCs contribute most efficiently to epiblast derivatives, meaning that they integrate and differentiate into all three embryonic germ layers and the germline. ESCs can also contribute to extraembryonic lineages but this usually occurs at low levels[225,241,247,261]. This extraembryonic contribution can be incremented by selecting particular subpopulations of cells within the culture[224,241] or by growing ESCs in the 2iL condition[241,247]. The latter is consistent with the idea that 2iL ESCs correlate with an earlier developmental stage compared to SL ESCs. One of the most significant observations in chimaeras is that live-born fertile mice can be generated entirely from ESCs when they are injected into developmentally compromised tetraploid embryos[265]. In these chimaeras, the placenta and the yolk sac endoderm are of tetraploid origin while the yolk sac mesoderm, amnion and foetus are almost entirely of ESC origin[265–267]. Finally, while the tetraploid complementation approach results in F0 generation mice with almost complete ESC origin, it is also inefficient, presents non-specific lethality and around 2% host contamination[265–267]. More novel techniques such as laser-assisted injection of ESCs into eight-cell stage embryos yield healthy F0 generation mice that are entirely ESC-derived and show total germline transmission[263].

Primed Epiblast stem cells

Derivation & maintenance of EpiSCs

EpiSCs can be derived from the epiblast of post-implantation embryos ranging from the E5.5 to the E8.0 stage[82,83,268]. Importantly, derivation efficiency declines as pluripotency declines[120]. EpiSCs can also be derived from ICMs of preimplantation embryos but these must first be grown in minimal medium conditions before being transferred to EpiSC medium given that ICM cells differentiate when cultured directly in EpiSC medium[82,134]. This intermediate step probably grants the window of opportunity that ICM cells need to developmentally progress and adapt their signalling networks towards the primed pluripotency state. Additionally, EpiSCs can also be derived directly from ESCs by growing them in EpiSC medium[163,269], however, this requires a long and tedious process involving a great deal of cell death and differentiation. This strongly contrasts with the *in vivo* situation where the naïve-to-primed transition occurs in less than 2 days. To date, no study has directly assessed the differences in functional potential or performed full transcriptome comparisons between ESC-derived EpiSCs and embryo-derived EpiSCs. Nonetheless, they have been reported to be both morphologically and transcriptionally alike.

EpiSCs are grown in a defined serum-free medium (N2B27) containing Activin A and FGF2[82,83]. Activin A is a member of the TGF-beta superfamily of proteins that has overlapping signalling features with NODAL, which, in the embryo, supports pluripotency and blocks premature neural differentiation[270,271]. Activin A is thought to be doing the same role as NODAL *in vivo* to maintain self-renewal *in vitro*[160,272]. Additionally, FGF2 blocks neural differentiation too and blocks rare events of reprogramming into the naïve state[82,83,160]. The role of FGF2 in the epiblast of post-implantation embryos remains unclear but some studies suggest it regulates cell proliferation[273].

In contrast to ESCs, which can be kept in a relatively steady state of self-renewal, EpiSCs show a high degree of spontaneous differentiation[274]. Additionally, ESCs can produce endogenous fibronectin when simply grown on gelatin[275], while EpiSCs cannot. In fact, EpiSCs require an exogenous supply of fibronectin which is key for their self-renewal. In line with this, dissociation of EpiSCs into single cells triggers extensive cell death and differentiation, thus, passages must be done as cell clusters by scraping cells or by mild enzymatic colony dissociation.

Molecular signatures of EpiSCs

EpiSCs express OCT4, SOX2 and NANOG, the core pluripotency transcription factors, also expressed in ESCs. However, the gene expression of *Oct4* changes from being regulated mainly by its distal enhancer element in naïve pluripotency to being mainly regulated through its proximal enhancer element in primed pluripotency[276]. Additionally, while NANOG is still present in EpiSCs, its expression is clearly lower than in ESCs, which actually reflects what occurs in the embryo upon implantation[277–279]. Along these lines, the expression of naïve pluripotency genes such as *Rex1*, *Stella*, *Klf4*, *Klf2*, *Prdm14*, *Nr0b1*, *Fgf4* and *Gbx2* is completely shut down while the expression of post-implantation epiblast genes such as *Oct6*, *Fgf5*, *Otx2*, *Lefty* and *Nodal* is turned on[82,83]. Additionally, genes that are not expressed in the post-implantation epiblast, are highly expressed in EpiSCs mainly due to their culture conditions. This is the case for *Tgf1-beta*, *Mapk* and *Wnt*-related genes[120].

Other molecular regulatory players also switch in the naïve-to-primed transition and their effects help shape the two identities. In fact, in comparison with ESCs, EpiSCs show a more closed chromatin structure[280], use different enhancer elements[276] and possess a different set of epigenetic regulators[281] and miRNAs[278]. Additionally, even if the regulatory players are in the same levels in both naïve and primed states, sometimes, their distribution changes. This is the case of H3K4me1, an active enhancer mark whose distribution changes in the naïve-to-primed transition[282]. This change in H3K4me1 distribution plays a key role in establishing the primed identity as perturbing it results in the resetting of EpiSCs towards a more naïve pluripotent state[282].

Additionally, the naïve-to-primed transition also comprises a metabolic switch. ESCs have a bivalent metabolism consisting of both oxidative phosphorylation and glycolysis, whereas EpiSCs have low rates of oxygen consumption and full dependence on glycolysis[283]. In this way, EpiSCs uncouple ATP synthesis from mitochondrial respiration to produce energy through the Krebs cycle by using amino acids and lipids. The resulting glycolytic intermediates may be consumed for one-carbon metabolism or anabolism[83,146,283,284]. Curiously, although mitochondrial electron transport genes are downregulated in EpiSCs, they have longer mitochondria with improved cristae definition compared to ESCs[283]. The trigger of this naïve-to-primed metabolic switch is still unknown, although some factors such as HIF1A[283] and RAS[285] have been implicated, it is thought to reflect the oxygen and metabolic substrate availability in the

embryo. Thus, EpiSC dependence on glycolysis might reflect the implantation of the embryo into the hypoxic uterine wall. Metabolites can directly influence transcriptional and epigenetic programs and hence influence also on cell identity and stem cell fate. Understanding the true roles of the metabolic switch in naïve-to-primed transition and finding the factors governing that metabolic switch will allow us to fully understand the drivers of cell fate and pluripotent state establishment (reviewed in[286]).

In vivo versus *in vitro*: capturing primed pluripotency

EpiSCs can be derived from a variety of embryonic stages. This does not facilitate the identification of their embryonic counterpart if any. Initially, they were thought to correspond to the E5.0-E6.0 epiblast, however, it was later detected that their transcriptional signature fitted more with cells emerging at gastrulation. In line with this, EpiSCs express mesodermal and primitive streak genes such as *Mixl1*, *Brachyury*, *Eomes*, *Gsc* and *Fgf8* but also endodermal genes such as *Gata6*, *Gata4*, *Sox17* and *Foxa2*[82,83,120,155,161]. In fact, E5.75 pre-gastrulation epiblasts show a stronger global transcriptional correlation with 2iL ESCs than with EpiSCs[121]. Actually, EpiSCs are more correlated to E7.25-8.0 embryos, where gastrulation has already started[120]. A more recent study used single-cell RNA-seq and found that the posterior epiblast cells in pre-gastrulating embryos at E5.25- E6.5 stage closely resembled EpiSCs[119], yet, later embryonic stages were not included in this study and this could change that interpretation. What is clear is that EpiSCs monoallelically express imprinted genes[287] and female EpiSC lines show an inactive X chromosome[83,158,163], both of which are traits found at embryonic stages from E6.5 on[194].

All in all, caution must be taken when interpreting EpiSCs traits as they could be a partial reflection of the spontaneous differentiation found in their colonies. In line with this, endodermal genes have been shown to be mainly expressed in differentiated cells within EpiSC colonies[288]. Also within EpiSC colonies, there are subpopulations of cells that coexpress pluripotency genes such as *Nanog*, *Oct4* and *Sox2* together with lineage markers such as *Foxa2* and *Brachyury*[288,289]. Moreover, mesodermal and endodermal markers have been shown to be inversely correlated with neuroectodermal markers in EpiSCs[289], which might be indicative of two other subpopulations within EpiSC colonies. Importantly, EpiSCs expressing *Brachyury* show a tendency to differentiate[274] but, if isolated, they re-establish mixed cultures[289] which points at the idea that the *Brachyury*-expressing subpopulation is not on a settled road towards

differentiation. Additionally, within EpiSC cultures, small subpopulations of cells also express naïve pluripotency genes, use the distal enhancer of Oct4 for its expression and downregulate *Brachyury* and *Fgf5* gene expression[214,277]. These subpopulations are transcriptionally distinct from ESCs[277], but potentially show a naïve-to-primed intermediate.

Finally, there is an extensive amount of variability in subpopulation composition across different EpiSC lines. However, the composition of individual lines is remarkably stable and barely varies as sub-clones are grown across different research labs[120]. Furthermore, this subpopulation variability across EpiSC lines does not seem to correlate with the derivation protocol with which they were obtained (meaning, there is no correlation with whether they were ESC- or embryo-derived or with the embryonic stage from which they were derived)[120]. Therefore, the origin of the variability that exists between individual EpiSCs remains a mystery.

Identity & developmental potential of EpiSCs

As all pluripotential cell types, EpiSCs can differentiate into all germ layers. This has been tested both by *in vitro* differentiation protocols and by grafting cells into embryos. Different subpopulations of EpiSCs have distinct biases towards specific lineages. *Brachyury*-expressing subpopulations, for instance, have a bias towards the mesodermal lineage[289]. Since different EpiSC lines show distinct subpopulation compositions, a variety of differentiation biases have been reported among different EpiSC lines.

Similarly to ESCs, *in vitro* differentiation protocols of EpiSCs use a serum-based medium to direct differentiation toward mesodermal and endodermal lineages, and, serum-free medium for neuroectodermal differentiation[83]. Additionally, the withdrawal of FGF2 and Activin A from EpiSC media or the inhibition of their signalling pathways also results in neuroectodermal differentiation[82,83,268] whereas BMP and WNT signalling direct EpiSC differentiation into mesodermal and endodermal lineages[274]. BMP was also reported to direct EpiSC differentiation into extraembryonic lineages[82], however as mentioned earlier, extraembryonic markers can sometimes be confounding as they tend to also be markers for mesoderm (in the case of trophoblast) and epiblast-derived definitive endoderm (in the case of PrE). In EpiSCs, BMP activates neuroectodermal differentiation through the activation of the WNT pathway, which in turn induce the expression of *Nodal* and *Fgf8*[274].

Precisely, this same group of factors coordinate gastrulation initiation which suggests that *in vitro* differentiation of EpiSCs recapitulates, in part, what occurs during embryogenesis. In ESCs, however, BMP and WNT support self-renewal, not differentiation. This is yet another evidence showing that ESCs and EpiSCs are differently wired to respond to developmental cues.

EpiSCs have a very low capacity to integrate into preimplantation embryos. This is consistent with the fact that epiblast cells of E6.0-7.0 embryos cannot generate chimaeras if injected into blastocysts[126]. In the uncommon cases where EpiSCs integrate into preimplantation embryos, female EpiSCs reactivate their inactivated X-chromosome[290]. This possibly indicates that reprogramming to the naïve state is required for integration into preimplantation embryos.

The low capacity of EpiSCs to integrate into preimplantation embryos has been partially attributed to adhesion incompatibilities. Indeed, ESCs and EpiSCs show very different morphological features. ESCs grow in dome-shaped colonies while EpiSCs colonies have a flat morphology. Also, ESCs have homogenously high levels of E-CADHERIN at cell-cell junctions while EpiSCs show a low and irregular distribution of E-CADHERIN and an upregulation of N-CADHERIN[274]. This shift also occurs in the gastrulating embryo as E-CADHERIN expression is downregulated and coordinated with N-CADHERIN upregulation[291]. This allows for epiblast cells to delaminate from the primitive streak and grants mesoderm formation through EMT resulting in neural tube closure. Furthermore, E-CADHERIN has been reported to regulate naïve-to-primed transition as ESCs with E-CADHERIN depletion adopt a primed phenotype[292,293]. In line with this, overexpression of *E-cadherin* in EpiSCs increases their contribution to preimplantation embryos[290]. However, in this case, their contribution is still very small which indicates that *E-cadherin* expression is not the main discordancy between EpiSCs and preimplantation embryos.

EpiSCs have also been grafted into embryos at the post-implantation stage. Importantly, dissociated EpiSCs do not incorporate into the embryo but cell clumps do[264]. These also disperse from the graft site and initiate proper lineage marker expression. Importantly, EpiSC incorporation into post-implantation embryos is most efficient when injected into the anterior or mid primitive streak and yields unincorporated cell clumps when injected into the posterior primitive streak[120]. Furthermore, *Brachyury*-expressing EpiSCs preferentially form mesoderm and definitive endoderm when grafted into the primitive

streak while *Brachyury*-non-expressing EpiSCs fail to incorporate[264,289]. This agrees with the idea that EpiSCs that do not express *Brachyury* are the in vitro counterpart of the anterior neuroectoderm, and thus, do not enter the primitive streak. Finally, EpiSCs that are grafted into E8.5 embryos, when pluripotency is lost, can also survive but do not efficiently disperse from the graft site or initiate proper lineage marker expression[264].

EpiSCs are known to contribute to all germ layers but their contribution to the germline has not been demonstrated yet. This is mainly due to technical barriers in the testing of that competency. First, it is not possible to re-insert post-implantation chimaeras into recipient females for those embryos to terminate embryogenesis. In the rare cases where EpiSCs were able to integrate into preimplantation embryos, they did not contribute to the germline of the chimaeras[82,290]. In post-implantation chimaeras, EpiSCs contribute to cells that express Alkaline Phosphatase, a trait of primordial germ cells, precisely in the localization where primordial germ cells arise[264,289]. This suggests that EpiSCs might be able to contribute to the germline. On the other hand, EpiSCs can generate primordial germ cell-like cells *in vitro*, yet at very low efficiency[121,294]. This low efficiency might be a reflection of their *in vivo* counterpart as E7.25-8.0 embryos have virtually lost their capacity to generate primordial germ cells[295].

In this line, ESCs just after two days of differentiation in EpiSC medium with 1% KOSR supplement differentiate into a transient cell type termed Epiblast-like stem cells (EpiLCs)[121]. These cells display the hallmarks of pluripotency and are believed to be in an intermediate state between naïve and primed pluripotency[121]. They are transcriptionally similar to the epiblast of E5.75 embryos and downregulate naïve markers such as *Klf4*, *Stella* and *Rex1* while upregulating early primed markers such as *Fgf5* and *Oct6* to the levels of EpiSCs. In contrast, they do not upregulate differentiation markers such as *Brachyury* and *Foxa2* to the levels of EpiSCs[121]. Importantly, unlike EpiSCs, EpiLCs can efficiently generate primordial germ cell-like cells *in vitro*[121].

Shaping cell identities with master splicing regulators

One of the most unexpected and revolutionary findings in molecular biology came in 1977 with the discovery of the sequences that interrupt genes, namely the introns[296,297]. Before this major breakthrough, molecular biology seemed to follow one simple rule: DNA encoded the genetic information in the form of genes[298], these were transcribed into RNA and then translated into proteins[299]. Introns broke that rule as they generated an extra RNA sequence that must be removed by subsequent RNA processing to produce the protein-coding mature messenger RNA (mRNA) that will be translated. The removal of introns was denominated intron splicing and sequences that remained in the mature mRNA were denominated exons[300].

Introns and the molecular tools that remove them from the final transcript, namely the spliceosome, are ubiquitous in eukaryotes. In fact, most eukaryotic genes have at least one intron and generally many[301,302]. The mouse transcriptome, for example, contains on average ~8 introns per protein-coding transcript[303]. Curiously, in evolutionary history, there were long stages in which a small number of introns were gained or lost, separated by episodes of intense and rapid intron gain and loss[301,302]. While many hypotheses have been drawn towards how introns are gained[304–308], to date, only one has been experimentally validated at a genomic scale[309]. This hypothesis proposed that short non-autonomous DNA transposons carrying a splice site in one end independently generated hundreds of intron gain episodes by inserting a novel splice site and co-opting another from the genomic sequence that was duplicated upon transposon insertion[309]. Accordingly, this mechanism intrinsically guarantees the splicing out of the intronic RNA and likely explains some episodes of rapid intron gain in the evolutionary history of eukaryotes.

A year after introns were discovered, Walter Gilbert wrote a seminal perspective article that anticipated the vast implications of having "genes in pieces"[300]. The concepts he presented still invite to think more profoundly on evolution and the functional meaning of genome architecture. Firstly, the different decisions of intron splicing and exon re-joining, in other words, the decision of which exons to join with which, can result in different mature mRNA sequences originating just from a single gene. This process was later named alternative splicing and the different mature mRNA sequences generated were referred to

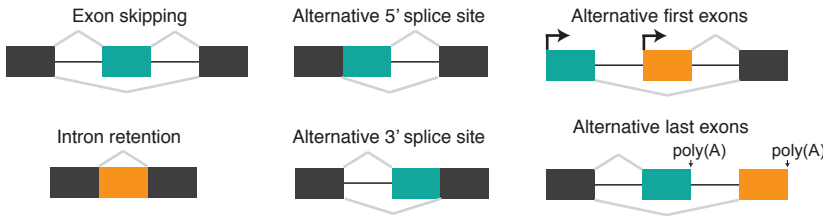
as alternative splice isoforms. Most importantly, alternative splicing provides a window of opportunity for greatly expanding the molecular diversity of the cell. In fact, in his perspective article, Gilbert already foresaw the impact that this molecular diversity could have precisely on promoting differentiation pathways. He proposed that new differentiation pathways could be determined by "splicing enzymes" that could call forth a specific set of splice isoforms. This view is a central principle in my thesis.

Two other key features of "genes in pieces" were proposed by Gilbert[300]. First, that single mutations have the potential to dramatically alter the coding sequence by changing the splicing patterns. Second, that alternative splicing enables genes to sample new functions without affecting the ancestral function of the gene, or as Gilbert cleverly stated: "*evolution can seek new solutions without destroying the old*". Introns provide a mechanism primed to sample evolutionary innovations while alternative splicing grants that evolutionary innovation potential.

The types of alternative splicing choices & its implications

An alternatively spliced exon can be recognized as a single unit and be either included or skipped in the final transcript; in contrast, constitutively spliced exons are always included in the final transcript. This distinction, however, is empirical and a growing number of exons that were once considered to be constitutively spliced have been found to be alternatively spliced as whole-transcriptome sequencing data keeps growing[310,311]. The alternative choice of exon-intron boundaries, termed splice sites, can give rise to diverse types of alternative splicing (Introduction figure 5). In most of the cases, different mutually exclusive splice sites are simultaneously present on the pre-mRNA. Consequently, alternative splicing cannot be simply viewed as a "spliced or not" decision process, where only the intrinsic properties of splice sites are determinating the final outcome, but it should be conceived as a competition where the splicing machinery must discriminate and choose among different splice sites.

Ultimately, alternative splicing plays a major role in expanding transcriptomic and proteomic diversity by generating a wide range of isoforms[312,313]. These isoforms might have similar, slightly different or even opposite functions. Alternative splice site location within the transcript determines whether isoforms will differ in their coding sequence (CDS) or in their untranslated



Introduction Figure 5: Types of alternative splicing patterns. Overview of splicing decisions, or "events". From top left: exon skipping events are produced when a full exon is recognized as an individual unit and consequently is included or skipped in the final transcript; intron retention events are produced when full introns are unrecognized by the splicing machinery and are kept in the final transcript; alternative 5' and 3' events are produced when consecutive 5' and 3' splice sites are alternatively used, respectively; alternative start sites are often generated when splicing is coordinated with alternative promoters; and, alternative terminations are generated when splicing is coordinated alternative poly-adenylation events, poly(A). Adapted from Park et al. 2018[583].

regions (UTRs). The usage of alternative splice sites might affect the translation, the stability or the localization of the final transcript. Additionally, alternative splice site usage can change the reading frame of the transcript resulting in different protein isoforms with different functionalities or localizations.

In vertebrates, a large percentage of multiexonic genes undergo alternative splicing[314,315]. In humans, for example, it is estimated to be up to 95% of multiexonic genes[316]. Many of these alternative splicing decisions, or events, contribute to proteome complexity. In humans, for example, high-resolution mass spectrometry analyses revealed that around 37% of the protein-coding genes generate multiple protein isoforms[317]. Despite this huge contribution to proteome complexity, we know very little on the functional impact of the vast majority of alternative splicing events[311]. Notably, alternatively spliced exons located in protein-coding transcripts are often situated within predicted intrinsically disordered protein regions, which are usually sites of protein-protein interactions and protein post-translational modifications[318–320]. This grants alternative splicing the potential to diversify protein interaction capabilities. A smaller number of conserved alternatively spliced exons directly overlap with critical protein domains and affect various important protein functions such as transcriptional regulation or enzymatic activity[321–323].

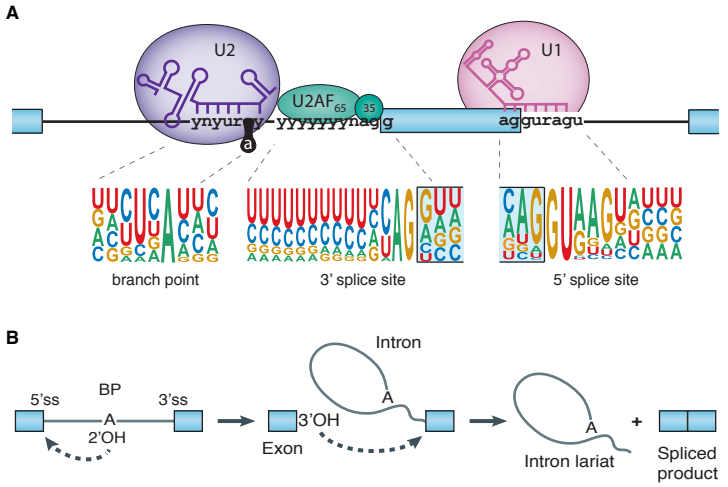
Moreover, not all alternative splicing events produce functional proteins. Firstly, because the transcript could be non-coding and thus, not be translated into protein; secondly, because the stability of the transcript could be affected; thirdly, because transcript localization could be affected and this could modify its function or translation; and finally, because the translated protein could include a domain that renders it non-functional. In line with this, it is important to point out that alternatively spliced transcripts that are non-coding can still have a function as, for example, they could be modulating other RNAs by competing for their regulators.

In recent years, the alternative splicing field has moved from reporting single splicing events and their impact on protein function to describing whole alternative splicing networks and how they are regulated. Despite the great advances in uncovering regulatory networks, we still have a very limited view of their true functional consequence.

The mechanistic insights of splicing

The identification of exon-intron boundaries and the subsequent removal of the intron is guided by four consensus sequence motifs in the pre-mRNA: the 5' splice site, the 3' splice site, the branch point and the polypyrimidine tract (Introduction figure 6a). The 5' splice site is the upstream boundary of the intron and determines where the intron starts while the 3' splice site is the downstream boundary, and determines where it ends. The branch point contains an adenosine and catalyses the first reaction. Finally, the polypyrimidine tract is at the 3' end of the intron and helps in its recognition. The remainder of the intronic sequence is generally variable, evolutionarily unconstrained and, in some cases, can extend up to a million nucleotides[324].

Intron removal occurs in two transesterification steps, starting with the nucleophilic attack of the 5' splice site by the adenosine in the branch point and the cleavage of the phosphodiester bond between the last nucleotide of the upstream exon and the first nucleotide of the intron (Introduction figure 6b). This generates an intron lariat intermediate and a free 3'OH group at the end of the 5' exon. This free 3'OH group then carries out a nucleophilic attack on the 3' splice site which leads to the release of the intron and the joining of the two exons[325,326].



Introduction Figure 6: Intron recognition and removal. (A) Depiction of consensus sequence motifs mediating the identification of exon-intron boundaries and the factors that recognize them. The height of the residues represented in the motif indicates its relative frequency in a given position within the sequence. **(B)** The two transesterification reactions occurring in pre-mRNA splicing. BP: branch point, ss: splice site. Adapted from Scotti and Swanson 2016[584].

The spliceosome is a remarkably complex ribonucleoprotein machinery that executes intron removal in eukaryotes. These introns, named spliceosomal introns, are removed from the pre-mRNA in the nucleus soon after they are transcribed. Spliceosomal introns originated from pre-existing self-splicing introns, and expanded abruptly and at the time of the origin of eukaryotes[301,327]. More specifically, it is thought that the ancestors of these spliceosomal introns are self-splicing group II introns[301]. These are mobile ribozymes present in bacterial, mitochondrial and chloroplast genomes but absent from eukaryotic nuclear genomes that self-splice through a very similar process as that of spliceosomal introns[327,328].

The spliceosome is composed of five small ribonucleoproteins complexes: U1, U2, U4, U5 and U6 snRNP and over two hundred proteins[329]. The core components of the spliceosome recognise the sequence motifs in the pre-RNA and mediate the structural rearrangements to catalyse the splicing reaction. Unlike many other enzymes, the spliceosome does not possess a preformed catalytic centre and must build it *de novo* on the pre-mRNA in a step-wise manner for every splicing event. First, the U1 snRNP recognises

and binds the 5' splice site while the U2-auxiliary factors U2AF1 and U2AF2 (also known as U2AF35 and U2AF65) recognise and bind the 3' splice site and the polypyrimidine tract, respectively[329–332] (Introduction figure 6a). Concomitantly, SF1 binds to the branch point sequence[329]. The binding of these snRNA and proteins forms what is known as the E complex. Next, U2 snRNP is recruited to the branch point in an ATP-dependent manner, this displaces SF1 and forms the A complex[332]. Then, the U4/U6.U5 tri-snRNP is recruited, forming the pre-B complex. This leads to conformational changes that displace both the U1 and the U4 snRNPs and generate a catalytically active complex, named the B* complex[333]. Splicing proceeds with the two transesterification steps (C and C* complex) after which the spliceosome components disengage and are recycled, the intron lariat dissociates and the two exons are ligated[333–335]. Very recently, cryo-electron microscopy has validated years of genetic and biochemical research providing major insights into the assembly, activation, catalysis and disassociation of the spliceosome (reviewed in [336]).

One of the most fascinating things in the splicing field is that, despite having so much knowledge on many of the smallest details of spliceosome assembly, activation, catalysis and disassembly, we still do not know the process by which the spliceosome components at the splice sites interact with each other and recognise the primary unit to splice. Two models have been suggested: the intron definition model and exon definition model[337–339]. The use of one or the other has been proposed to depend on intron-exon architectures. In the intron definition model, the splicing machinery pairing occurs across introns, and thus, the intronic unit is recognized. In principle, this type of recognition would be impaired in long introns suggesting that intron-defined introns could be under evolutionary selection to remain short. However, vertebrates have markedly long introns along with shorter exons in comparison to other animals[340,341]. This prompted the need for an alternative model, the exon definition model, where the splicing machinery forms across the exon[337]. Importantly, in this model, the exon unit is recognized instead of the intron unit. This then leads to a second step that swaps the splice site pairing configuration from one where the splice sites pair across the exon to one where the splice sites pair across an intron. This conversion from exon definition to intron definition is essential for subsequent steps of spliceosome assembly. If it is obstructed, it could result in exon skipping[342].

The Rules of Alternative Splicing Regulation

The regulation of alternative splicing determines when and where a specific isoform is produced. Several players are key in this regulation. On one hand, sequence features on the pre-mRNA, termed *cis*-acting features, promote or repress splicing at particular sites. On the other hand, RNA-binding proteins (RBPs), termed *trans*-acting regulators, bind on the pre-mRNA at particular *cis*-regulatory sites and regulate splicing (Introduction figure 7). The regulatory roles of *trans*-acting factors are highly context-dependent, in part due to the combinatorial, synergistic and competitive effects among them[343]. In this section, will give a brief overview of the relative contributions of both *cis*- and *trans*-acting factors.

The Regulatory role of *cis*-acting sequence features

The *cis*-acting features involved in alternative splicing can be grouped into: (i) the splice sites needed for spliceosome assembly; (ii) the motifs recognized by splicing enhancers or silencers; and (iii) other sequence features, such as exon-intron architecture and mRNA secondary structure.

Splice sites are highly conserved motifs with leading sequences (GU and AG dinucleotides at the 5' and 3' splice sites, respectively) that are essential for splicing and with a surrounding nucleotide context that can vary. Splice sites can be described as being "weak" or "strong" depending on how favourable the sequence is for splice site recognition, which, in turn, greatly depends on the surrounding nucleotide context. Weak splice sites are particularly common in alternative 3' and alternative 5' splicing in which exons have a different start or end coordinates, respectively. Strong splice sites are common in constitutively spliced exons. Splice site strength can be calculated using entropy-based modelling of the exon-intron boundary[344], which largely serves as a measure of how dissimilar a splice site sequence is to the complementary sequences in the snRNAs that bind them.

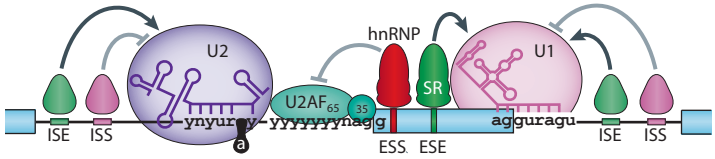
The motifs acting as binding sites for *trans*-factors also play a crucial role in promoting or repressing spliceosome assembly by recruiting RBPs[345]. Additionally, they can alter the accessibility of nearby features. These motifs are classified depending on their effect in splicing (enhancer or silencer) and their location (exonic or intronic). Importantly, these sequences may not be motifs in the classical sense as they can be quite short and degenerate or simply

consist of repeat-rich elements. As a consequence, the identification of these motifs has proven to be quite challenging. Ongoing efforts are systematically mapping the binding sites of RBPs in the transcriptome[346] and could shed some light on the binding preferences of RBPs. However, some motifs are multivalent and recruit a diverse number of RBPs. In fact, many RBPs have been found to act cooperatively to interact with repetitive multivalent motifs that are located within or close to their regulated alternative exons[347–350]. Biophysical studies showed that these multivalent interactions can trigger condensation into dynamic complexes through liquid-liquid phase separation[351–353]. This condensation, termed RNP condensation, likely enables RBPs to bind with more stability to multivalent motifs in pre-mRNA transcripts than to a single RBP binding motif[353].

Exon-intron architecture and mRNA secondary structure also influence splice site recognition. In the case of exon-intron architecture, early observations reported that expanding the size of mammalian exons led to exon skipping and that these same exons were also excluded if the flanking introns were made smaller[354]. This is likely to do with the mechanism by which splice sites pair but, while efforts have been done in trying to understand how exon-intron architecture impacts splice site recognition[341,355,356], it remains largely unknown. On the other hand, mRNA secondary structure affects splice site recognition by modulating the accessibility to splice sites or RBP binding motifs[357]. One simple example would be and RNA hairpin concealing a splice site and keeping it from being recognized by the spliceosome. In this case, mRNA secondary structure would drive exon exclusion[358,359]. Another example would be the formation of RNA duplexes, which have been found to "loop out" entire exons[360,361] and to mediate long-range interactions (RNA-bridges) with other parts of the transcript[362]. These structures can help in bringing splicing elements into close proximity[363] or in exposing enhancer or silencer motifs[364]. Moreover, RBPs sometimes recognize RNA structures rather than linear sequences, which implies that sometimes RNA structure can be critical in alternative splicing regulation[365].

The regulatory role of *trans*-acting splicing regulators

There are two main RBP families involved in alternative splicing regulation: the heterogeneous nuclear ribonucleoproteins (hnRNPs)[366] and the serine/arginine-rich proteins (SR proteins)[367]. RBPs from both families generally bind to the pre-mRNA in a sequence-specific manner[368]. In most cases, hnRNPs



Introduction Figure 7: Regulation of alternative splice site recognition. The binding of splicing regulators to *cis*-acting sequence features can influence spliceosome assembly and splice site recognition. ESS: Exonic splicing silencer, ESE: Exonic splicing enhancer, ISS: Intronic splicing silencer, ISE: Intronic splicing enhancer, hnRNP: heterogeneous nuclear ribonucleoprotein, SR: serine/arginine-rich protein. Adapted from Scotti and Swanson 2016[584].

contain various types of RNA-binding domains and interact with other proteins through unstructured regions[369]. On the other hand, SR proteins usually have one or two RNA-recognition motifs (RRM) domains at their N-terminus and an arginine/serine-rich (RS) domain with which they interact with other proteins at the C-terminus[370]. SR proteins are mostly considered to promote exon inclusion. This largely occurs through the recruitment of either U1 snRNP to the 5' splice site or of U2AF to the 3' splice site[370,371]. On the other hand, hnRNPs are commonly considered to promote exon skipping by binding to intron splicing silencers[343]. However, this view of positive and negative splicing regulators has turned out to be overly simplistic. Firstly, because the role of these proteins changes considerably depending on their position within the pre-mRNA. Secondly, because there are potent competitive and cooperative binding effects among SR and hnRNP family members[343,372]. These include strong auto- and cross-regulatory loops among the proteins themselves.

There are many different modes of action among SR proteins and hnRNPs but they mainly involve recruiting spliceosomal components, changing the spatial proximity between exons and obstructing splice site access[349,373]. Additionally, one splicing regulator may possess multiple mechanisms to regulate splicing. One good example of this is the many ways the polypyrimidine tract-binding protein (PTBP1), an hnRNP family member, regulates splicing. When PTBP1 binds within introns, it blocks the recruitment of factors needed for spliceosomal assembly[374], but it can also result in the "looping out" of entire exons to promote exon skipping[375]. Additionally, when it binds to the 3' or 5' end of introns flanking an alternative exon, it mostly promotes its exclusion but if it binds closer to the adjacent downstream exon, it promotes its inclusion[376].

Furthermore, PTBP1 has been reported to block the switch from exon to intron definition showing that it can also inhibit splicing even if both the 5' and 3' splice sites have already been properly bound by U1 and U2 snRNPs, respectively[377].

Building cell identities with alternative splicing regulatory networks

Isoform composition changes profoundly across tissues[315,316,378]. In fact, to date, dynamic alternative splicing networks have been identified in adult tissues such as muscle, brain, heart, liver, kidney, adipose tissue, testis, adrenal glands and ovaries[311,378–383]. Moreover, specific splicing programs have been reported in embryonic stem cells and precursor cells[384,385], throughout differentiation and reprogramming of different cell lineages[386–389], and in epithelial-to-mesenchymal transitions[390,391]. Additionally, other splicing networks have been associated with particular physiologic cellular states such as thermal regulation or stress response[392–395]. Many splicing regulators function in a tissue-, cell type-, or condition-specific manner and can coordinate functionally coherent networks of introns and exons. This is partially due to different tissue-dependent expression patterns of some splicing regulators, but also due to RBP binding preferences, as specific RBP motifs can also be specifically enriched in genes involved in the development of a specific cell type or tissue (reviewed in[382,396,397]).

Alternative splicing in the maintenance & establishment of pluripotency

Several pluripotency factors have distinct isoforms that vary in function, stability or intracellular localization due to their different alternative splicing events affecting either their coding sequences or their untranslated regulatory regions. One example is OCT4, which has two isoforms. One, OCT4A, is the canonical pluripotency transcription factor expressed in pluripotential cells; the other, OCT4B, is a cytoplasmic protein with unidentified functions and expressed in non-pluripotential cells[398]. Other examples include TCF3, which has two isoforms with different degrees of *Nanog* and *Oct4* transcriptional inhibition capabilities[399]; NANOG, which has three isoforms with varying efficiencies in pluripotency maintenance[400]; and, DNMT3B which has over 40 isoforms that are differentially expressed in pluripotent cells and differentiated cell types[401]. However, to date, the most captivating example of regulated isoforms in pluripotency is that of FOXP1, a transcription factor. The transcript of *Foxp1* bears a conserved exon that is alternatively spliced and uniquely

included in ESCs[384]. This alternative splicing event changes the DNA-binding preference of FOXP1. When this ESC-specific exon is included in the final transcript, FOXP1 stimulates the expression of pluripotency transcription factor genes including *Oct4* and *Nanog*, represses the transcription of genes involved in differentiation and lineage specification, promotes self-renewal as well as maintenance of ESC pluripotency and contributes to efficient somatic cell reprogramming to induced pluripotent stem cells[384].

Different high-throughput analyses have profiled alternative splicing networks during somatic cell reprogramming to pluripotency[389,402–405]. Importantly, all have shown a multiphasic rewiring of alternative splicing landscapes from diverse differentiated cell states towards a pluripotent state, suggesting that rewiring of alternative splicing networks is required for pluripotency acquisition. Various splicing regulators have been shown to facilitate the acquisition of pluripotency during reprogramming, these are SON, RBFOX2, SRSF2, SRSF3, MYC, GCN5, ZCCHC24, U2AF1, RBM47, ESRP1 and ESRP2[389,402,404,406–409]. In contrast, MBNL1, MBNL2, RBM24 and SFRS11 have been shown to promote the differentiated cell state and act as repressors for reprogramming[385,405,410]. Additionally, the splicing factor SFRS11 regulates alternative splicing events in MBNL1, SFRS3 and U2AF1 which indicates that it could be regulating reprogramming via the splicing of downstream splicing factors[405].

While all previous studies have broadened our understanding of how alternative splicing can shape pluripotency, none of them has ever made distinctions between naïve and primed pluripotency. Over the course of this thesis, one study found that the splicing factor HTATSF1 played a role in naïve-to-primed pluripotency transition[411]. More specifically, HTATSF1 mediates intron removal from around 45 ribosomal protein transcripts and thus mediates ribosomal abundance and protein-synthesis. The authors reported that HTATSF1 is essential for the naïve pre-implantation epiblast to be able to transition into the primed post-implantation epiblast as *Htatsf1 KO* embryos failed to undergo amniotic cavity fusion. Furthermore, they reported that downregulation of HTATSF1 is crucial in the establishment of primed pluripotency as they were not able to derive EpiSCs from *Htatsf1*-overexpressing ESCs. Conversely, they showed that overexpressing *Htatsf1* facilitated the conversion of EpiSCs into reverted ESCs. To our knowledge, this has been the only study reporting alternative splicing differential regulation in naïve and primed pluripotency. Thus, for years, the naïve and primed pluripotency alternative splicing

networks have been overlooked. We believe that alternative splicing networks are part of what defines a cell state and that their regulation not only helps in building robust transcriptomes but also drives critical cell state transitions. Therefore, we consider that by identifying the alternative splicing regulatory networks of naïve and primed pluripotent stem cells, we will be able to learn more about these two pluripotent cell identities.

Objectives

Objectives

Objectives

To determine alternative splicing dynamics in the naïve-to-primed transition, uncover how they are regulated and understand their impact on cell identity, we established the following aims in this thesis:

1- Profile the alternative splicing landscapes of naïve and primed pluripotency using transcriptomics

2- Identify potential alternative splicing regulators responsible for establishing these alternative splicing landscapes:

   a. Candidate-based approach for identification of alternative splicing regulators in mESCs

   b. High-throughput identification of alternative splicing regulators using SPAR-seq

3- Detailed *in vitro* characterization of the impact of master regulators on alternative splicing and of their function in pluripotency, naïve-to-primed transition and cell differentiation.

Results

Chapter 1: Transcriptomic alternative splicing profiling of naïve and primed pluripotency

To examine the alternative splicing profiles of naïve and primed pluripotency, we performed deep coverage paired-end polyA-selected RNA sequencing (RNA-seq) of mouse embryonic stem cells (ESCs) cultured in two different media with either two inhibitors/LIF (2iL) or serum/LIF (SL); and of mouse epiblast stem cells (EpiSCs) cultured with Activin A and bFGF (Figure 1A). We used three independent well-characterized ESC cell lines: E14[412], 46C [250] and DizFx [413] and the corresponding mEpiSC cell lines derived directly from them. On average, we sequenced ~90 million reads per sample.

First, we assessed the overall transcriptomic similarity between samples based on their gene expression profiles. Principal component analysis (Figure 1B) and sample-to-sample correlation distances (Figure 1C) consistently clustered our nine samples based on culture conditions and not by cell line. Additionally, mESCs grown in 2iL and SL were more similar to each other than they were to EpiSCs (Figure 1C), in line with their presumed naïve and primed pluripotency cell identities. The expression of naïve and primed markers across our nine samples further corroborated their naïve and primed pluripotency cell identities (Figure 1D).

Alternative splicing profiling in naïve and primed samples

We used *vast-tools*[380] for detecting alternative splicing events and quantifying their inclusion levels in each sample using the Percent Spliced-In [PSI] metric. We compared transcriptome-wide inclusion levels of alternatively spliced (AS) cassette exons across our samples through principal component analysis (Figure 2A) and sample-to-sample correlation distances (Figure 2B) and found that the nine samples clustered according to their pluripotency state. We obtained similar results with transcriptome-wide retention levels of alternatively retained introns (Figure 2C, D). Importantly, the percentage of the variance explained by the clustering according to cell culture condition was lower for alternative splicing than for gene expression, which pointed at higher variability across cell lines at the level of alternative splicing (Figure 1B, C). This was especially evident for transcriptome-wide inclusion levels of AS

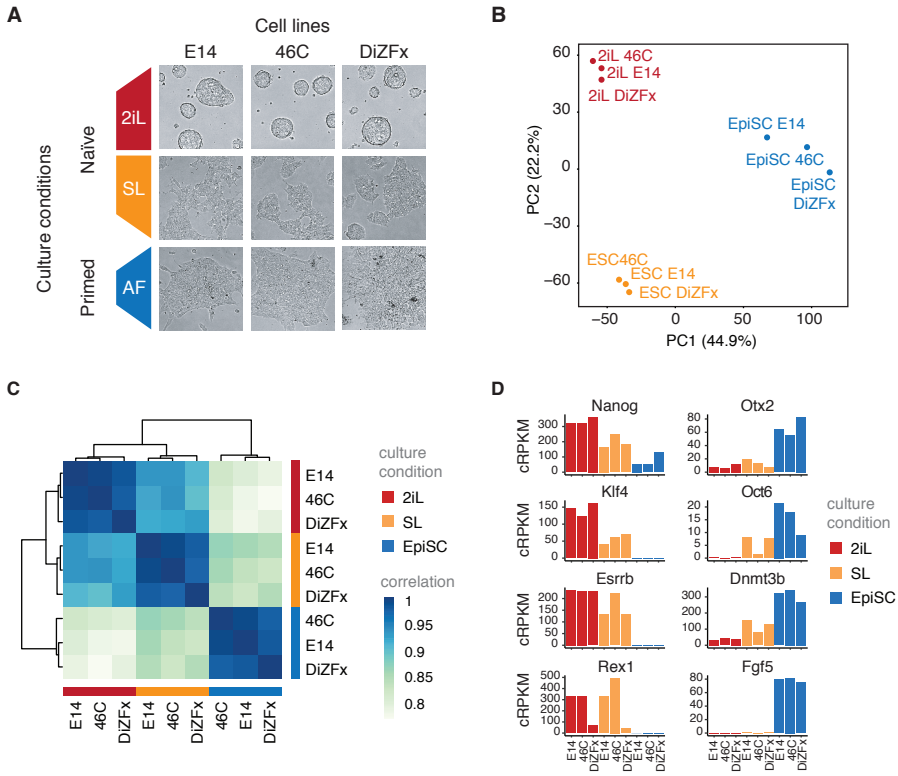


Figure 1: Experimental strategy to characterize the alternative splicing profile of naïve and primed pluripotency (A) Three independent cell lines were used: E14, 46C and DizFx. These were grown in either naïve (2iL mESCs or SL mESCs) or primed conditions (AF EpiSCs). A representative image of each condition and each cell line is shown. **(B)** Principal component (PC) analysis of the 9 RNA-seq samples generated, based on their gene expression profiles. The two components accounting for most of the variability between samples are represented. **(C)** Heatmap of sample-to-sample gene expression correlation distances. **(D)** Expression levels of several classical markers of naïve (left) and primed (right) pluripotency. cRPKM: reads per kilobase and million mapped reads, corrected for mappability. 2iL: 2i/LIF ESCs; SL: Serum/LIF ESCs; AF: Activin/bFgf EpiSCs.

cassette exons, where the DizFx cell line consistently had lower correlation levels with the other two cell lines across all culture conditions (Figure 2A, B).

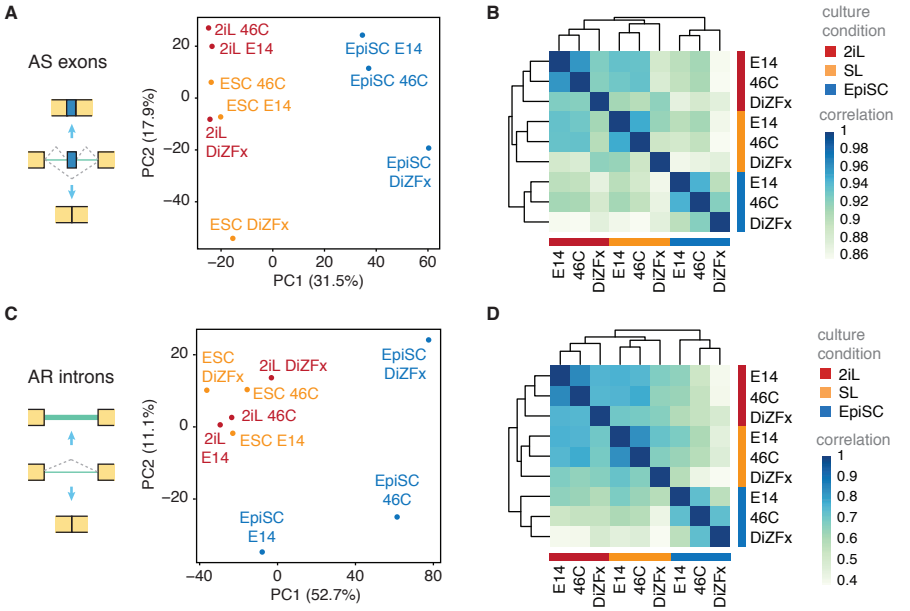


Figure 2: Global alternative splicing landscape in naïve and primed pluripotency **(A)** Principal components analysis based on the transcriptome-wide inclusion levels of alternatively spliced (AS) cassette exons. The two first components (PC) are represented. **(B)** Sample-to-sample spearman correlation distance matrix based on the inclusion profile of alternatively spliced exons. **(C)** Principal components analysis based on the retention levels of alternatively retained (AR) introns transcriptome-wide. The two components (PC) explaining most of the variability are represented. **(D)** Sample-to-sample spearman correlation distance matrix based on the genome-wide profile of alternatively retained introns. 2iL: 2i/LIF ESCs; SL: Serum/LIF ESCs.

Next, we performed pairwise comparisons between each of our culture conditions. We defined differentially spliced events following two rules. First, the average PSI change (dPSI) between the two compared conditions in the three cell lines had to be of at least 15%. Second, we performed a paired comparison by cell line, requiring a minimum dPSI of 5% in the same direction in the three cell lines independently. With these cut-offs, we detected 1499 differentially regulated alternative splicing events across all pairwise comparisons, with over 84% of them (1266 events) being AS between naïve and primed conditions, i.e. events that are differential when comparing 2iL ESCs vs. EpiSCs and/or SL ESCs vs. EpiSCs (Figure 3A). The majority of these events are AS cassette exons, namely, exon skipping events, with no clear bias

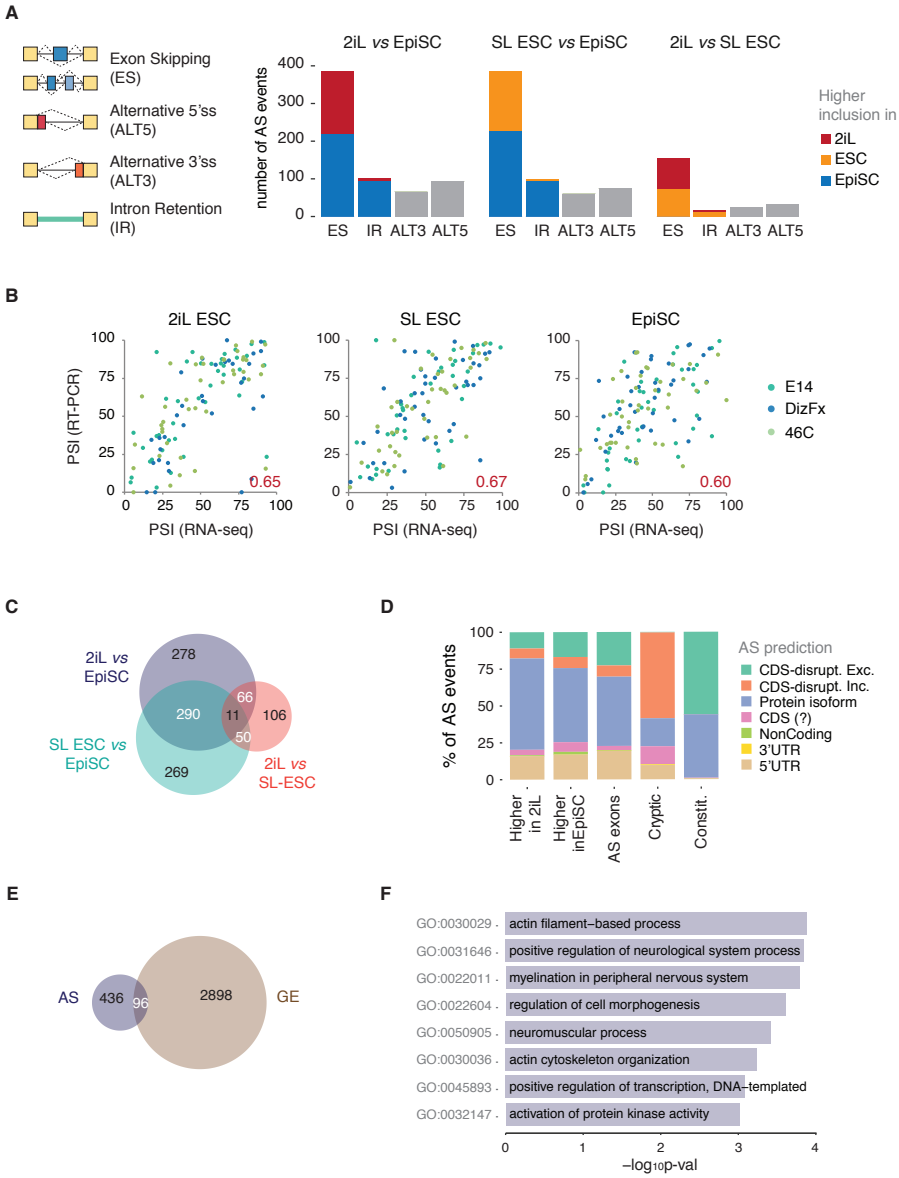


Figure 3: Alternative splicing switch between naïve and primed pluripotency **(A)** Genome-wide numbers of alternative splicing (AS) events for each pairwise comparison of pluripotent samples. Each of the comparisons was paired by cell line, requiring a minimum difference in percentage of inclusion/retention of 15% and a range of 5% in each cell line. **(B)** Scatter plot of differential cassette exon inclusion estimated using RT-PCR or RNA-seq across the different cell lines and cell culture

towards inclusion or exclusion between the two pluripotent states. In contrast, over 95% of the identified differentially retained introns are more retained in primed pluripotency (Figure 3A).

We analysed by RT-PCR 41 naïve-to-primed differentially spliced cassette exons in each of the three cell lines and each of the three culture conditions using independent samples from the ones used for RNA-seq analysis. Comparison with RNA-seq PSI estimations displayed overall satisfactory correlations in all three cell lines and cell states (Figure 3B). Thus, we concluded that our *in vitro* culture system reproducibly recapitulated changes in alternative splicing matching their state of pluripotency.

When comparing alternative splicing events that were differentially regulated between culture conditions - 2iL ESCs vs. EpiSCs; SL ESCs vs. EpiSCs and 2iL ESCs vs. SL ESCs - we found significantly fewer differential events occurring in the 2iL ESCs vs. SL ESCs comparison (16% of the total events identified) (Figure 3A). This suggests that the alternative splicing landscape in our dataset is shaped by the pluripotency state more than it is driven by media conditions. Additionally, we found that over 50% of differentially spliced events overlap when comparing 2iL ESCs vs. EpiSCs and SL ESCs vs. EpiSCs, further supporting this finding (Figure 3C).

---

conditions. Pearson correlation coefficient between estimated exon inclusion values in RT-PCR and RNA-seq is shown on the bottom right corner of each plot. All RT-PCR validations were performed on independent samples with respect to the RNA-seq. **(C)** Venn diagram of overlapping alternative splicing events among the three different cell culture media comparisons showing considerable overlap between 2iL vs. EpiSC and SL vs. EpiSC comparisons. **(D)** Predicted effect of differentially spliced exons on their cognate transcripts for 5 exon categories: higher inclusion in 2iL samples, higher inclusion in EpiSC samples, alternative exons not differentially spliced (AS exons), exons with very low inclusion levels or "cryptic" and exons with very high inclusion across samples or "constitutive" exons. CDS-disrupt: disruption of coding sequence, Exc: exon exclusion or skipping, Inc: exon inclusion, "?": uncertain effect, UTR: untranslated region. **(E)** Venn diagram of number of genes regulated by differential gene expression (GE) and differential alternative splicing (AS) in the 2iL vs EpiSC comparison. **(F)** Gene Ontology (GO) terms enriched in the genes with alternative exons differentially spliced between the 2iL and EpiSC conditions. Bars correspond to the p-values of the enrichment of each GO term.

To evaluate the potential impact on the proteome of the identified alternative splicing isoform diversity, we predicted the effect of cassette exon inclusion on their corresponding transcripts. Over 70% of the differentially spliced 2iL ESCs vs. EpiSCs cassette exons mapped to the coding sequence (Figure 3D). Around one-third of them were disruptive, i.e. altered the reading frame of the transcript, but with no bias towards being disruptive either in the naïve or the primed state. Still, the majority of differentially spliced 2iL ESCs vs. EpiSCs cassette exons potentially code for alternative protein isoforms, in a higher proportion than the average AS exon for the case of exons higher included in naïve samples (Figure 3D, p-val from Fisher's exact test = $6.6x10^{-4}$). Remarkably, only 18% of differentially spliced 2iL ESCs vs. EpiSCs events occurred in transcripts that were differentially expressed (Figure 3E), suggesting that regulation at the level of gene expression and of alternative splicing are two independent regulatory layers in the naïve-to-primed transition.

To understand if specific biological processes or pathways are affected by alternative splicing in the naïve-to-primed transition, we performed Gene Ontology (GO) analyses of the genes with differential 2iL ESCs vs. EpiSCs frame-preserving AS exons. We found significant enrichment in gene ontology terms related to neuronal biology or classically associated with neural splicing such as actin cytoskeleton[311], as well as a mild enrichment in genes associated with transcriptional regulation (Figure 3F).

Dynamics of differentially spliced cassette exons in the naïve-to-primed transition and beyond

For subsequent analyses, we focused on cassette exons differentially spliced in the 2iL ESCs vs. EpiSCs comparison (Figure 2A), termed "2iL-EpiSC differential AS exons". We examined the regulation of these 2iL-EpiSC differential AS exons across embryonic and adult differentiated tissue samples to understand if there was any tendency towards a specific differentiated cell type alternative splicing profile. As expected, hierarchical clustering based on the correlation in the inclusion levels of these exons clusters the two naïve conditions together, and separately from primed EpiSCs (Figure 4A; each condition is represented by the average PSI across the three lines). Curiously, the primed samples cluster together with neural tissues and particularly of those of early embryonic origin and separately from the rest of differentiated tissues and pluripotent samples. Another interesting observation is that the inclusion levels of 2iL-EpiSC differential AS exons had a very poor correlation

between neural tissues and naïve ESCs, notably much lower than with any other differentiated tissue.

We then investigated the distribution of PSI levels of 2iL-EpiSC differential AS exons across tissues. We found that while 2iL ESCs have intermediate inclusion levels of these exons, the distribution becomes bimodal with a tendency towards higher inclusion in EpiSCs, with an even more pronounced bimodality in differentiated tissues, where more extreme inclusion and exclusion values are predominant (Figure 4B). Importantly, 2iL ESCs and SL ESCs have evident differences in the average distribution of inclusion levels of differential alternatively spliced 2iL-EpiSC exons: although both show mostly intermediate PSI levels, there is a clear bimodal distribution of inclusion levels in SL ESCs. SL ESCs are morphologically heterogeneous and show transcriptional fluctuations of some pluripotency factors compared to 2iL ESCs. Some claim that these fluctuations represent a dynamic equilibrium between self-renewal and differentiation and that SL ESCs are in fact in a metastable pluripotency state preceding primed pluripotency [239] (see introduction section). Our results may reflect this phenomenon also at the level of alternative splicing.

To better understand the inclusion patterns of differential cassette exons in the naïve-to-primed transition, we separated 2iL-EpiSC differential exons into six groups according to their inclusion level dynamics (Figure 4C). The vast majority of 2iL-EpiSC differential exons had state-specific dynamics, either primed or naïve. On the other hand, 2iL-specific events (either towards inclusion or exclusion) represented only 16% of the total 2iL-EpiSC differential AS exons. This agrees with our previous observations pointing towards a naïve vs primed specific AS landscape, modestly affected by the choice of naïve cell culture conditions.

To explore the inclusion dynamics beyond pluripotency, we compared the inclusion levels of 2iL-EpiSC differential AS exons in EpiSCs with their average PSI in differentiated tissues, for each of the six groups independently. Interestingly, for the majority of exons, the inclusion levels of differentiated tissues tended to change towards the 2iL inclusion patterns (Figure 4D). This suggested that the primed AS landscape was far from being an intermediate towards the AS profiles of differentiated cell states, at least in respect to 2iL-EpiSC differential AS exons.

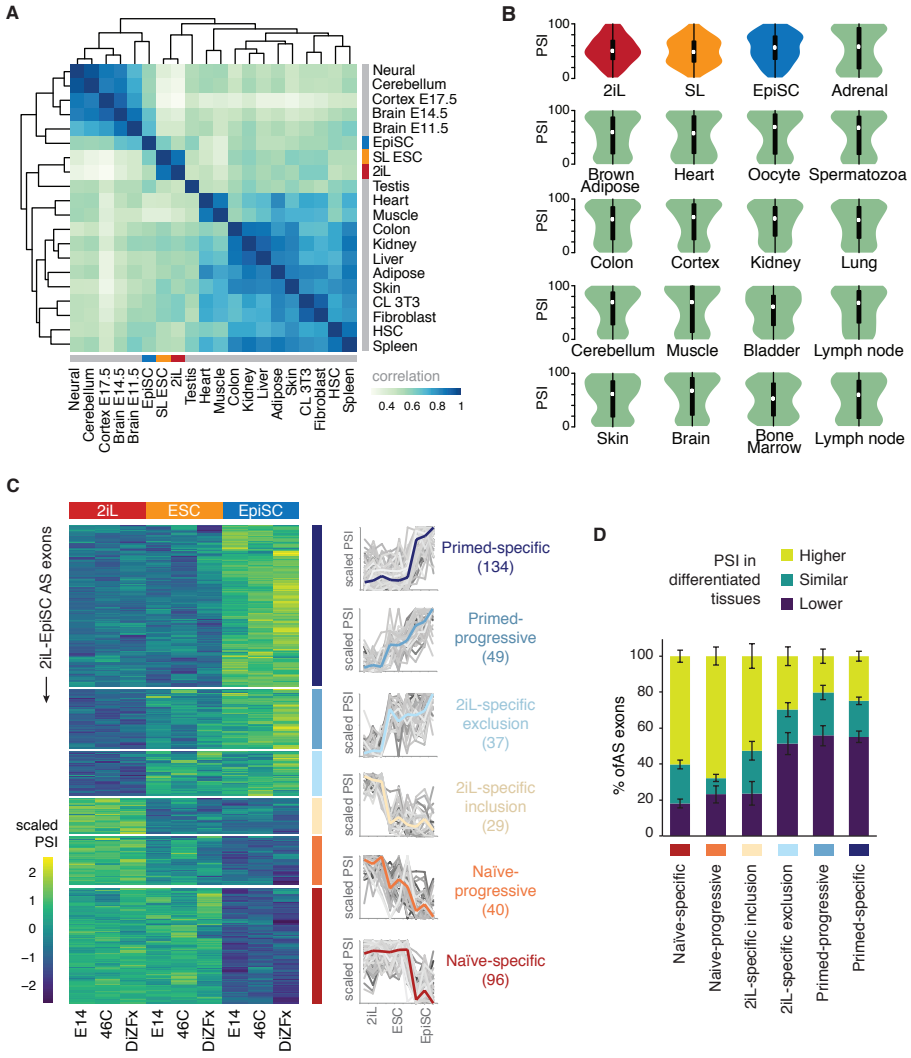


Figure 4: Inclusion dynamics of 2iL-EpiSC differential AS exons. **(A)** Heatmap of Spearman correlation values across pluripotent samples and differentiated tissues generated using 2iL-EpiSC differential AS exon PSI values. **(B)** Distribution of average 2iL-EpiSC differential AS exon PSIs in pluripotent samples and differentiated tissue samples shows bimodality among differentiated samples and central distribution among 2iL ESCs. **(C)** Heatmap depicting scaled PSI levels of 2iL-EpiSC differential AS exons across all pluripotent samples shows separation into 6 independent groups according to their inclusion "dynamics". On the right, each plot represents the PSI dynamics for each exon group. Grey lines represent the scaled inclusion dynamic of

Genomic features of differentially included naïve-to-primed alternatively spliced exons

We examined the intronic and exonic features of naïve and primed differentially included cassette exons to shed light on how they are potentially regulated. For this, we subdivided the 2iL-EpiSC differential AS exons into two categories: higher inclusion in 2iL and higher inclusion in EpiSCs. We found that exons that were highly included in EpiSCs (in blue), were significantly longer ($p < 0.01$, Mann-Whitney U test), had shorter flanking introns and had a higher GC content than those that were more included in 2iL ESCs (Figure 5A-C). These features have been previously associated with exon vs intron-defined splicing and would suggest that exons more included in 2iL ESCs have a strong signature of exon-definition whereas those more included in EpiSC show features closer to those associated with intron-definition. The exonic and intronic length patterns of exons with higher inclusion in 2iL ESCs were similar to those of alternative non-differential exons whereas those exons which inclusion is higher in EpiSCs are more similar to constitutive exons (Figure 5A, B).

We sought to find whether this association between exon-intron lengths and inclusion in naïve vs primed conditions occurred genome-wide for all AS exons. Indeed, we detected that longer exon length as well as shorter intron lengths were associated with increased exon inclusion levels in EpiSCs compared to 2iL ESCs (Figure 5D, E). Additionally, higher exonic GC content was also associated with higher inclusion in EpiSCs compared to 2iL ESCs (Figure 5F). We tested these traits for AS genome-wide in the three cell lines and found the same results in each of them independently (Figure 5G). Finally, we constructed an exon-vs-intron definition score that considers exon GC-content and the difference in intron and exon lengths to summarize these features in a single metric (see materials and methods). Using this score, we investigated whether 2iL-EpiSC differential AS exons with different dynamics could have different exon vs. intron definition biases and thus, potentially have intrinsically different splicing regulatory features. This analysis revealed

---

one exon across the 9 samples. Coloured lines represent the mean scaled inclusion dynamic of each group. Numbers of cassette exons belonging to each group are indicated between brackets. **(D)** Stacked plot illustrating the percentage of 2iL-EpiSC differential AS exons that either increase, decrease or maintain their average PSI levels in differentiated tissues with respect to their EpiSC PSI levels.

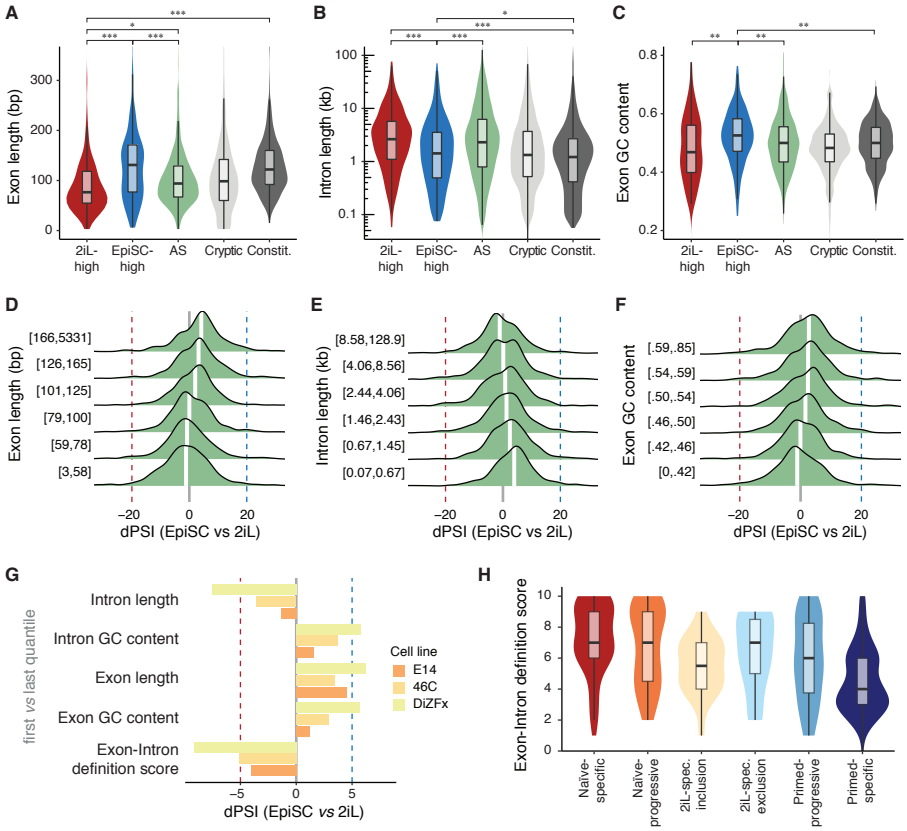


Figure 5: Splicing regulatory features of alternative splicing events in naïve and primed pluripotency (A) Boxplot of exon lengths across 5 exon categories: 2iL-high: AS exons that have an inclusion level in 2iL at least 15% higher than in EpiSC. EpiSC-high: AS exons that have an inclusion level in EpiSCs at least 15% higher than in 2iL. AS: all other alternatively spliced exons (PSI between 10% and 90% in any sample). Cryptic: lowly included exons with a maximum PSI lower than 10% across all samples. Constitutive: highly included exons with a minimum PSI higher than 90% across all samples. bp: base pairs. **(B)** Boxplot of the average surrounding intron lengths across different exon groups. kb: kilobases. **(C)** Boxplot of exonic GC content (fraction of bases in the exon corresponding to guanine or cytosine residues) across different exon groups. P-values from Mann-Whitney U-tests are indicated for several comparisons. * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$. **(D-F)** Distribution of changes in PSI (dPSI) values between EpiSC and 2iL conditions, for all alternative exons genome-wide classified in six quantiles based on the exonic length (D), intronic length (E) and exonic GC content (F). White lines correspond to the median value of each sextile group. In brackets, the interval values for each corresponding to each sextile. **(G)** Bar plot depicting the average difference in dPSI (between EpiSC and 2iL ESCs) between the first and last

a strong gradual association between the inclusion dynamics in 2iL vs EpiSCs and the exon-vs-intron definition score, with the highest difference found between the naïve-specific vs primed-specific exons (Figure 5H).

Conservation of naïve-to-primed alternative splicing profile in human pluripotency

To test whether similar changes in the alternative splicing profile also occur in human naïve and primed pluripotency, we first compiled all studies to date that have *in vitro* naïve and primed pluripotent human samples isolated from embryos [146,150,152,414–416]. We selected those with enough coverage and number of replicates for robust alternative splicing quantification and used *vast-tools* for mapping, profiling and comparing alternative splicing events in each dataset. To ensure robust comparisons across species, we then ranked the different datasets according to their differential gene expression (DGE) similarity with our mouse naïve and primed dataset. For this, we selected the top 1200 differentially expressed genes between our 2iL ESCs and EpiSCs samples and studied if there was a gene expression change in the same direction of the orthologous genes in naïve and primed human samples for each independent experiment (Figure 6A). We selected the three top datasets with more satisfactory results (over 70% of differentially expressed genes changed in the same direction as in our naïve vs. primed mouse samples). These samples came from the following studies: Collier et al 2017[414], Takashima et al 2014[146] and Guo et al 2017[152].

To investigate whether the genomic features associated with 2iL-EpiSC differential AS exons are conserved in the naïve-to-primed transition in humans, we scored intron length, exon length, intron GC content, exon GC content and exon-intron definition features for differentially spliced exons in the three selected datasets (Figure 6B). Remarkably, the observed patterns were exceptionally well conserved in all three human naïve-primed comparisons. Indeed, flanking introns are significantly longer in alternative exons more

---

quantiles for each genomic feature, separated by cell line. We calculated the "Exon-intron definition score" considering the combined effect of exon GC content and the difference in intron and exon lengths (see methods for further details and complete definition). **(H)** Boxplot of exon-intron definition scores across the different exon categories based on their "dynamics" across pluripotent samples (as in Figure 4C).

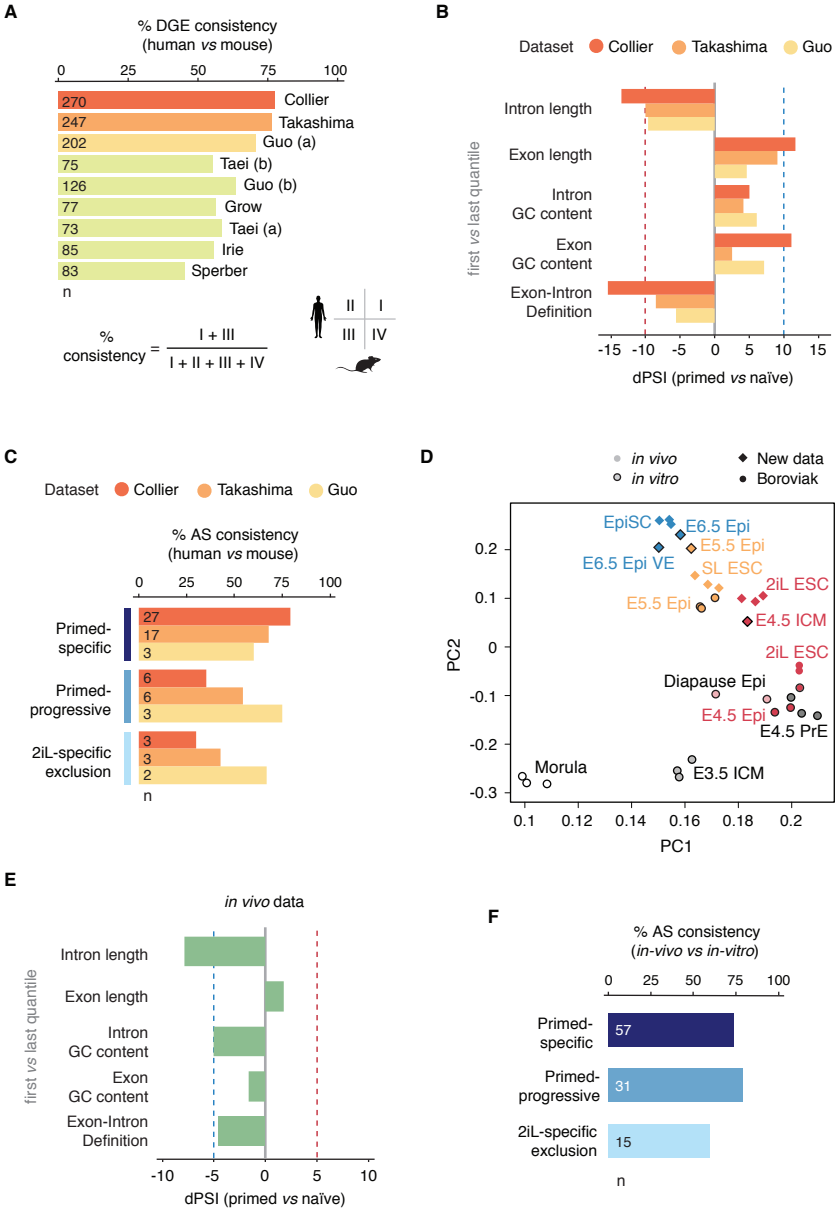


Figure 6: Conservation of naïve-to-primed alternative splicing profile in human pluripotency and overlap with *in-vivo* pluripotency (A) Bar plot depicting the degree of mouse naïve-to-primed differential gene expression (DGE) conservation across human naïve and primed stem cell comparisons coming from different

included in human naïve samples and these exons were also shorter and had a lower GC content than exons more included in human primed samples. This indicates that the association between exon-vs-intron definition and lower inclusion levels in primed pluripotency could potentially be a common feature for the naïve-to-primed transition at least since the last common ancestor of mice and humans.

Beyond the mode of splicing regulation, one key remaining question was whether the orthologue exons in human were regulated in the same way as their mouse counterparts in naïve vs primed samples. Overall, 67% of regulated exons in the naïve-to-primed transition in mouse have a human ortholog (ranging from 62.3% in Primed-specific to 75% in progressive exons). Interestingly, for the orthologous exons that were differentially spliced between naïve and primed in the two species, we found that regulatory consistency was different across the six defined groups based on their splicing dynamics (Figure 6C). On average, primed or naïve-specific exons were more consistent than progressive and 2iL-specific events (69%, 55% and 47% average consistency across studies, respectively), with the experiment from Collier et al showing the highest rate of conservation for primed-specific exons.

---

studies (indicated as the first author name). Numbers at the base of the bars indicate the number of consistent genes in each comparison: genes changing expression with the same directionality in human and mouse, corresponding to genes laying on the I and III quadrants, divided by the total number of DGE genes. **(B)** Bar plot of splicing regulatory feature scores for human naïve-primed dPSI comparisons across three different studies (as in Figure 5G, see Methods for further details and complete definition of each score). **(C)** Bar plot illustrating percentage of differential naïve-to-primed cassette exon alternative splicing consistency in human and mouse pluripotency shows higher consistency among primed-specific and progressive events. The number at the base of each bar represents the number of orthologue alternative exons identified. Percentage of consistency was calculated as depicted on (A). **(D)** Principal component analysis of *in-vitro* and *in-vivo* naïve and primed samples and other stages of early mouse development using gene expression data for the top 5000 most variable genes. VE: Visceral endoderm. Epi: Epiblast. PrE: Primitive endoderm. ICM: Inner cell mass. **(E)** Bar plot of splicing regulatory feature scores for *in-vivo* naïve-primed PSI comparisons shows intron length is significantly higher in exons differentially included in E4.5 epiblast vs. E6.5 epiblast. **(F)** Bar plot displaying percentage of consistent changes in naïve-to-primed cassette exon alternative splicing across *in-vivo* and *in-vitro* samples. The number at the base of each bar represents the number differential alternative exons identified.

Transcriptomic profiling of *in vivo* naïve-to-primed alternative splicing variants during mouse early embryo development

We set out to find whether the naïve-to-primed alternative splicing switch that we observed *in vitro*, also occurs *in vivo*. To rigorously interrogate exclusively epiblast lineage identity and avoid transcriptional signals coming from other lineages, we performed pure ICM isolation from E4.5 mouse blastocysts as well as pure epiblast isolation (discarding visceral endoderm and extraembryonic ectoderm) from E5.5 and E6.5 embryos (see Figure 1 in Methods section). We then performed deep coverage RNA sequencing to identify *in vivo* naïve-to-primed alternative splicing variants. Principal component analysis of differential gene expression across *in vivo* and *in vitro* samples collected by us and Boroviak et al[248], clustered samples according to their pluripotency state (Figure 6D). This agreement at the level of gene expression led us to ask whether the alternative splicing profiles that we had observed in naïve and primed pluripotency *in vitro* were also consistent with the ones occurring *in vivo*.

We first identified alternative exons that were differentially spliced between naive and primed conditions *in vivo*. For this, we compared E4.5 and E6.5, finding 1,789 exons with a dPSI of at least 15%. Of all alternatively spliced cassette exon switches changing between E4.5 and E6.5 epiblast that are also changing between 2iL ESCs and EpiSCs, 73% are in the same direction of change ($p= 4.1 \times 10^{-8}$, Binomial test). These changes were more consistent for the primed and naïve-specific and progressive groups, similar to the results obtained when comparing mouse vs human AS exons in this transition (Figure 6E). With respect to the characterized splicing regulatory features, and in line with what we found *in vitro*, intron length was significantly higher in exons that were differentially more included in the inner cell mass (Figure 6F). However, we did not find any other significant enrichment in the rest of the splicing-associated genomic features we looked at. This suggests that alternative splicing regulation of pluripotent cells might occur at different levels *in vivo* and *in vitro*, while still maintaining a certain degree of similarity and consistency, potentially linked to regulation driven by the action of state-specific RNA binding proteins.

Chapter 2: Alternative splicing regulation by RNA Binding Proteins in naïve and primed pluripotency

We set out to find the regulators that were shaping the alternative splicing profiles of naïve and primed pluripotency. For this, we first looked at how the expression levels of known alternative splicing factors, spliceosomal components and splicing-associated factors change during the naïve-to-primed transition (Figure 7A) and validated the change in protein levels of some of these factors via western blot (Figure 7B). We found no bias in the expression of core spliceosomal components between naive and primed pluripotent cells. It was thought-provoking to see that among the most changing factors, some regulator pairs with allegedly redundant or complementary roles were positioned on opposing states of pluripotency. This was the case for two regulatory axes: Rbmx/RbmxL2 and Rbm47/Esrp1.

In the case of RBMX (RNA Binding Motif protein, X-linked) and RBMXL2, the latter is a eutherian-specific retrocopy of *Rbmx* with over 83% amino acid similarity. Both have been involved in the regulation of alternative splicing [417–424] but only three studies to date have tested both factors in parallel to check their relative effect and only on a few specific splicing events[417,423418]. Two of these studies, both using the minigene system on a particular exon, showed that RBMX and RBMXL2 impact exon inclusion in the same direction. Particularly, they found that both RBMX and RBMXL2 promote inclusion of an alternative exon of the human transducing-like enhancer of split 4 (*TLE4*) transcript, albeit RBMXL2 did so more efficiently[417]. Additionally, both RBMX and RBMXL2 promote the exclusion of a cryptic exon in the mouse lysine-specific demethylase 4d (*Kdm4d)* transcript with similar efficiency[423]. Another study found that overexpression of *Rbmx* but not of *RbmxL2* induced exclusion of an exon in the thioredoxin-like 4A (*TXNL4a*) transcript and inclusion of exons in nucleolar Protein 5A (*NOL5A*) and in Tyrosine-Protein Kinase Fyn (*FYN*) transcripts[418]. No study has ever reported RBMX and RBMXL2 having opposing effects on splicing. However, we hypothesized that naïve-to-primed transition may have laid the stage for retrocopy subfunctionalization in the case of *Rbmx* and *RbmxL2* and thus that they may have opposing regulatory features in the context of pluripotency.

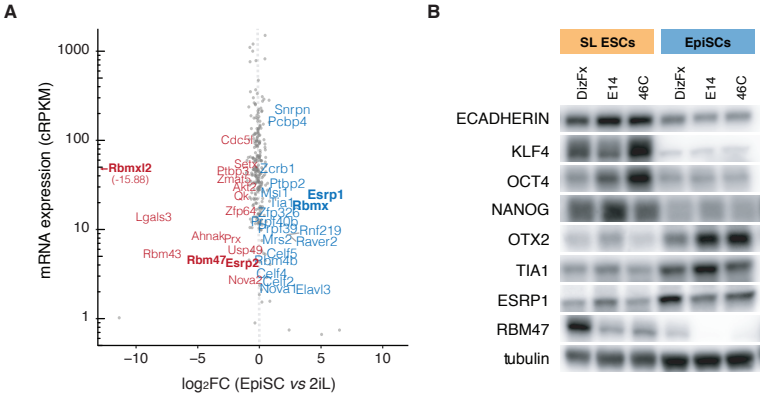


Figure 7: Expression level changes in splicing factors, splicing-associated factors and spliceosome components in naïve and primed pluripotency. (A) Scatterplot depicting the fold-change difference in expression between 2iL ESCs and EpiSCs for genes under the "splicing" gene ontology (GO) category. Only genes with a false discovery rate value below 0.01 are labelled. Expression levels in cRPKMs correspond to the state where each gene is more expressed: 2iL for red-labelled genes and EpiSC for blue ones. **(B)** Western blot in SL ESCs and EpiSCs of three independent cell lines showing changes in proteins levels in general pluripotency markers (Oct4), naïve markers (eCadherin, Klf4, Nanog), primed markers (Otx2) and identified splicing factors (Tia1, Esrp1 and Rbm47). Representative image of experiment repeated in triplicate.

RBM47 and ESRP1 were also a fascinating case primarily because both are key regulators of the EMT splicing program with known coregulated events [425] and with known enhancing effect on induced pluripotency reprogramming [389]. Notably, although RBM47 and ESRP1 regulate many alternative cassette exon events in the same direction of inclusion/exclusion and thus, are considered to have a functional combinatorial co-regulation promoting epithelial traits, some other events seem to be regulated in opposing directions[425]. However, to date, no study has explored this further than speculating a more complex pattern of interactions among these two factors during EMT. Given their expression levels in the different states of pluripotency, we hypothesized that their contrasting roles might acquire protagonism in naïve and primed pluripotency.

Moreover, we sought to identify additional factors that could be playing a role in regulating the alternative splicing profile of naïve and primed pluripotency. We wanted to go a step further from our differential expression-based

approach to find factors that were not notably changing at the level of gene expression or that were simply not known splicing regulators to begin with. Hence, we carried out a quantitative functional genomics screen that allowed us to interrogate the role of hundreds of transacting factors on the inclusion levels of differentially spliced naïve-to-primed alternative splicing events.

Part I and Part II of this chapter will centre on the candidate-based approach that we followed. We will describe the results obtained on the Rbmx/RbmxL2 axis and Rbm47/Esrp1 axis, respectively. Subsequently, Part III will aim to describe the screening, the research direction that emerged from it, and the results we obtained.

Part I: Rbmx/RbmxL2 regulatory axis

We found that there is a complete switch in *Rbmx* and *RbmxL2* expression levels in the naïve-to-primed pluripotency transition. While *Rbmx* expression levels in naïve pluripotency are low, they rise in primed pluripotency (Figure 8A). The contrary occurs for *RbmxL2* whose expression drops to zero in primed pluripotency (Figure 8A) and was consistently by far the RNA binding protein (RBP) with most changing expression levels in the naïve vs. primed state that we detected (Figure 7A). Additionally, *Rbmx* has another known retrocopy in mice named *RbmxL1*. The expression levels of *RbmxL1* remained constant across naïve and primed pluripotency.

Interestingly, the sequence similarity among these proteins is remarkable- in comparison to RBMX, RBMXL2 has 83% similarity (75% identity) and RBMXL1 has 97% similarity (96% identity). All three have an RNA recognition motif (RRM) with two conserved RNP motifs (RNP1 and RNP2) at the N-terminus and a C-terminus highly enriched in disordered regions (Figure 8B). This C-terminus has a very high content of serine, arginine, glycine and proline in its amino acid composition. Specifically, 68%, 68% and 65% in RBMX, RBMXL1 and RBMXL2 respectively. Additionally, RBMX has two RGG boxes on each side of this region each containing three RGG repeats. Curiously, RBMXL1 lacks one of these RGG repeats and RBMXL2 lacks two, one at each side (Figure 8B).

Rbmx itself has an alternative splicing event occurring in the last coding exon, which precisely codes for the region containing the last RGG box (Figure 8B). The skipping of this exon predictably results in a stop codon at the start of the C2 exon that could predictably result in a truncated protein lacking the last C-terminal 102 amino acids. This exon is alternatively spliced in the naïve-to-primed transition. In the naïve state, it has an intermediate inclusion whereas, in the primed state, it remains highly included in the majority of transcripts (Figure 8C). Therefore, naïve pluripotent stem cells not only have lower *Rbmx* expression, but also around half of its transcripts produce a shorter version of the protein if translated. We found that the human ortholog of this cassette exon is also alternatively spliced in naïve and primed human pluripotency (Figure 8D).

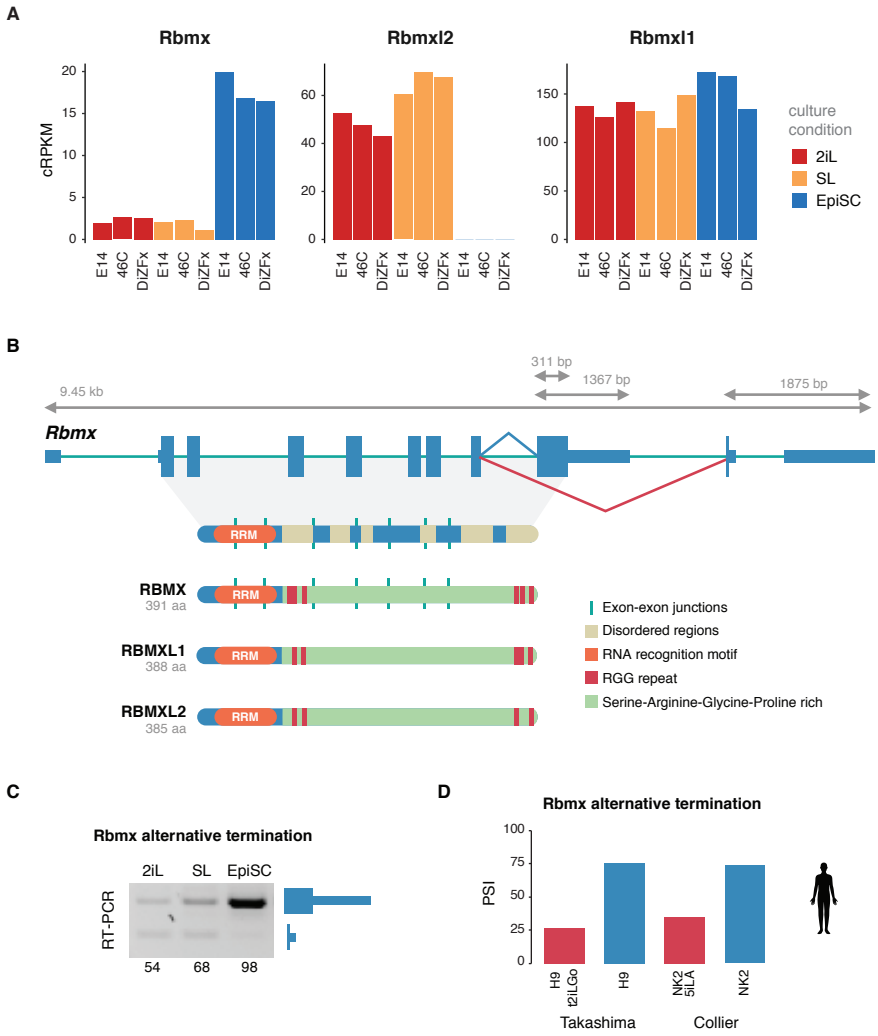


Figure 8: Expression level and protein features of Rbmx and its retrocopies RbmxL2 and RbmxL1 in naïve and primed pluripotency. **(A)** Expression levels of Rbmx, RbmxL2 and RbmxL1 across three independent cell lines in naïve and primed cell culture conditions. **(B)** Genomic landscape of the *Rbmx* gene and protein features of RBMX, RBMXL1 and RBMXL2. The alternative splicing event of the last exon of Rbmx is depicted. In the genomic landscape, coding sequences are illustrated with thick boxes, untranslated sequences with thin boxes and introns with lines. Angles in blue join the main exon-exon junction used in EpiSCs, in red, the alternative exon-exon junction found in naïve ESCs. RGG repeat: Arginine-Glycine-Glycine repeat. **(C)** RT-PCR validation of inclusion levels of the alternative splicing event in the last coding exon of Rbmx showing differential inclusion in naïve and primed pluripotency.

Altering the Rbmx/RbmxL2 regulatory axis in naïve pluripotency

To understand the functional roles of *Rbmx*, *RbmxL2* and *RbmxL1* on pluripotent cell identity, we made a set of full gene knockout (KO) ESC lines using the CRISPR/Cas9 gene-editing system on E14 ESCs. In all cases when making ESC KO lines, we generated control ESC lines transfected in parallel with an empty vector (named px, from px459/330). Specifically, we first generated an RbmxL2 KO ESC line by deleting ~1000bp of the *RbmxL2* genomic sequence. This deletion was directed by the use of two guide RNAs (gRNAs) targeting two distant sites in the coding sequence of the gene (Figure 9A). The strategy allowed us to generate three full RbmxL2 KO clones. One had an extremely aberrant karyotype with most of the cells having a high chromosome number (33% had 76-82 chromosomes), thus we discarded that clone for further analysis. The other two remaining clones had normal karyotypes (see material and methods) but one, clone 2.8, had an extreme propensity for differentiation, with most of the cells showing a differentiated morphology. This clone had a low number of ESC colonies and overall very low AP staining after few days in SL culture (Supplemental Figure 1).

Given the high protein sequence similarity across RBMX, RBMXL1 and RBMXL2, it is possible that some regulatory effects of their depletion may be masked by the other retrocopies undertaking functionalities that normally are carried out by the depleted factor. Thus, using gRNAs that simultaneously target both Rbmx and RbmxL1 coding sequences, we first generated Rbmx RbmxL1 DKO cell lines (Figure 9A). Next, we used the RbmxL2 clone with the milder phenotype, clone 2.5, as background to generate double and triple KOs. Interestingly, while we obtained multiple Rbmx/RbmxL2 KO clones, after an intense screening of clones we were not able to obtain a single Rbmx RbmxL1 RbmxL2 triple KO clone, suggesting the triple KO may be lethal. Finally, to understand the role of the differentially included last coding exon of Rbmx, we generated a cell line that lacked that last exon. This line, from now on referred to as Rbmx Exon KO, was generated using the CRISPR/Cas9 gene-editing system targeting the introns flanking the exon.

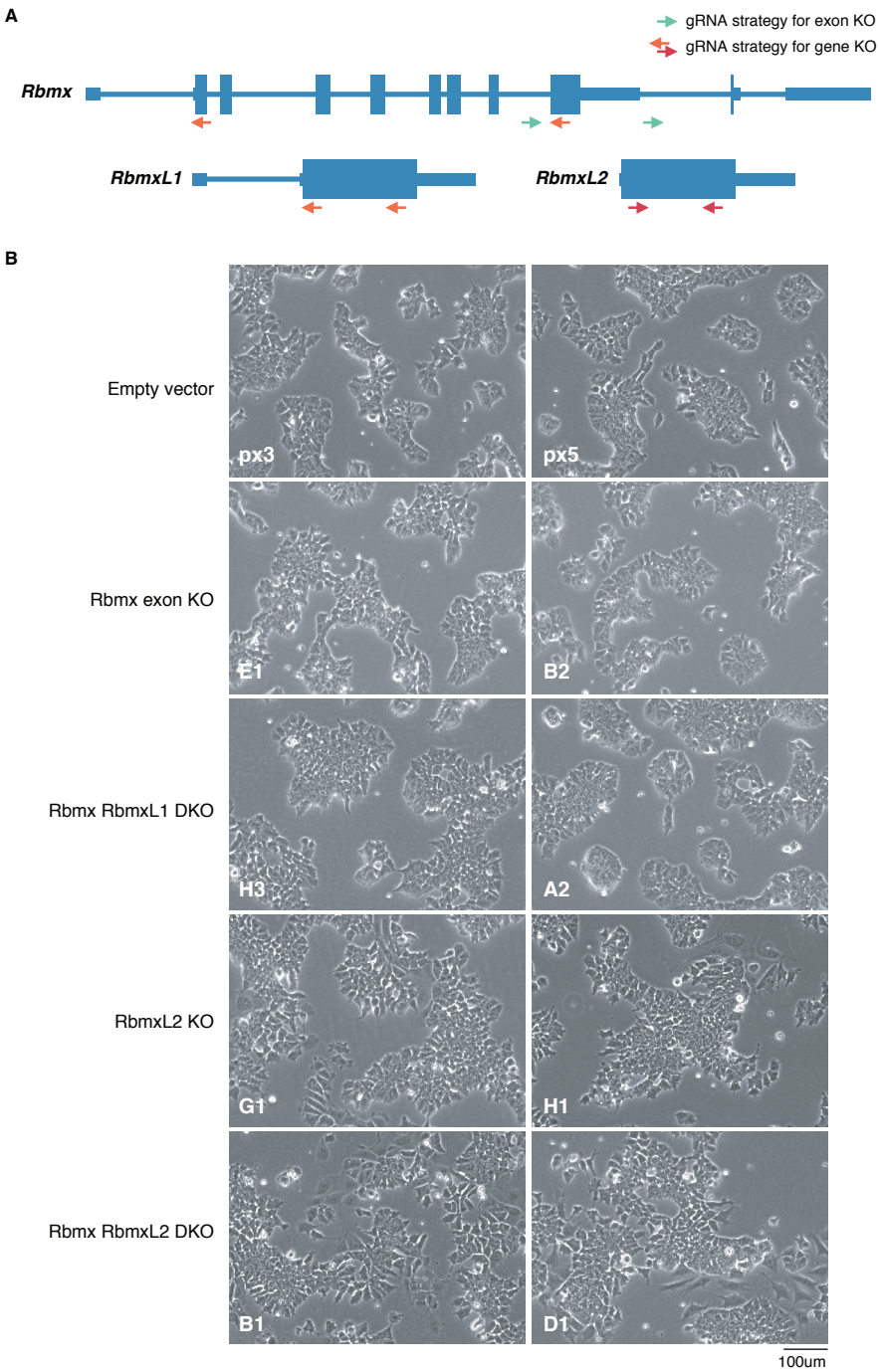
Numbers at the bottom are PSI levels based on direct gel quantifications. **(D)** Bar plot illustrating the inclusion of the conserved Rbmx alternative splicing switch in the last coding exon from human naïve and primed samples shows conservation of the pattern of regulation of this event. PSI values for naïve (red) and primed (blue) human stem cells were calculated using RNA-seq from Takashima et al. 2014[146] and Collier et al. 2017[414].

For all of the above lines, various clones were obtained. All of them were karyotyped and expanded for subsequent experiments (see Materials and methods for further details). The following results were generated using two clones for each genotype to ensure solid reproducible conclusions.

We expanded our KO cell lines in SL conditions. As discussed before, this condition maintains ESCs in the naïve state but is permissive for primed pluripotency and occasional differentiation which precisely allows us to observe whether any of our KOs has an evident bias towards differentiation. Indeed, one of the most obvious phenotypes that we observed in our cell lines was that clones that had depleted *RbmxL2* showed a higher number of cells with a differentiated morphology (Figure 9B; as mentioned before, this was particularly extreme in the RbmxL2KO clone 2.8, which could not be maintained in SL culture and was not used for further analyses). Another striking observation was that one of the Rbmx RbmxL1 DKO clones had rounder and more dome-shaped colonies than the rest of the clones. The remaining clones did not show any discernible morphological phenotype.

We then wondered whether there was any compensation of expression levels by the retrocopies when one or the other were depleted. Certainly, we observed that upon *RbmxL2* depletion, *Rbmx* expression levels raised (Figure 10A). Naturally, this could be due to the differentiated phenotype that RbmxL2 KO ESCs showed and therefore might not compose a direct compensation of *RbmxL2* levels. However, we also detected higher levels of *RbmxL2* in Rbmx RbmxL1 DKOs (Figure 10A). Considering that SL

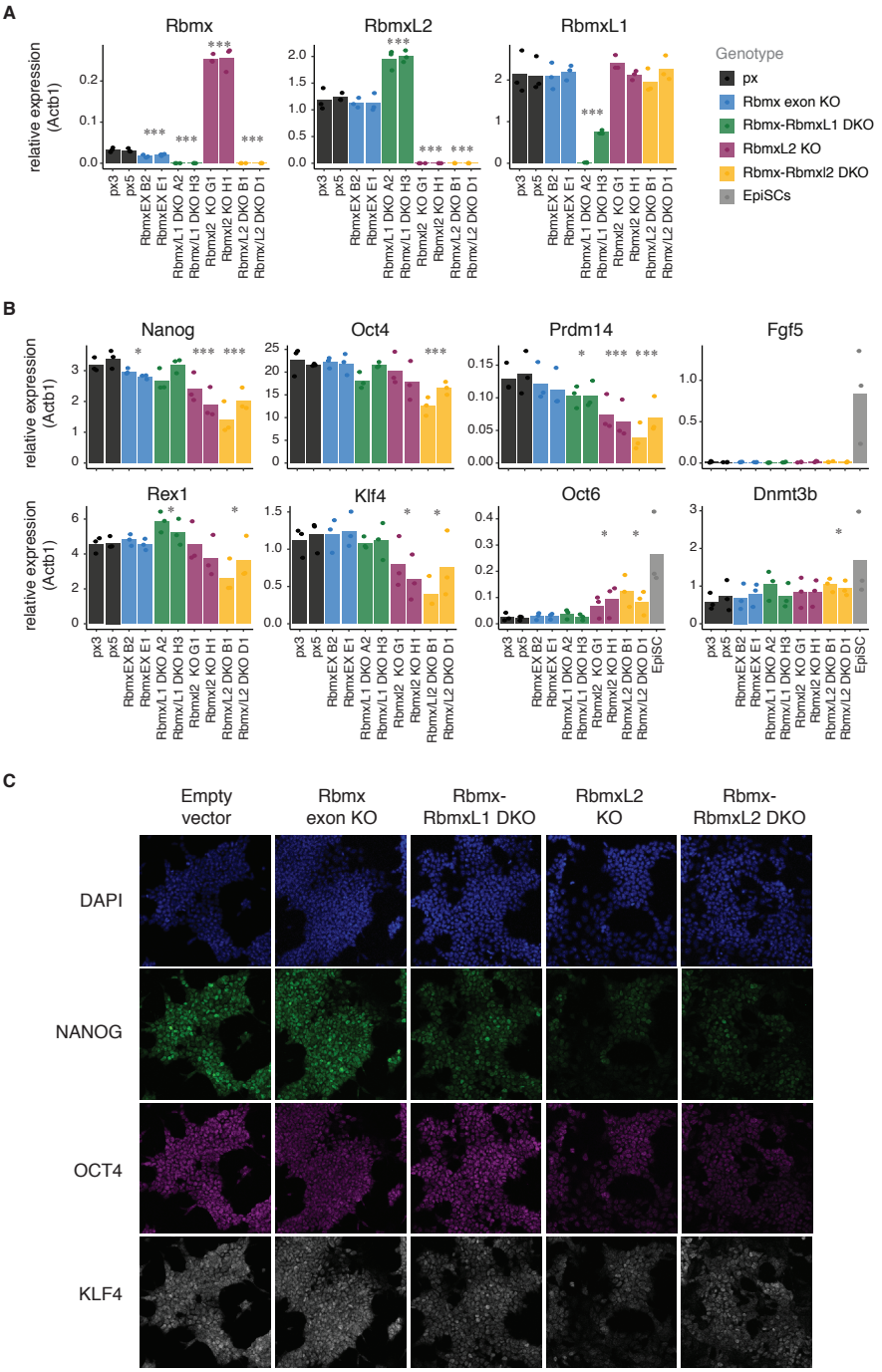
Figure 9: Generation of ESC lines with Rbmx, RbmxL1 and/or RbmxL2 CRISPR/Cas9 gene and alternative exon depletions. (A) CRISPR/Cas9 strategy for depletion of *Rbmx*, *RbmxL2* and *RbmxL1* genes and *Rbmx* alternatively spliced exon. Red arrows illustrate the gRNAs that were used to delete ~1000bp from *RbmxL2* coding sequence between them. Green arrows illustrate the gRNAs that were used to delete the alternatively spliced last coding exon from *Rbmx*. These gRNAs targeted intronic sequences flanking the exon of interest and generated deletions of ~1600bp between them. Orange arrows illustrate the gRNAs that were used to delete both *Rbmx* and *RbmxL1*. These were two promiscuous guides targeting *Rbmx,* generating deletions of ~4300bp between them, but with only 1 or 2 mismatches with the *RbmxL1* sequence, where they generated deletions of ~900bp. **(B)** Bright-field images of KO cell lines grown in SL conditions show discernible aberrant morphology in all RbmxL2 KO cell lines and dome-shape morphology of the colonies from Rbmx RbmxL1 DKO A2 clone.


A
gRNA strategy for exon KO
gRNA strategy for gene KO
Rbmx
RbmxL1
RbmxL2
B
Empty vector
Rbmx exon KO
Rbmx RbmxL1 DKO
RbmxL2 KO
Rbmx RbmxL2 DKO
px3
px5
E1
B2
H3
A2
G1
H1
B1
D1
100um

conditions maintain a heterogeneous population of pluripotency states, these results could be pointing us that Rbmx RbmxL1 DKOs would have a more prevalent naïve population of ESCs as this would translate into a higher level of *RbmxL2* expression compared to the rest. Nonetheless, 2iL ESCs are a more homogenously naïve cell population and they maintain the same expression levels of *RbmxL2*, actually, even lower (Figure 8A). Therefore, in this case, this might be suggesting a clear compensation among retrocopies (see discussion). Finally, the deletion of the terminal alternative *Rbmx* exon downregulated the expression levels of *Rbmx* but did not seem to affect the expression levels of *RbmxL2* or *RbmxL1* (Figure 10A). In fact, *RbmxL1* levels did not seem to markedly change throughout the various KOs.

To understand the effect that the depletion of *Rbmx*, *RbmxL1* or *RbmxL2* had on naïve pluripotency, we monitored the expression levels of naïve pluripotency markers. Compared to controls, Rbmx RbmxL2 DKOs had significantly lower levels of all naïve markers we studied (Figure 10B). RbmxL2 KOs also showed lower levels of most of the naïve markers we studied, but unlike Rbmx RbmxL2 DKOs, did not show lower expression levels of *Oct4*, a general pluripotency marker. This seemed to suggest that the depletion of *RbmxL2* might affect the pluripotency state and the simultaneous depletion of *Rbmx* might exacerbate this effect. On the other hand, both Rbmx exon KOs and Rbmx RbmxL1 DKOs did not seem to have substantial differences in the expression of most of the naïve markers compared to controls. We therefore wondered whether part of the ESCs population of these KO clones grown in SL had exited naïve pluripotency

Figure 10: Naïve pluripotency markers upon depletion of Rbmx/RbmxL2 regulatory axis in ESCs. (A) Bar plots of *Rbmx*, *RbmxL1* and *RbmxL2* relative gene expression levels measured by quantitative PCR across all KO lines show high levels of *Rbmx* in RbmxL2KOs and high levels of *RbmxL2* in Rbmx RbmxL1 DKOs. **(B)** Bar plots depicting relative gene expression levels of general pluripotency (*Oct4*), naïve pluripotency (*Prdm14*, *Klf4*, *Rex1*, *Nanog*), formative pluripotency (*Oct6* and *Dnmt3b*) and primed pluripotency (*Fgf5*) markers measured by quantitative PCR in all KO lines show lower levels of naïve pluripotency genes and a modest increment of *Oct6* in RbmxL2 KOs and Rbmx RbmxL2 DKOs. P-values from two-sided Student's t-tests joining together the two clones per genotype and comparing them with controls are indicated: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$. **(C)** Microscopy images of KO clones immunostained for naïve pluripotency markers shows higher heterogeneity in their expression among Rbmx RbmxL1 DKO colonies and lower heterogeneity in their expression among RbmxL2 KO and Rbmx RbmxL2 DKO colonies.


A
Rbmx
RbmxL2
RbmxL1
relative expression (Actb1)
Genotype
px
Rbmx exon KO
Rbmx-RbmxL1 DKO
RbmxL2 KO
Rbmx-Rbmxl2 DKO
EpiSCs
px3
px5
RbmxEX B2
RbmxEX E1
Rbmx/L1 DKO A2
Rbmx/L1 DKO H3
Rbmxl2 KO G1
Rbmxl2 KO H1
Rbmx/L2 DKO B1
Rbmx/L2 DKO D1
B
Nanog
Oct4
Prdm14
Fgf5
Rex1
Klf4
Oct6
Dnmt3b
EpiSC
C
Empty vector
Rbmx exon KO
Rbmx-RbmxL1 DKO
RbmxL2 KO
Rbmx-RbmxL2 DKO
DAPI
NANOG
OCT4
KLF4

and was primed pluripotent instead. We monitored the expression levels of *Oct6* and *Dnmt3b*, two formative and primed state markers; and *Fgf5*, a primed state marker in all our KOs and found that RbmxL2 KOs and Rbmx RbmxL2 DKOs had a modest increase of *Oct 6* but no noticeable increase in *Fgf5* (Figure 10B). Moreover, Rbmx RbmxL2 DKOs also showed a modest increase in *Dnmt3b* expression, again pointing at a more exacerbated phenotype than RbmxL2 KOs. As SL ESC colonies are heterogeneous, the modest increase in formative state marker levels seen upon *RbmxL2* depletion could mean that a small population of cells were, in fact, exiting naïve pluripotency but not yet primed. However, the considerable drop in naïve pluripotency markers upon RbmxL2 depletion could not solely be explained by the modest increase in formative pluripotency and probably points towards a proportion of the population being differentiated cells.

The heterogeneity level of a population of cells can only be assessed by looking at the single-cell level, thus we performed immunostainings for naïve pluripotency markers across all our KO clones. Interestingly, Rbmx RbmxL1 DKOs showed higher heterogeneity of NANOG and KLF4 levels than controls, with cells showing a wide spectrum of levels among the population (Figure 10C). Additionally, we found that most cells of RbmxL2 KOs and Rbmx RbmxL2 DKOs had generally lower levels of naïve markers, and thus were more homogenous in that respect. Sadly, we were not able to successfully perform immunostainings using a primed pluripotency marker to identify the potential primed colonies in RbmxL2 KOs and Rbmx RbmxL2 DKO clones.

To elucidate the general pluripotency state and the extent of the depletion effect of *Rbmx*, *RbmxL1* and *RbmxL2*, we performed RNA-seq in SL KO ESCs. Again, we chose to perform the experiment in SL conditions as it presents a permissive setting that is ideal to detect a bias towards a given cell state. Using Gene Set Enrichment Analysis (GSEA)[426], we found that RbmxL2 KO and Rbmx RbmxL2 DKOs had a very strong bias towards a primed pluripotency gene expression profile and downregulated the naïve pluripotency gene expression profile (Figure 11 A, B). Oppositely, Rbmx exon KOs were more enriched in the differential gene expression profile of naïve pluripotency and showed significantly lower primed pluripotency gene expression profile compared to controls (Figure 11A, B). This suggested a bias in these cells towards a more naïve state of pluripotency. Finally, we found that Rbmx RbmxL1 DKOs downregulated both naïve and primed gene expression (Figure 11A, B). This suggested the induction of a different cell identity and probably reflected

the cell population heterogeneity that we had observed by immunostaining (Figure 10C).

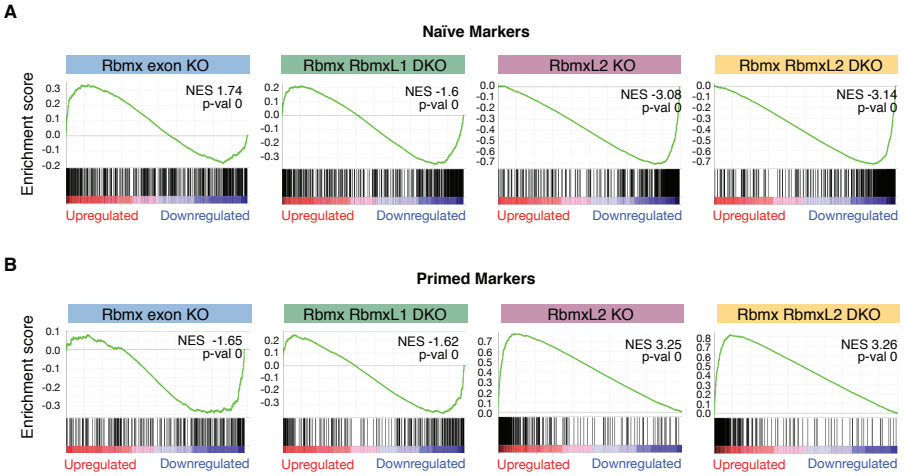


Figure 11: Gene set enrichment analysis shows RbmxL2 KOs and Rbmx RbmxL2 DKOs upregulate primed pluripotency genes whereas Rbmx exon KOs upregulate naïve pluripotency. GSEA depicting enrichment in naïve (A) or primed (B) gene expression profiles for the differentially expressed genes across *Rbmx* alternative exon, *Rbmx*, *RbmxL1* and *RbmxL2* depleted lines. Ranked lists of naïve and primed markers were defined by establishing the differential gene expression profile across paired comparisons of 2iL vs. EpiSCs in our three independent cell lines (E14, 46C, DizFx) and E4.5 vs. E6.5 epiblasts (see methods for further details). RbmxL2 KOs and Rbmx RbmxL2 DKOs downregulate naïve genes and upregulate primed ones; Rbmx exon KOs significantly upregulate naïve pluripotency genes and downregulate primed ones; and Rbmx RbmxL1 DKOs show significant downregulation of both naïve and primed pluripotency genes. NES: Normalized enrichment score.

We then set out to find whether the effect of *Rbmx*, *RbmxL1* and *RbmxL2* depletion on gene expression of pluripotency state profiles reflected their effect on the alternative splicing landscape. More importantly, this could potentially hint at whether alternative splicing landscapes in pluripotency were consonant with the gene expression profiles driven by pluripotency state. Hierarchical clustering of all samples by differential alternative splicing profiles grouped samples according to their genotype (Figure 12A). Interestingly, there was a

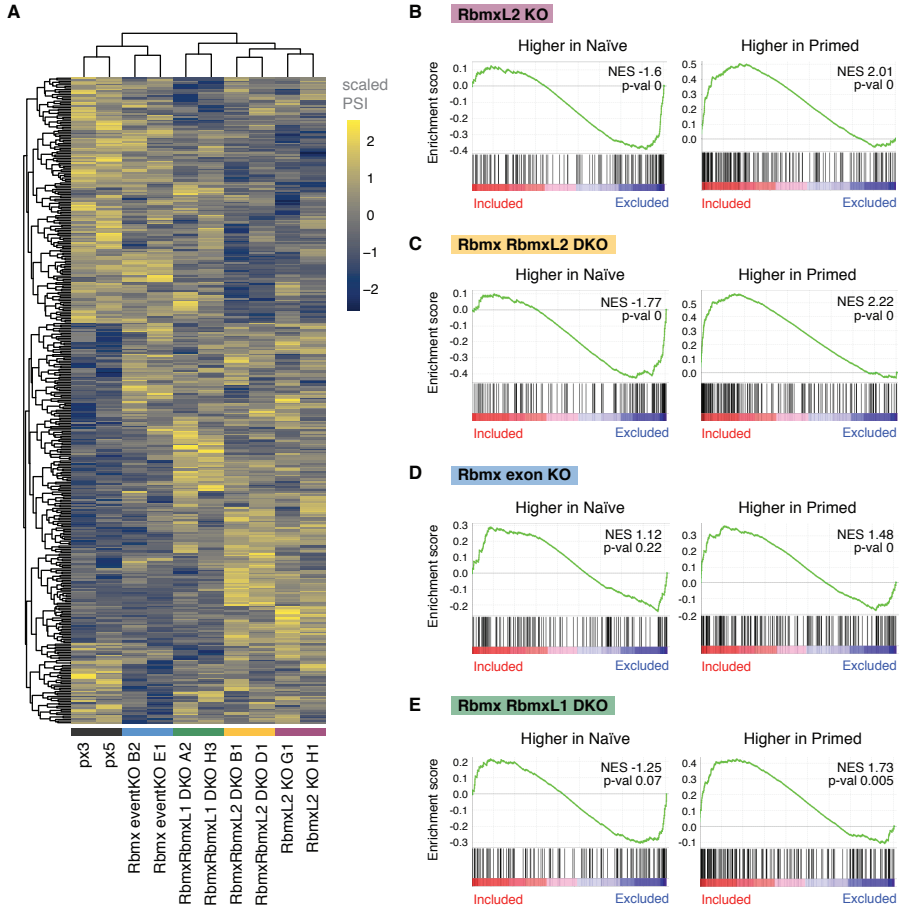


Figure 12: Differential alternative splicing profiles upon depletion of Rbmx/ RbmxL2 regulatory axis in ESCs. (A) Heatmap depicting scaled PSI levels of differential AS events (cassette exon, intron retention, alternative 5'ss or alternative 3'ss events) occurring across *Rbmx* alternative exon, *Rbmx*, *RbmxL1* and *RbmxL2* depleted lines, depicting unsupervised clustering of samples in the top dendrogram. **(B-E)** Alternative Splicing Set Enrichment Analysis (ASSEA) using our previously identified differentially spliced exons between naïve and primed pluripotency. (B-C) RbmxL2 KOs and Rbmx RbmxL2 DKOs show enrichment for the exclusion of exons with higher inclusion in naïve pluripotency (higher in naïve) and for the inclusion of exons with higher inclusion in primed pluripotency (higher in primed). (D) Rbmx exon KOs show enrichment for the inclusion of exons more highly included in primed pluripotency. (E) Rbmx RbmxL1 DKOs show enrichment in inclusion of exons more highly included in primed pluripotency. For ASSEA, ranked lists of differentially

cluster of all cell lines with depleted *RbmxL2*. This cluster in turn also clustered with Rbmx RbmxL1 DKOs, indicative of having a more closely related alternative splicing profile with them than with controls. Additionally, Rbmx exon KOs clustered with controls, separately from gene KOs, suggestive of a potentially milder and contrasting effect on alternative splicing.

To impartially evaluate the transcriptome-wide changes in alternative splicing among the different splicing factor depletions, we designed a novel method which we termed Alternative Splicing Set Enrichment Analysis (ASSEA). This method determines whether a defined set of alternative splicing events holds statistically significant inclusion differences between two biological states. In this case, we defined the set of alternative splicing events using naïve and primed pluripotency differentially included exons, more specifically using the 2iL ESCs vs. EpiSCs comparison (see materials and methods). With this, we were able to test whether the effect of each KO on naïve-to-primed differential AS exons was significant, and more importantly, whether the isoform landscape of the ESCs would be more similar to the one found in 2iL ESCs or EpiSCs.

Compared to controls, RbmxL2 KOs and Rbmx RbmxL2 DKOs showed an enrichment in inclusion of exons that are more highly included in primed pluripotency and an enrichment in exclusion of exons that are more highly included in naïve pluripotency (Figure 12B, C). This suggested a clear profile of isoform change towards primed pluripotency, albeit with a greater enrichment score in the Rbmx RbmxL2 DKOs compared to RbmxL2 single KOs. This hinted at the possibility that a set of naïve-to-primed exons might be co-regulated by both RbmxL2 and Rbmx and thus depleting the two factors has a greater effect. However, it must be noted that the increase in enrichment is minimal and thus is probably due to very few events. Additionally, we found that Rbmx exon KOs were significantly enriched in inclusion of alternatively spliced exons that were also more included in EpiSCs (Figure 12D). This came as a surprise given that, as we discussed earlier, in terms of differential gene expression Rbmx exon KOs were closer to the naïve state (Figure 11A, B). This might be a case where the regulation of alternative splicing and gene expression seem to be independent from one another. Rbmx RbmxL1 DKOs also were

---

included exons were defined using paired comparisons of 2iL vs. EpiSCs across our three independent cell lines (E14, 46C, DizFx; see methods for further details). NES: Normalized enrichment score.

significantly enriched in the inclusion of alternatively spliced exons that were also more included in EpiSCs (Figure 12E), which in this case, did coincide with the trend observed in gene expression where they were more primed.

Impact of Rbmx/RbmxL2 regulatory axis on cell differentiation

As discussed in the introduction, LIF is indispensable to maintain the pluripotent state of ESCs in serum-based media. LIF signalling pathway promotes self-renewal and inhibits differentiation but when it is removed from the culture media that inhibition is cleared away and differentiation starts. Thus, to explore the influence of the Rbmx/RbmxL2 regulatory axis on cell differentiation, we performed simple LIF-removal differentiation assays testing *Rbmx* alternative exon, *Rbmx*, *RbmxL1* and *RbmxL2* depleted cell lines.

Firstly, we performed clonal monolayer differentiation assays. These consisted on plating KO ESC lines and controls in SL media and at clonal density and later removing LIF for two or four days. After this time, alkaline phosphatase (AP) staining was performed to detect the proportion of pluripotent colonies that remained (see material and methods). Importantly after four days of LIF removal, cell lines that have a limited capacity for differentiation will maintain a high proportion of pluripotent colonies which are highly AP stained, whereas lines with a tendency to differentiate will have a high proportion of differentiated colonies and thus low AP stained colonies (Figure 13). We found that RbmxL2 KOs and Rbmx RbmxL2 DKOs had a very strong tendency to differentiate, with most of the colonies showing low AP staining by day four of differentiation (Figure 13). This goes in line with what we had observed before and suggested that depletion of *RbmxL2* primes ESCs for differentiation. Additionally, we found that one of the Rbmx RbmxL1 DKO clones, clone A2, had a delayed exit from pluripotency as many highly AP stained colonies still remained by day four of LIF removal. This was the same clone that we reported to have an unusual morphology, with rounder dome-shaped colonies compared to the other Rbmx RbmxL1 DKO clone and with controls (Figure 9B). These dome-shaped colonies are typical of 2iL ESCs and could be indicating a more naïve state, however, we did not detect an increase in most naïve pluripotent markers for this specific clone (Figure 10B).

To better pinpoint the dynamics of exit of naïve pluripotency, entry and exit of primed pluripotency as well as to identify any biases towards specific lineages in our KO lines, we performed embryonic body (EB) assays. We took samples

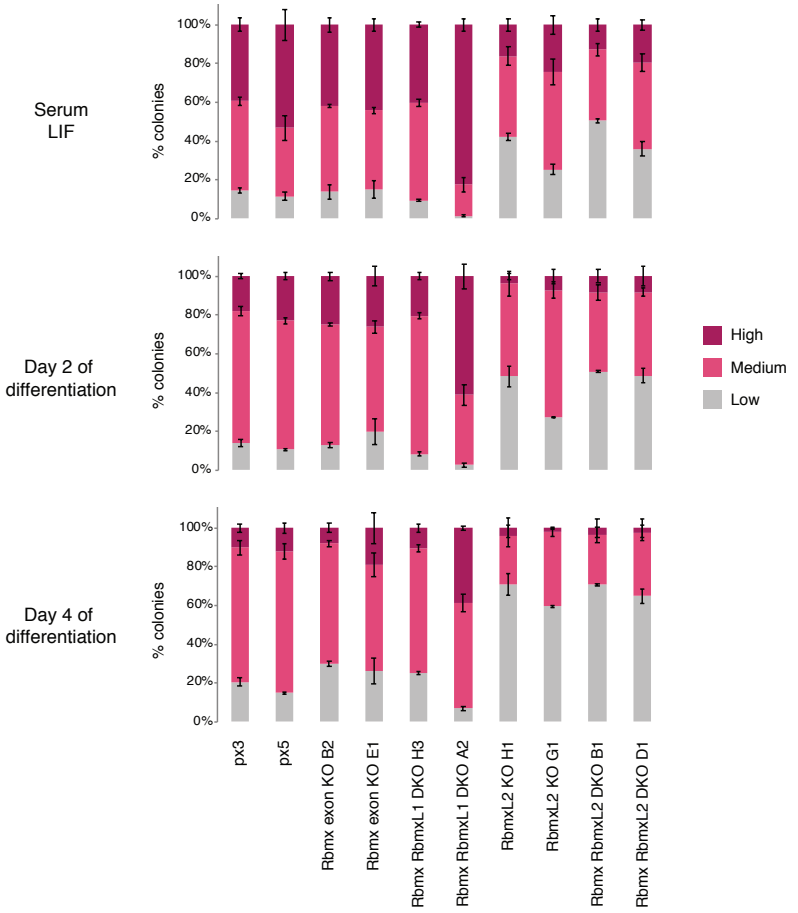


Figure 13: Pluripotency decrease assessed by AP staining during LIF removal clonal differentiation assay in ESCs depleted of Rbmx/RbmxL2 regulatory axis. ESCs were plated in SL at clonal density, LIF was removed for two or four days after which colonies were stained with AP staining. Colonies were counted and classified according to their level of AP staining (high, medium or low). Stacked plot depicts the population distribution of stained colonies and shows a clear delay in loss of pluripotency for Rbmx RbmxL1 KO clone A2 and a faster acquisition of differentiated characteristics in RbmxL2 KOs and Rbmx RbmxL2 DKOs. A minimum of 60 colonies were counted per condition per clone. n=3.

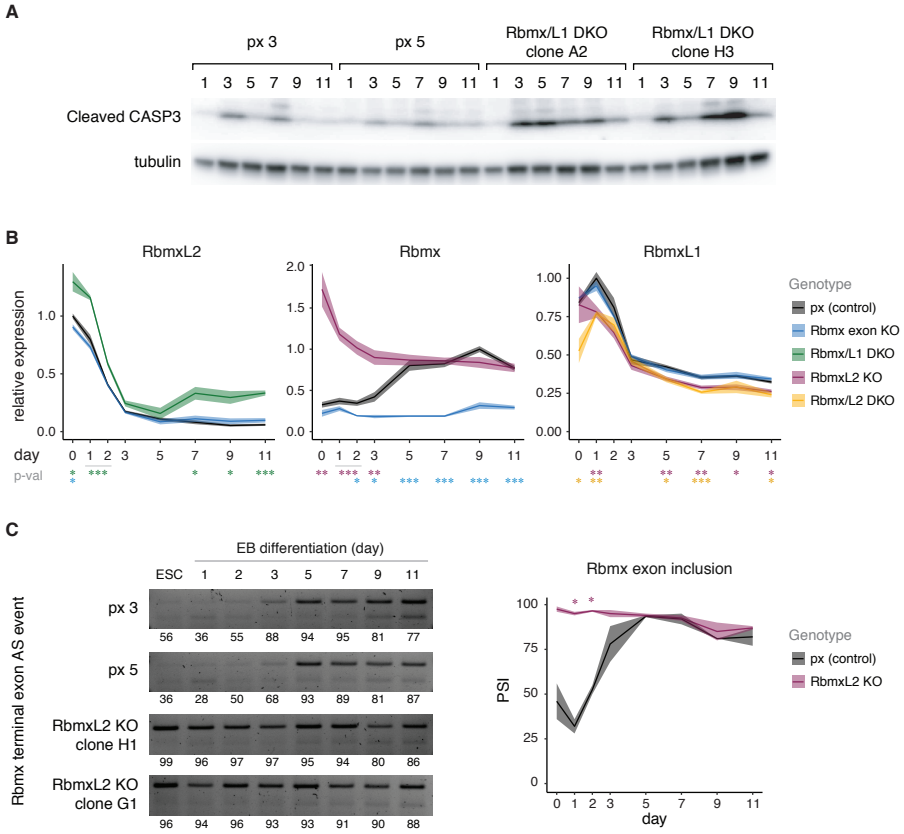


Figure 14: Cell death, *Rbmx*, *RbmxL1* and *RbmxL2* gene expression and *Rbmx* alternative exon inclusion dynamics during embryoid body differentiation. (A) Western blot analysis displaying protein levels of cleaved Caspase-3 as a marker of apoptosis during embryoid body differentiation shows higher levels of cell death in Rbmx Rbmxl1 DKO clones compared to controls. **(B)** Line charts depicting *Rbmx*, *RbmxL1* and *RbmxL2* gene expression levels throughout EB differentiation as measured by RTqPCR show differential expression dynamics of *RbmxL2* in Rbmx RbmxL1 DKOs and of *Rbmx* in Rbmx exon KOs. Expression values are normalized relative to the expression of Beta-actin (*Actb1*) and further normalized by the average maximum expression in the control cell lines across each differentiation experiment. For each genotype, central lines indicate mean values of both clones and ribbons, the standard error of the mean. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated for each time point: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, coloured by genotype. Experiment repeated in duplicate. **(C)** Left, gel images showing RT-PCR amplification of *Rbmx* alternative exon usage across controls and RbmxL2 KOs during EB differentiation. Numbers at the bottom of each lane under each image indicate PSI levels based on gel quantification. Right, line chart illustrating

every day for the first three days and every other day in the following 8 days and monitored gene expression levels by reverse transcription quantitative polymerase chain reaction (RTqPCR). One thing that was particularly striking when we performed these experiments was that both Rbmx RbmxL1 DKO clones had increased cell death (Figure 14A) and low attachment rates at later time points. Indeed, although the normal EB cell aggregates were formed, these persisted for longer with a low plate attachment rate and did not grow to the size that other KOs and controls did. This, together with the high levels of cell death observed, suggested that Rbmx RbmxL1 DKOs were refractory to differentiation.

First, we examined the relative expression levels of *Rbmx*, *RbmxL1* and *RbmxL2* during EB formation. Interestingly all KOs downregulated *RbmxL2* upon differentiation (Figure 14B). This included the Rbmx RbmxL1 DKOs that had higher expression levels of *RbmxL2* to begin with, as we had previously observed in ESCs (Figure 10A). Importantly, Rbmx RbmxL1 DKOs start from higher levels of *RbmxL2* but they adjust to the same expression levels as controls and Rbmx exon KOs by day 2 of differentiation (Figure 14B). We observed, however, a succeeding increase in *RbmxL2* expression levels in Rbmx RbmxL1 DKOs at day seven of differentiation and these levels were then stably higher than controls for the rest of the differentiation. Nonetheless, we cannot rule out the possibility that this increase in *RbmxL2* expression levels is a consequence of the high cell death that we observed. Certainly, it could be that those cells that remain alive are remaining in a pluripotent state and that those that die are cells that are differentiating. Indeed, this hypothesis could potentially explain what seems like a regain of *RbmxL2* gene expression levels. Additionally, we observed that it was precisely the Rbmx RbmxL1 KO clone A2 that had higher levels of *RbmxL2*, therefore, a more acute phenotype than the other clone (clone H3). This more extreme phenotype could be reflecting what we saw in our previous LIF withdrawal experiments where clone A2 had clear and distinct augmented maintenance of pluripotent colonies during differentiation (Figure 13).

---

inclusion levels based on gel quantifications for two clones of each genotype. Central line represents mean value per genotype and ribbons, the maximum and minimum (range) values.

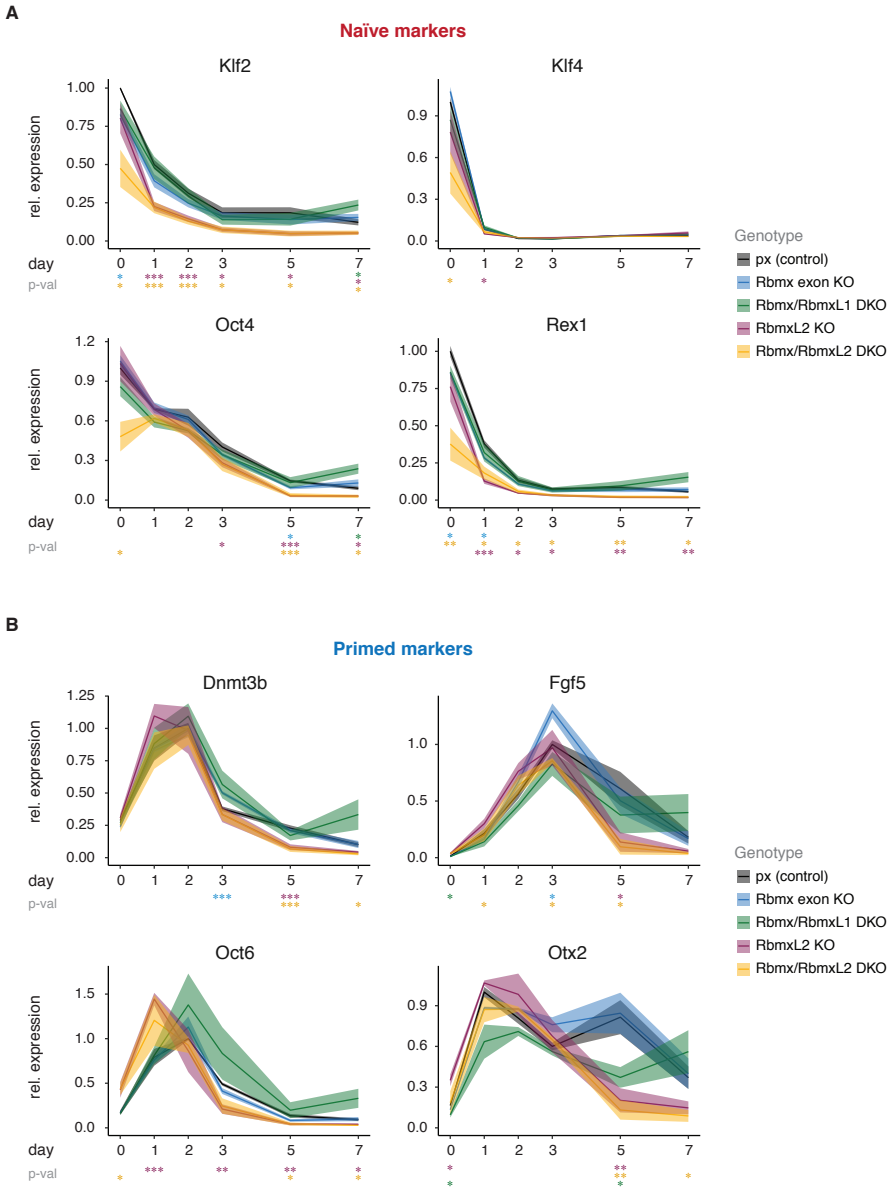


Figure 15: Gene expression profiles of naïve and primed markers during embryoid body differentiation in cell lines depleted of the Rbmx/RbmxL2 regulatory axis. **(A)** Line charts depicting *Oct4* (general pluripotency marker), *Klf4* (early naïve pluripotency marker), *Klf2* and *Rex1* (naïve pluripotency markers) gene expression levels throughout EB differentiation as measured by RTqPCR showing differential

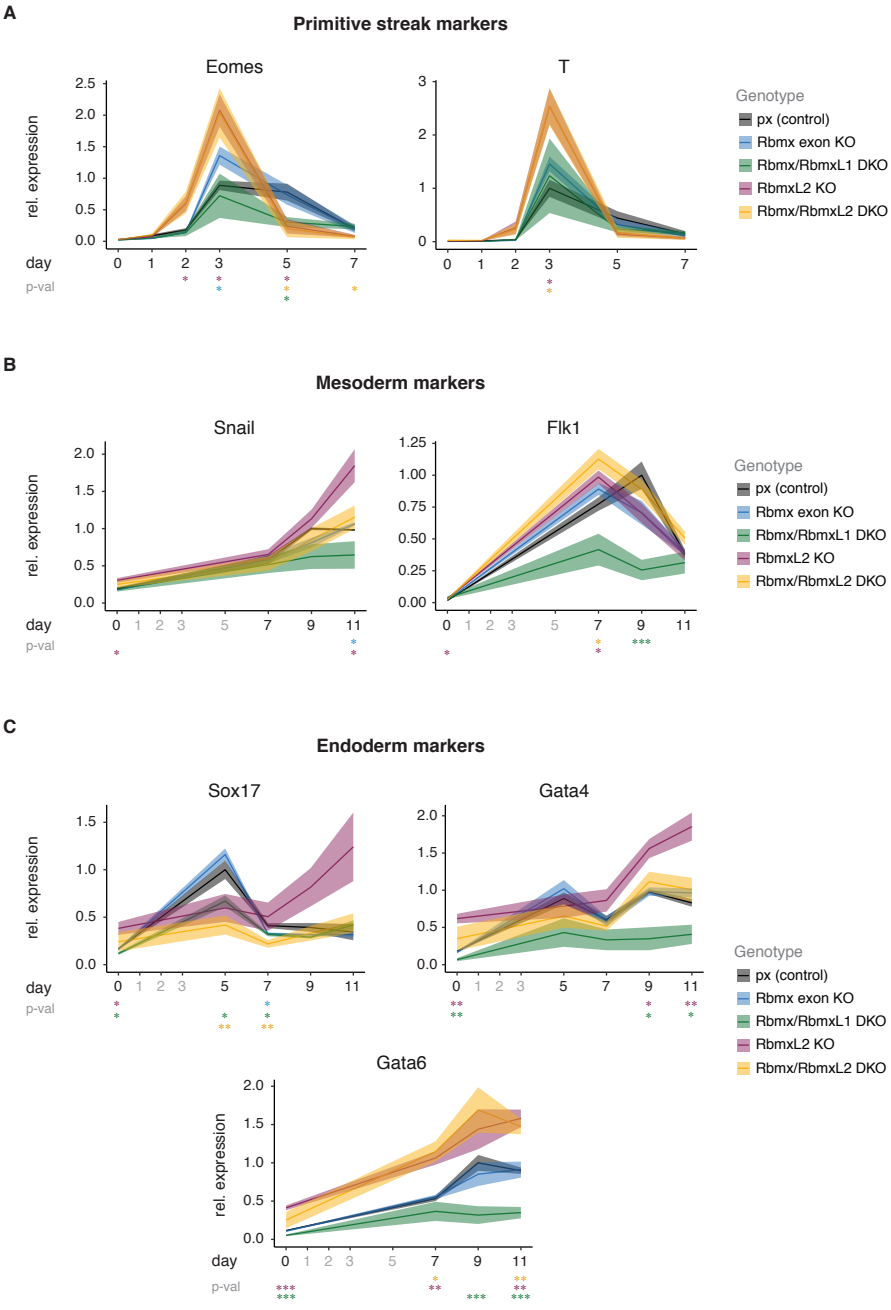
Rbmx expression levels increase upon differentiation until they stabilize at around day 5 (Figure 14B). Curiously, although *Rbmx* expression levels in RbmxL2 KOs start at a higher level at ESCs (Figure 10A, Figure 14B), they tend to stabilize at the same level as controls also by day 5. This could suggest that a regulatory mechanism is in place to stabilize *Rbmx* expression levels to a specific range during differentiation and that this regulatory mechanism is not present in ESCs, allowing for higher *Rbmx* levels in RbmxL2 KOs in that state. Another interesting observation is that in Rbmx exon KOs, *Rbmx* expression levels are maintained low throughout differentiation hinting at a possible regulatory mechanism executed by the alternative splicing of this exon.

RbmxL1 gene expression levels, on the other hand, seem to show a similar pattern throughout differentiation in KOs and controls, albeit *RbmxL2* depleted clones steadily maintain lower levels than the rest (Figure 14B). Interestingly, *RbmxL1* expression levels go down upon pluripotency exit and stabilize by day seven.

In the light of *Rbmx* gene expression dynamics during EB differentiation, we decided to look at the inclusion dynamics of *Rbmx* terminal alternatively spliced exon. We found that exon inclusion is also differentially regulated upon differentiation, starting at intermediate levels of PSI and increasing inclusion at the exit of naïve pluripotency (Figure 14C). By day 5, exon inclusion is at its highest and then seems to stabilize at high inclusion levels for the rest of the differentiation. Curiously, this exon is also highly included in RbmxL2 KO ESCs and this high inclusion is maintained thought EB formation. This compelling observation could potentially indicate a regulatory role of *RbmxL2* in the inclusion of this exon.

---

expression dynamics in KO cell lines. **(B)** Line charts depicting *Otx2*, *Oct6*, *Dnmt3b* (formative state, early primed pluripotency markers) and *Fgf5* (primed pluripotency marker) gene expression levels throughout EB differentiation as measured by RTqPCR show differential expression dynamics of KO cell lines. Expression values are first normalized relative to the expression of Beta-actin (*Actb1*) and further normalized by the average maximum expression in the controls across each differentiation experiment. Central lines depict the mean value for two replicates and two clones per genotype and ribbons, the standard error of the mean. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated for each time point: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, coloured by genotype.


A
Primitive streak markers
Eomes
T
rel. expression
day
p-val
B
Mesoderm markers
Snail
Flk1
C
Endoderm markers
Sox17
Gata4
Gata6
Genotype
px (control)
Rbmx exon KO
Rbmx/RbmxL1 DKO
RbmxL2 KO
Rbmx/RbmxL2 DKO

To explore the potential differences in exit from naïve pluripotency, entry into primed pluripotency and the following exit from primed pluripotency, we examined the levels of naïve and primed markers during EB differentiation in our KOs and controls. First, the expression levels of *Oct4*, a general pluripotency marker showed that by day five RbmxL2 KOs and Rbmx RbmxL2 DKOs had completely exited pluripotency whereas controls still showed detectable levels (Figure 15A). Expression levels of *Klf4*, an early marker of naïve pluripotency exit showed no difference among KOs. Importantly, the 24-hour time window we set might be too lengthy to look at the dynamics of early naïve-exit markers. On the other hand, *Klf2* and *Rex1*, two naïve pluripotency markers clearly showed that RbmxL2 KOs and Rbmx RbmxL2 DKOs have faster naïve pluripotency exit dynamics than controls. Furthermore, *Otx2* and *Oct6*, two formative stage and early primed pluripotency markers, and *Fgf5* and *Dnmt3b*, two primed pluripotency markers, also showed faster exit from primed pluripotency of RbmxL2 KOs and Rbmx RbmxL2 DKOs compared to controls (Figure 15B). Notably, *Oct6* gene expression dynamics also suggested an earlier entry into primed pluripotency for RbmxL2 KOs and one of the Rbmx RbmxL2 DKO clones, however, we did not detect this trend when looking at the other primed pluripotency markers.

Gene expression profiles of primitive streak markers *Brachyury* (T) and *Eomes* showed that RbmxL2 KOs and Rbmx RbmxL2 DKOs had substantial differences in dynamics compared to controls (Figure 16A). While at day three of EB differentiation RbmxL2 KOs and Rbmx RbmxL2 DKOs had a higher

---

Figure 16: Gene expression profiles of primitive streak, mesoderm and endoderm markers during embryoid body differentiation in cell lines depleted of the Rbmx/RbmxL2 regulatory axis. (A) Line charts depicting primitive streak markers *Eomes* and *Brachyury (T)* expression levels throughout EB differentiation as measured by RTqPCR showing differential expression dynamics in KO cell lines. **(B)** Line charts depicting mesodermal markers *Snail* and *Flk1* expression levels throughout EB differentiation as measured by RTqPCR show differential expression dynamics of KO cell lines. Line charts depicting endodermal markers *Sox17*, *Gata4* and *Gata6* expression levels throughout EB differentiation as measured by RTqPCR show differential expression dynamics of KO cell lines. Expression values are first normalized relative to the expression of Beta-actin (*Actb1*) and further normalized by the average maximum expression in the controls across each EB differentiation experiment. Central lines depict the mean value for two replicates and two clones per genotype and ribbons, the standard error of the mean. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated for each time point: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, coloured by genotype.

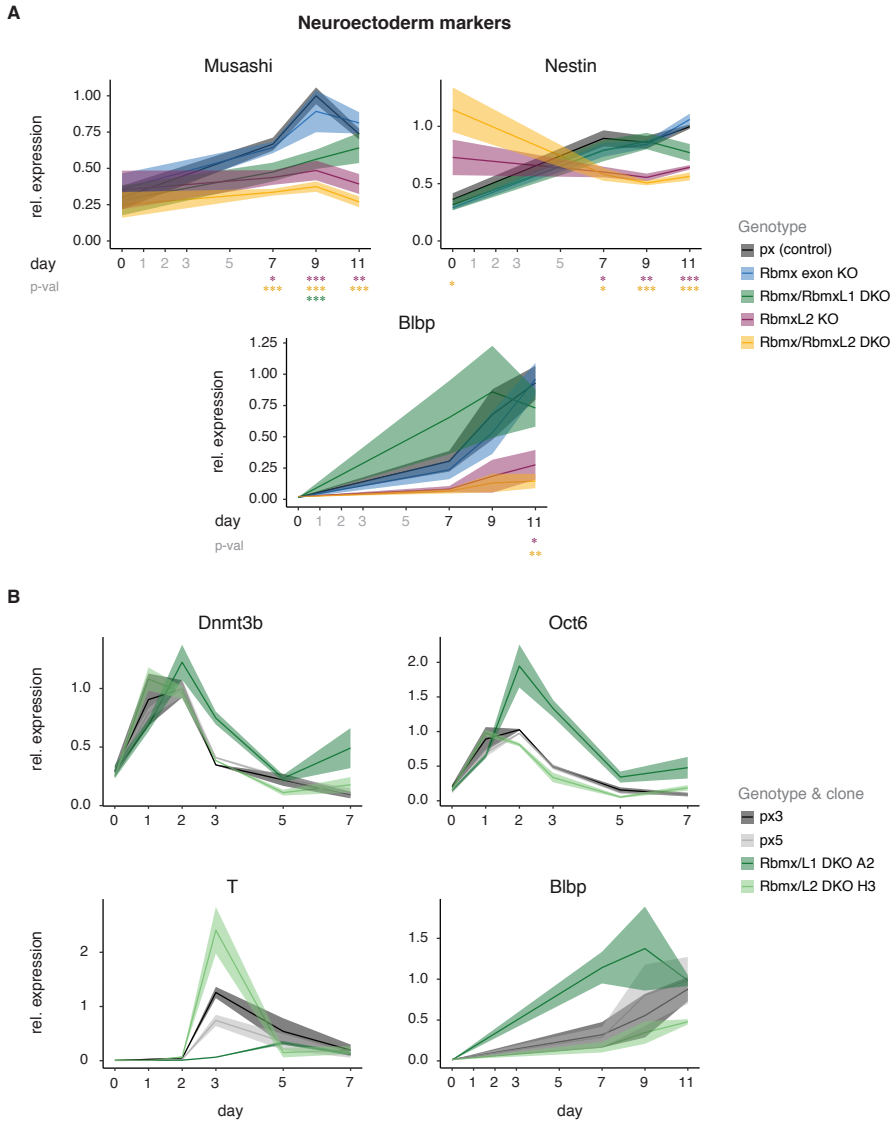


Figure 17: Gene expression profiles of neuroectoderm markers during embryoid body differentiation in cell lines depleted of the Rbmx/RbmxL2 regulatory axis and gene expression profile differences among Rbmx RbmxL1 KO clones. (A) Line charts depicting *Blbp*, *Nestin* and *Musashi* expression levels throughout EB differentiation as measured by RTqPCR show differential expression dynamics of KO cell lines. Expression values are first normalized relative to the expression of Beta-actin (*Actb1*) and further normalized by the average maximum expression in the controls

expression, by day five this expression was lower than controls. This change in the amplitude and duration of primitive streak marker expression levels could suggest greater synchronicity of RbmxL2 KOs and Rbmx RbmxL2 DKOs upon exit of pluripotency. This could be explained by a greater homogeneity of pluripotency factor levels across RbmxL2 KO and Rbmx RbmxL2 DKO cell populations as we observed previously (Figure 10C).

We then investigated whether depletion of the components of the Rbmx/RbmxL2 regulatory axis had any bias towards differentiation into any specific lineage. For this, we studied the gene expression of mesodermal, endodermal and neuroectodermal markers. In the case of mesoderm markers, we found that RbmxL2 KOs had differentially higher expression of *Snail*, a key inducer of epithelial-to-mesenchymal transition (EMT)[427] and a lateral plate mesoderm marker by day eleven of differentiation (Figure 16B). Additionally, RbmxL2 KOs and Rbmx RbmxL2 DKOs showed an earlier *Flk1* expression peak than controls. *Flk1* is a marker for mesoderm and angiogenic precursors, it supports an endothelial fate and is expressed in the lateral plate mesoderm but more restricted to blood vessels and cardiac lineages[428–431] thus these results suggested an earlier determination towards that fate.

Regarding induction of the endodermal lineage during EB differentiation, we found a clear bias of RbmxL2 KOs towards overexpressing *Gata4* and a trend to overexpress *Sox17* (Figure 16C). Additionally, RbmxL2 KOs and Rbmx RbmxL2 DKOs seemed to have an increased expression of *Gata6* throughout EB differentiation compared to controls. Together these results seemed to indicate that RbmxL2 KOs have a clear bias towards endomesodermal lineage whereas this bias is more restricted in Rbmx RbmxL2 DKOs. This might point to the need for a right balance between *Rbmx* and *RbmxL2* expression for an unbiased lineage differentiation to occur.

---

across each EB differentiation experiment. Central lines depict the mean value for two replicates and two clones per genotype and ribbons, the standard error of the mean. **(B)** Line charts showing the high level of variability between the two clones of Rbmx RbmxL1 DKOs for some of the studied markers. Central lines indicate the average of two experiments and the ribbons mark the values for each individual replicate. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated for each time point: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, coloured by genotype.

Concerning the neuroectodermal lineage, we observed that both RbmxL2 KOs and Rbmx RbmxL2 DKOs had clear difficulties to differentiate into it (Figure 17A). On the other hand, as with the rest of the lineage markers, we did not observe substantial differences in neuroectodermal marker gene expression profiles of Rbmx exon KOs.

Rbmx RbmxL1 DKO clone A2 showed a deviating phenotype compared to the other Rbmx RbmxL1 DKO clone, clone H3. In clone A2, we detected a regain of pluripotent markers by day seven and a prolonged expression of *Dnmt3b* and *Oct6* (Figure 17B). Having a delay in primed pluripotency downregulation potentially goes in line with the hypothesis we proposed previously in which differentiating cells were dying and thus the expression signals could come from the pluripotent cells that remained, however this hypothesis clashes with the fact that both Rbmx RbmxL1 DKO clones behave differently. Furthermore, Rbmx RbmxL1 DKO clone H3 showed an increase in amplitude in primitive streak marker expression dynamics compared to control that was similar to what we saw for RbmxL2 KOs and Rbmx RbmxL2 DKOs (Figure 16A, Figure 17B). Rbmx RbmxL1 DKO clone A2, however, did not seem to activate gene expression of *Brachyury* or *Eomes* to the levels of control cell lines although there was a slight gain in expression by day five of EB differentiation (Figure 17B). Also, *Snail* and *Flk1* hardly changed expression and maintained very low levels during EB differentiation in the A2 clone whereas in clone H3, *Snail* levels were similar to controls and *Flk1* seemed to be downregulated sooner than controls (data not shown). Lastly, *Blbp* expression also showed a contrasting profile across the two RbmxL1 DKO clones with clone A2 upregulating *Blbp* expression sooner than controls and clone H3 (Figure 17B).

Part II: Rbm47/Esrp regulatory axis

Epithelial splicing regulatory protein 1 (ESRP1), Epithelial splicing regulatory protein 2 (ESRP2), and RNA-binding protein 47 (RBM47) are three regulators that play a combinatorial regulation over a splicing program that is reverted during epithelial to mesenchymal transition (EMT)[425,432,433]. Although RBM47 mostly promotes epithelial splicing patterns by coregulating targets with ESRP1 and ESRP2, it also promotes different or opposite inclusion patterns in some alternative splicing events[425]. So far, these contrasting patterns have been overlooked.

We found that *Rbm47* is expressed in naïve pluripotency while it is very lowly expressed in primed pluripotency (Figure 18A). *Esrp1* expression however is low in 2iL ESCs but increases in SL ESCs and EpiSCs (Figure 18A). This led us to hypothesize that there could be a functional substitution for the regulation of RBM47/ESRP1 co-targets in the naïve-to-primed transition. This substitution could be responsible for partly switching the alternative splicing landscape from a naïve pluripotency state to a primed pluripotency state.

Esrp2 is a paralog of *Esrp1*; however, ESRP1 has been shown to be the primary driver of the ESRP-regulated alternative splicing programme[390,432,433]. *Esrp2* is expressed in both naïve and primed pluripotency, however, in 2iL ESCs the mean levels are twice the expression levels in EpiSCs (Figure 18A). Importantly, the expression levels of *Esrp1* in both SL ESCs and EpiSCs are much higher than *Esrp2* (4 times and 12 times higher, respectively). However, *Esrp2* expression levels in 2iL ESCs are more than 8 times higher than *Esrp1* (Figure 18A). ESRP1 and ESRP2 have a similar amino acid sequence (60% identity) and possess three highly similar RRM domains (89%, 68% and 76% identity) (Figure 18B). These three RRM domains have a high degree of conservation in ESRP orthologs of chicken, *Drosophila melanogaster* (*Fusilli*) and *Caenorhabditis elegans* (*sym-2*), specially RRM1[434], suggesting similar RNA target sequences and interaction partners of splicing regulatory complexes. Additionally, ESRP2 has been shown to partially compensate for *Esrp1* loss[390] suggesting functional redundancy.

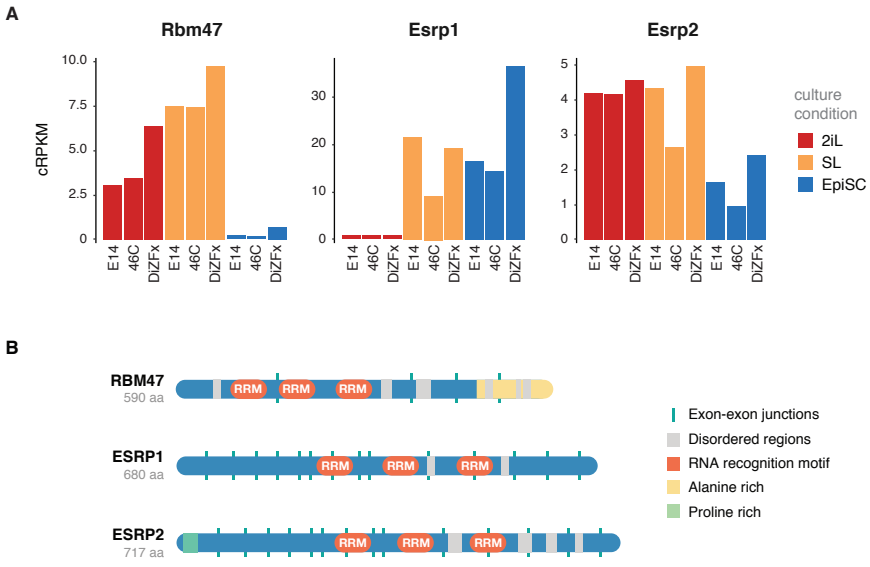


Figure 18: Expression level and protein features of RBM47, ESRP1 and ESRP2 in naïve and primed pluripotency. (A) Expression levels of *Rbm47*, *Esrp1* and *Esrp2* across three independent cell lines in naïve and primed cell culture conditions show differing expression patterns. **(B)** Protein features of RBM47, ESRP1 and ESRP2. Blue lines depict exon-exon junctions. RRM: RNA recognition motif.

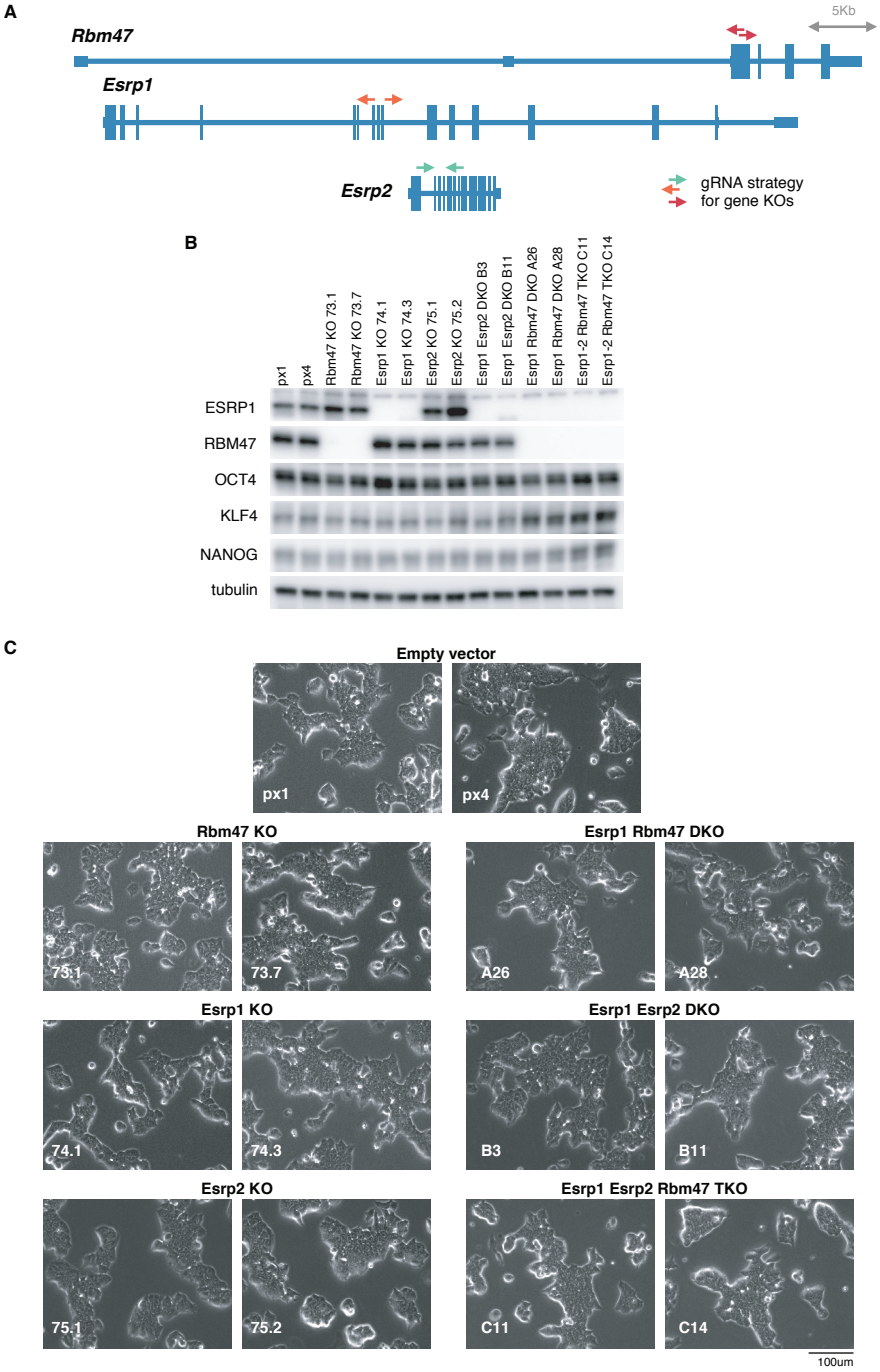
Altering the Rbm47/Esrp regulatory axis in naïve pluripotency

To understand the functional roles of RBM47 and ESRP1 on pluripotent cell identity, we made a set of full gene knockout (KO) ESC lines using the CRISPR/Cas9 gene-editing system. Additionally, since it was possible that in our system ESRP2 could be compensating for *Esrp1* depletion too, we also generated full gene KO ESC lines combining *Esrp1* and *Esrp2* deletions. Specifically, *Rbm47* deletions were directed by the use of two guide RNAs (gRNAs) targeting two different sites in the first coding exon of the gene (Figure 19A). This strategy generated deletions of ~750bp. *Esrp1* deletions were directed by two gRNAs targeting exon 7 and intron 9 (Figure 19A). This generated *Esrp1* deletions of over 550bp. *Esrp2* was targeted by two gRNAs that bound to introns 2 and 5 (Figure 19A). This resulted in deletions of over 900bp.

We generated single KO ESC cell lines for each splicing factor *Rbm47*, *Esrp1* and *Esrp2* to analyse the roles of each axis component separately. Additionally, we generated double KO ESC cell lines for *Esrp1* together with *Esrp2* to study the compensatory role of ESRP2, and double KO ESC cell lines for *Esrp1* together with *Rbm47* to examine the impact of this regulatory axis on naïve and primed pluripotency. Moreover, we generated triple KO ESC cell lines for *Rbm47*, *Esrp1* and *Esrp2* to abolish any possible compensatory effect that *Esrp2* expression could render. All of the above cell lines were generated in parallel, and various clones were obtained. Furthermore, we generated control ESC lines transfected in parallel with an empty vector (termed px). We validated the depletion of ESRP1 and RBM47 by western blot analysis (Figure 19B). Sadly, we were not able to validate ESRP2 depletion via western blot analysis due to the absence of a specific antibody. However, *Esrp2* transcript was clearly diminished in all confirmed Esrp2 KOs (Figure 20A). All of the cell lines were subsequently karyotyped and expanded for the following experiments (see Materials and methods for further details). We carried out subsequent experiments using two clones for each genotype to ensure solid reproducible conclusions. Morphologically, all twelve KO cell lines had a similar appearance to controls when grown in the permissive SL conditions (Figure 19C).

We found that *Rbm47* depletion in ESCs under SL conditions leads to a slight increase in *Esrp1* mRNA and protein levels (Figure 19B, Figure 20A). Additionally, both Esrp1 Esrp2 DKO clones show a slight downregulation of *Rbm47* expression at the transcript (Figure 20A) and at the protein level (Figure 19B). Finally, one Esrp2 KO clone, clone 75.2, consistently shows an increase in *Esrp1* expression both at the transcript level (Figure 20A) and at the protein level (Figure 19B). Together, these results might suggest a moderate compensatory regulation across the components of the Rbm47/Esrp regulatory axis.

To investigate the effect that the depletion of *Rbm47*, *Esrp1* and/or *Esrp2* had on pluripotency, we tested the general expression patterns of naïve and primed pluripotency markers in our KO ESC lines in SL conditions and found no significant differences (Figure 20A). Yet, we did detect higher KLF4 protein levels in Esrp1 Esrp2 DKOs and Esrp1 Esrp2 Rbm47 TKOs (Figure 19B) suggesting that the interplay between these splicing factors may affect naïve pluripotency state. However, this was not reflected at the transcript level (Figure 20A). Furthermore, all KO cell lines grown in SL conditions display similar levels of heterogeneity in the expression of NANOG, OCT4 and KLF4 with


A
Rbm47
5Kb
Esrp1
Esrp2
gRNA strategy
for gene KOs
B
px1
px4
Rbm47 KO 73.1
Rbm47 KO 73.7
Esrp1 KO 74.1
Esrp1 KO 74.3
Esrp2 KO 75.1
Esrp2 KO 75.2
Esrp1 Esrp2 DKO B3
Esrp1 Esrp2 DKO B11
Esrp1 Rbm47 DKO A26
Esrp1 Rbm47 DKO A28
Esrp1-2 Rbm47 TKO C11
Esrp1-2 Rbm47 TKO C14
ESRP1
RBM47
OCT4
KLF4
NANOG
tubulin
C
Empty vector
px1
px4
Rbm47 KO
73.1
73.7
Esrp1 Rbm47 DKO
A26
A28
Esrp1 KO
74.1
74.3
Esrp1 Esrp2 DKO
B3
B11
Esrp2 KO
75.1
75.2
Esrp1 Esrp2 Rbm47 TKO
C11
C14
100um

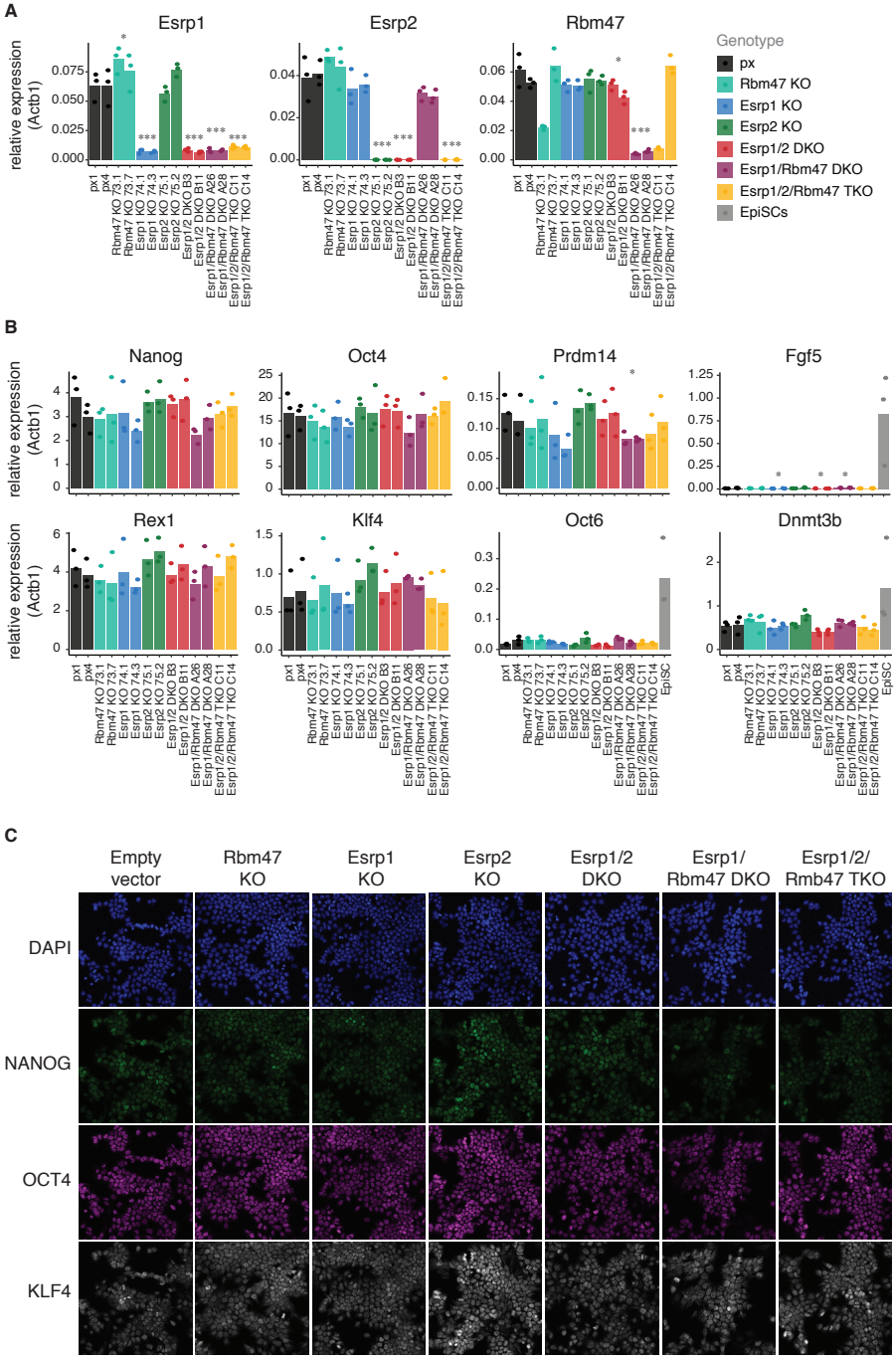
the sole exception of Esrp2 KOs which show higher heterogeneity among the population of cells (Figure 20B). More specifically, in Esrp2 KOs more cells in the population have high levels of OCT4, NANOG and KLF4 showing a wider spectrum of levels among the population compared to controls. Interestingly, these higher levels of naïve pluripotency markers in subpopulations of Esrp2 KOs did not translate into higher levels in the whole population as seen by western blot (Figure 19B).

We performed RNA-seq of Esrp1 Esrp2 DKOs, Rbm47 KOs and Esrp1 Esrp2 Rbm47 TKOs ESCs grown in parallel with controls in SL conditions to examine the effect of these gene depletions on the general pluripotency state. Gene Set Enrichment Analysis (GSEA)[426] shows that Esrp1 Esrp2 DKOs are significantly more enriched in the differential gene expression profile of naïve pluripotency compared to controls (Figure 21A). Additionally, they display a significant downregulation of primed pluripotency gene expression profile (Figure 21B). In contrast, Rbm47 KOs exhibit an upregulation of primed pluripotency gene expression profile (Figure 21B) and a downregulation of naïve pluripotency gene expression profile (Figure 21A). These results support the hypothesized opposing roles for *Esrp1/2* and *Rbm47* in the control of pluripotency states. Finally, disrupting the whole Rbm47/Esrp axis elicits a gene expression profile that is neither enriched in naïve pluripotency upregulated genes (Figure 21A) or in primed pluripotency upregulated genes (Figure 21B).

Hierarchical clustering of all samples by differential alternative splicing profiles determined the effect of *Esrp1 Esrp2* and/or *Rbm47* depletion on alternative splicing landscapes. Interestingly, Esrp1 Esrp2 DKOs cluster separately from Esrp1 Esrp2 Rbm47 TKOs and Rbm47 KOs (Figure 21C). This indicates that

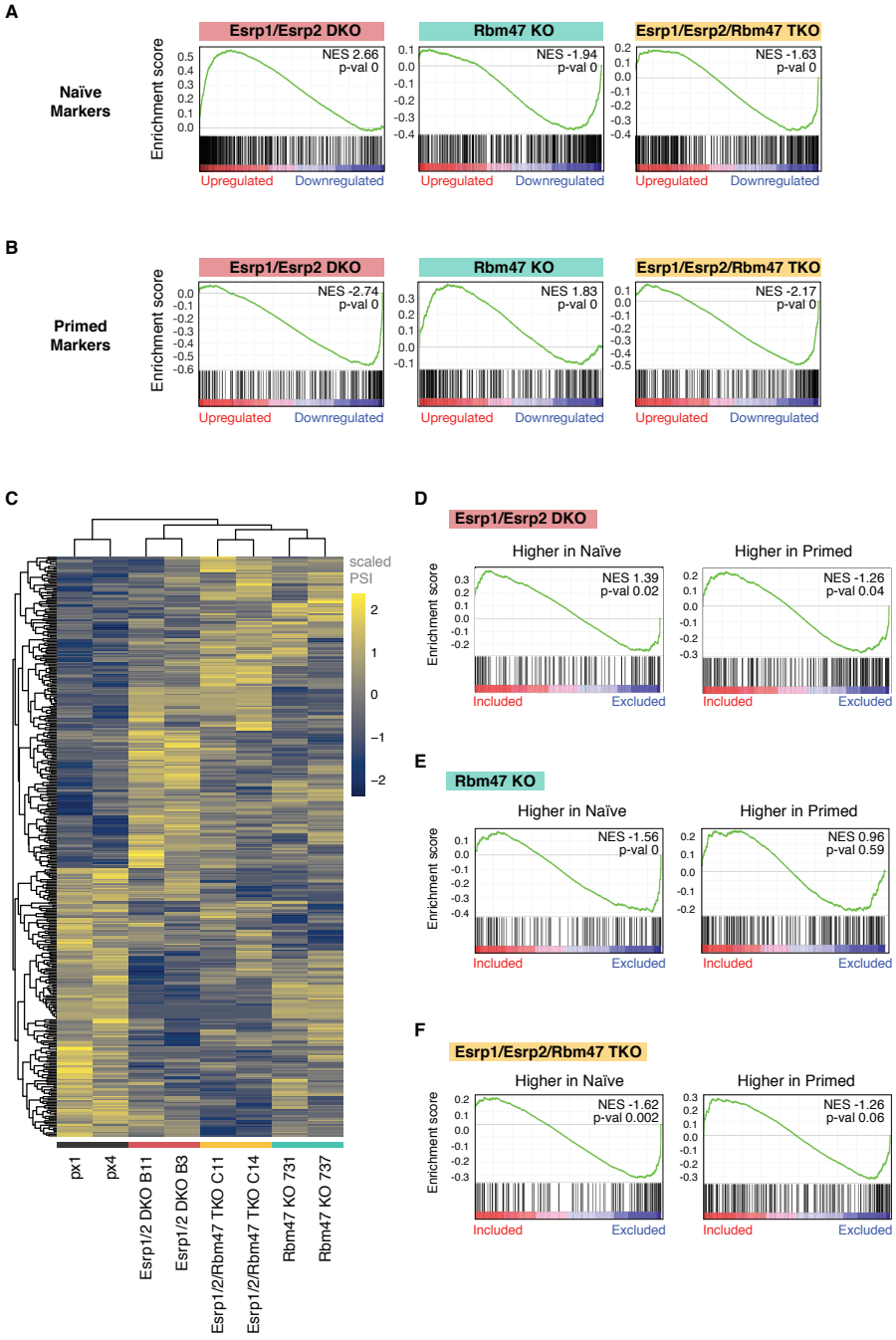
---

Figure 19: Generation of ESC lines with *Rbm47*, *Esrp1* and/or *Esrp2* CRISPR/Cas9 gene depletions. (A) CRISPR/Cas9 strategy for depletion of *Rbm47*, *Esrp1* and *Esrp2* genes. Red arrows illustrate the gRNAs that were used to delete a region of over 750bp in the first coding exon of *Rbm47.* Orange arrows illustrate the gRNAs that were used to delete a region of over 550bp from exon 7 to intron 9 of *Esrp1*. Green arrows illustrate the gRNAs that were used to delete a region of over 900bp between intron 7 and 9 of *Esrp2*. **(B)** Western blot analysis displaying protein levels of ESC cell lines grown in SL conditions shows complete depletion of ESRP1 and RBM47 proteins in KO cell lines. Protein levels of pluripotency markers OCT4, KLF4 and NANOG are also shown. Empty vector controls are labelled as px1 and px4. Representative image of experiment repeated in triplicate. **(C)** Bright-field images of ESC KO cell lines grown in SL conditions show similar morphology across in all cell lines.


A
Esrp1
Esrp2
Rbm47
relative expression (Actb1)
Genotype
px
Rbm47 KO
Esrp1 KO
Esrp2 KO
Esrp1/2 DKO
Esrp1/Rbm47 DKO
Esrp1/2/Rbm47 TKO
EpiSCs
B
Nanog
Oct4
Prdm14
Fgf5
Rex1
Klf4
Oct6
Dnmt3b
C
Empty vector
Rbm47 KO
Esrp1 KO
Esrp2 KO
Esrp1/2 DKO
Esrp1/ Rbm47 DKO
Esrp1/2/ Rmb47 TKO
DAPI
NANOG
OCT4
KLF4

Rbm47 deletion drives the alternative splicing switch in Esrp1 Esrp2 Rbm47 TKOs over the potential effect of *Esrp1 Esrp2* depletion on isoform diversity. Alternative Splicing Set Enrichment Analysis (ASSEA) confirmed this observation. Firstly, Esrp1 Esrp2 DKOs are significantly enriched in a more naïve pluripotency alternative splicing landscape than controls (Figure 21D). This is accomplished by both including more naïve highly included exons and by excluding more primed highly included exons, suggesting that ESRP1 influences both exon inclusion and exclusion in the naïve-to-primed transition. RBM47, however, seems to affect mainly exon inclusion as ASSEA shows a significant enrichment in excluding alternatively spliced exons that were also more excluded in EpiSCs but does not show an enrichment in including exons that were more included in EpiSCs (Figure 21E). Finally, Esrp1 Esrp2 Rbm47 TKOs also show this trend towards a primed pluripotency exon exclusion profile, albeit with a greater enrichment score than Rbm47 KOs (Figure 21F). This hinted at the possibility that a set of naïve-to-primed exons might be co-regulated in the same direction by Rbm47 and Esrp1/Esrp2. Nevertheless, it must be noted that the increase in enrichment is modest which implies that probably these are not many events.

Figure 20: Naïve pluripotency markers upon depletion of Rbm47/Esrp regulatory axis in ESCs. (A) Bar plots displaying relative gene expression levels of *Rbm47*, *Esrp1* and *Esrp2* measured by quantitative PCR across all ESC KO lines grown in SL. Experiment performed in triplicate. **(B)** Bar plots depicting relative gene expression levels of general pluripotency (*Oct4*), naïve pluripotency (*Prdm14*, *Klf4*, *Nanog*, *Rex1*), formative pluripotency (*Oct6*, *Dnmt3b*) and primed pluripotency (*Fgf5*) markers measured by quantitative PCR in all KO lines. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$. Experiment performed in triplicate. **(C)** Microscopy images of KO clones immunostained for naïve pluripotency markers (NANOG and KLF4) and general pluripotency marker (OCT4) shows higher heterogeneity in their expression among Esrp2 KO colonies.


A
Naïve Markers
Esrp1/Esrp2 DKO
NES 2.66
p-val 0
Rbm47 KO
NES -1.94
p-val 0
Esrp1/Esrp2/Rbm47 TKO
NES -1.63
p-val 0
Enrichment score
Upregulated
Downregulated
B
Primed Markers
Esrp1/Esrp2 DKO
NES -2.74
p-val 0
Rbm47 KO
NES 1.83
p-val 0
Esrp1/Esrp2/Rbm47 TKO
NES -2.17
p-val 0
C
scaled PSI
px1
px4
Esrp1/2 DKO B11
Esrp1/2 DKO B3
Esrp1/2/Rbm47 TKO C11
Esrp1/2/Rbm47 TKO C14
Rbm47 KO 731
Rbm47 KO 737
D
Esrp1/Esrp2 DKO
Higher in Naïve
NES 1.39
p-val 0.02
Higher in Primed
NES -1.26
p-val 0.04
Included
Excluded
E
Rbm47 KO
Higher in Naïve
NES -1.56
p-val 0
Higher in Primed
NES 0.96
p-val 0.59
F
Esrp1/Esrp2/Rbm47 TKO
Higher in Naïve
NES -1.62
p-val 0.002
Higher in Primed
NES -1.26
p-val 0.06

Impact of Rbm47/Esrp regulatory axis on cell differentiation

We used clonal monolayer differentiation upon LIF-removal assays to test the influence of Rbm47/Esrp regulatory axis on cell differentiation. Firstly, we found that Esrp1 Rbm47 DKOs have a greater proportion of highly AP stained colonies compared to controls and all other *Esrp1*, *Esrp2* and/or *Rbm47* depletion mutants when grown in SL (Figure 22). This fits in with our previous observation of high KLF4 protein levels in Esrp1 Rbm47 DKOs compared to controls (Figure 19B). However, Esrp1 Esrp2 Rbm47 TKOs also showed higher levels of KLF4 but do not seem to have a greater proportion of highly AP stained colonies compared to controls (Figure 19B, Figure 22), thus pointing at another possible explanation. Importantly, after two days of LIF removal, Esrp1 Rbm47 DKOs still retain a greater proportion of highly AP stained colonies compared to controls. At this point, the Esrp2 KO clone 75.2 also shows the same high degree of AP stained colonies. This clone has

---

Figure 21: Gene set and alternative splicing set enrichment analysis shows Esrp1 Esrp2 DKOs upregulate naïve pluripotency whereas Rbm47 KOs upregulate primed pluripotency. (A) GSEA depicting enrichment in naïve gene expression profile of differentially expressed genes across *Esrp1 Esrp2* and/or *Rbm47* depleted SL ESC lines compared to controls shows Esrp1 Esrp2 DKOs significantly upregulate naïve pluripotency whereas Rbm47 KOs and Esrp1 Esrp2 Rbm47 TKOs downregulate it. **(B)** GSEA depicting enrichment in primed gene expression profile of differentially expressed genes across *Esrp1 Esrp2* and/or *Rbm47* depleted SL ESC lines compared to controls shows Esrp1 Esrp2 DKOs and Esrp1 Esrp2 Rbm47 TKOs significantly downregulate primed pluripotency whereas Rbm47 KOs upregulate it. Ranked lists of naïve and primed markers were made with the differential gene expression profile defined using paired comparisons of 2iL vs. EpiSCs across our three independent cell lines (E14, 46C, DizFx) and E4.5 vs. E6.5 epiblasts (see materials and methods for further details). **(C)** Heatmap depicting scaled PSI levels of differential AS events (cassette exon, intron retention, alternative 5'ss or alternative 3'ss events) occurring across *Esrp1 Esrp2* and/or *Rbm47* depleted SL ESC lines. Empty vector controls are labelled as px1 and px4. **(D)** ASSEA of Esrp1 Esrp2 DKOs shows enrichment in inclusion of primed excluded exons (higher in naïve) and in exclusion of primed included exons (higher in primed) compared to controls. **(E)** ASSEA of Rbm47 KOs shows enrichment in exclusion of primed excluded exons compared to controls. **(F)** ASSEA of Esrp1 Esrp2 Rbm47 TKOs shows enrichment in exclusion of primed excluded exons. In all cases for ASSEA we used ranked lists of differentially included exons made with the differential PSI profile defined using paired comparisons of 2iL vs. EpiSCs across our three independent cell lines (E14, 46C, DizFx) (see materials methods for further details). NES: Normalized enrichment score.

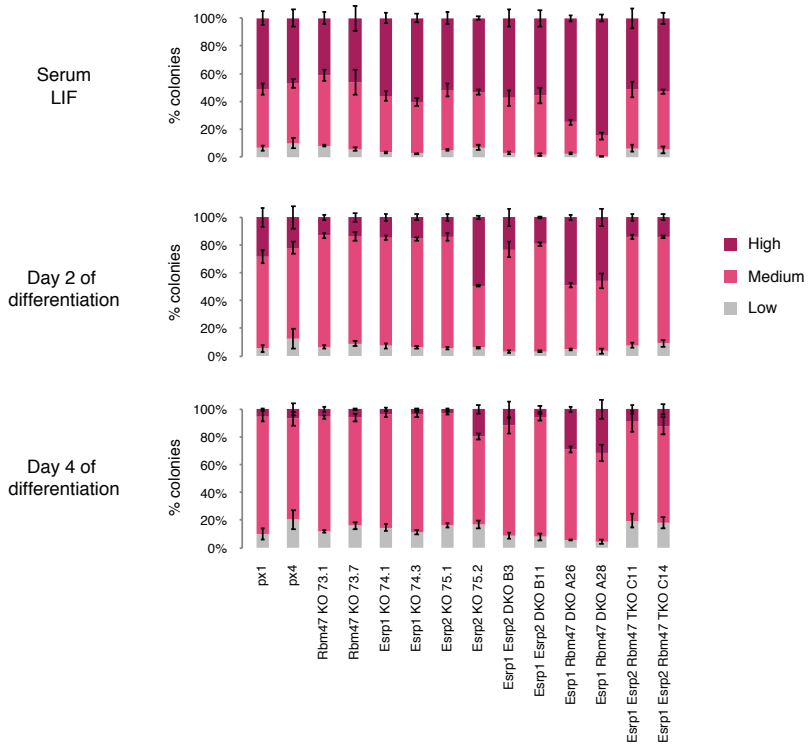


Figure 22: Pluripotency decrease assessed by AP staining during LIF removal clonal differentiation assay in ESCs depleted of Rbm47/Esrp regulatory axis. ESCs were plated in SL at clonal density, LIF was removed for two or four days after which colonies were stained with AP staining. Colonies were counted and classified according to their level of AP staining (high, medium or low). Stacked plots depict the population distribution of stained colonies and shows a significant delay in loss of pluripotency for Esrp2 KO clone 75.2 and Esrp1 Rbm47 DKOs. A minimum of 88 colonies were counted per condition per clone. n=3.

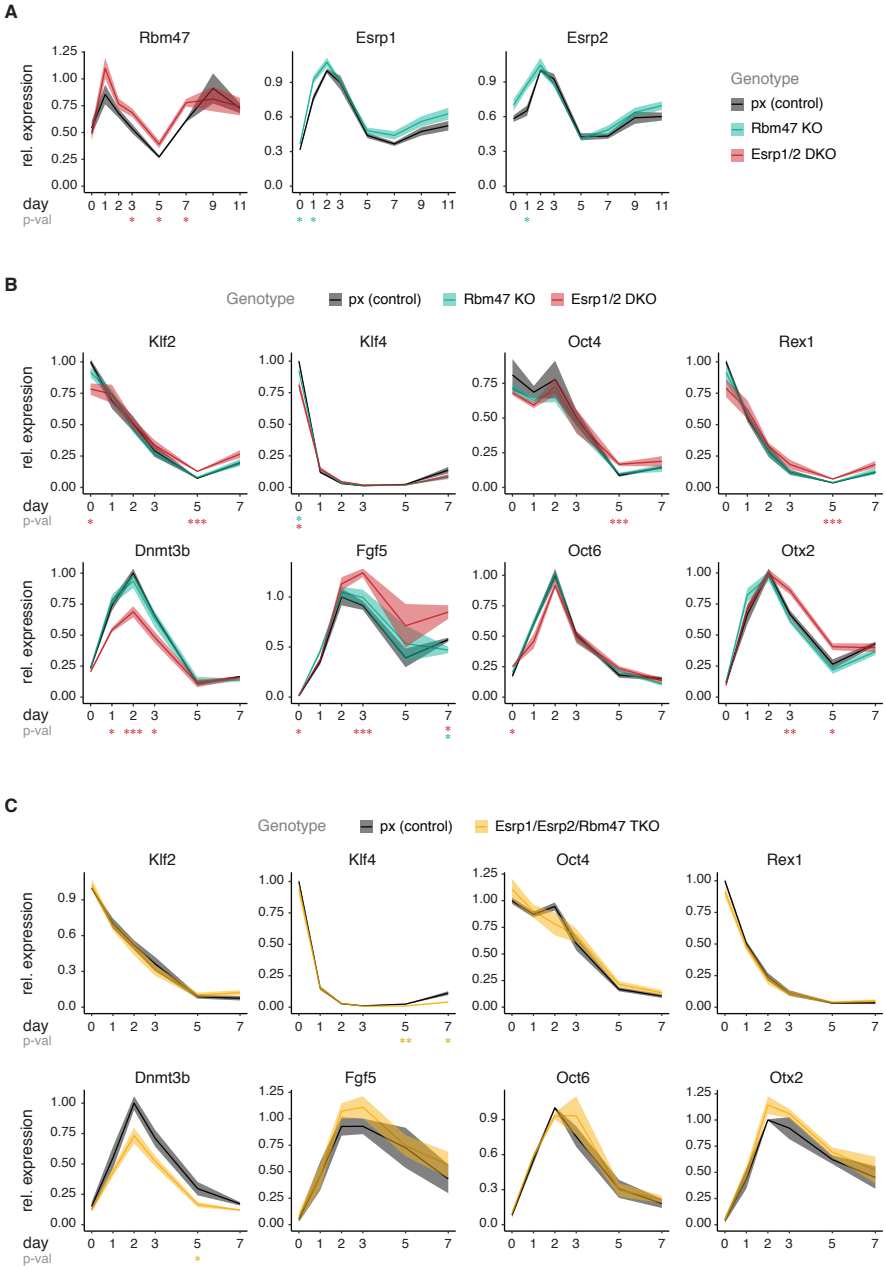
higher *Esrp1* expression levels than the other Esrp2 KO clone when cultured in SL (Figure 19B, 20A). This difference might account for the differences in pluripotency exit between the two. Curiously, however, these clones did not show differences in the proportion of AP stained colonies in SL conditions (Figure 22). By day four of differentiation, Esrp2 KO clone 75.2 retains this higher level of undifferentiated colonies in comparison to controls and the other Esrp2 KO clone (Figure 22). Additionally, Esrp1 Rbm47 DKOs retain

a greater proportion of undifferentiated colonies compared to controls and all other *Esrp1*, *Esrp2*, *Rbm47* depletion mutants. This indicates that Esrp1 Rbm47 DKOs have a delayed exit from pluripotency.

Given that Esrp1 Esrp2 DKOs display a more naïve differential gene expression profile and Rbm47 KOs display a more primed differential gene expression profile compared to controls (Figure 21A, B), it was unexpected to see that these cell lines do not show different pluripotency exit dynamics (Figure 22). Moreover, Esrp1 Esrp2 Rbm47 TKOs seem to display an alternative differential gene expression profile relative to controls, which does not relate to either naïve or primed pluripotency (Figure 21A, B) yet they also don't seem to have different pluripotency exit dynamics (Figure 22). This prompted us to study the levels of naïve and primed markers during the EB differentiation time-course of these KOs relative to controls. We did this in two independent rounds of experiments. First, testing Esrp1 Esrp2 DKOs, Rbm47 KOs and controls, and separately, testing Esrp1 Esrp2 Rbm47 TKOs and controls.

Importantly, during the first days of EB differentiation, *Rbm47*, *Esrp1* and *Esrp2* expression levels seem to follow a similar pattern: first they increase, reaching a peak at 1 or 2 days of differentiation and then decrease until day 5 or 7 (Figure 23A). Throughout most of it, *Rbm47* expression levels are steadily higher in Esrp1 Esrp2 DKOs compared to controls, whereas *Esrp1* and *Esrp2* expression levels show a slight tendency to be higher in Rbm47 KOs compared to controls (Figure 23A). This might suggest a moderate compensatory regulation across the components of the Rbm47/Esrp regulatory axis.

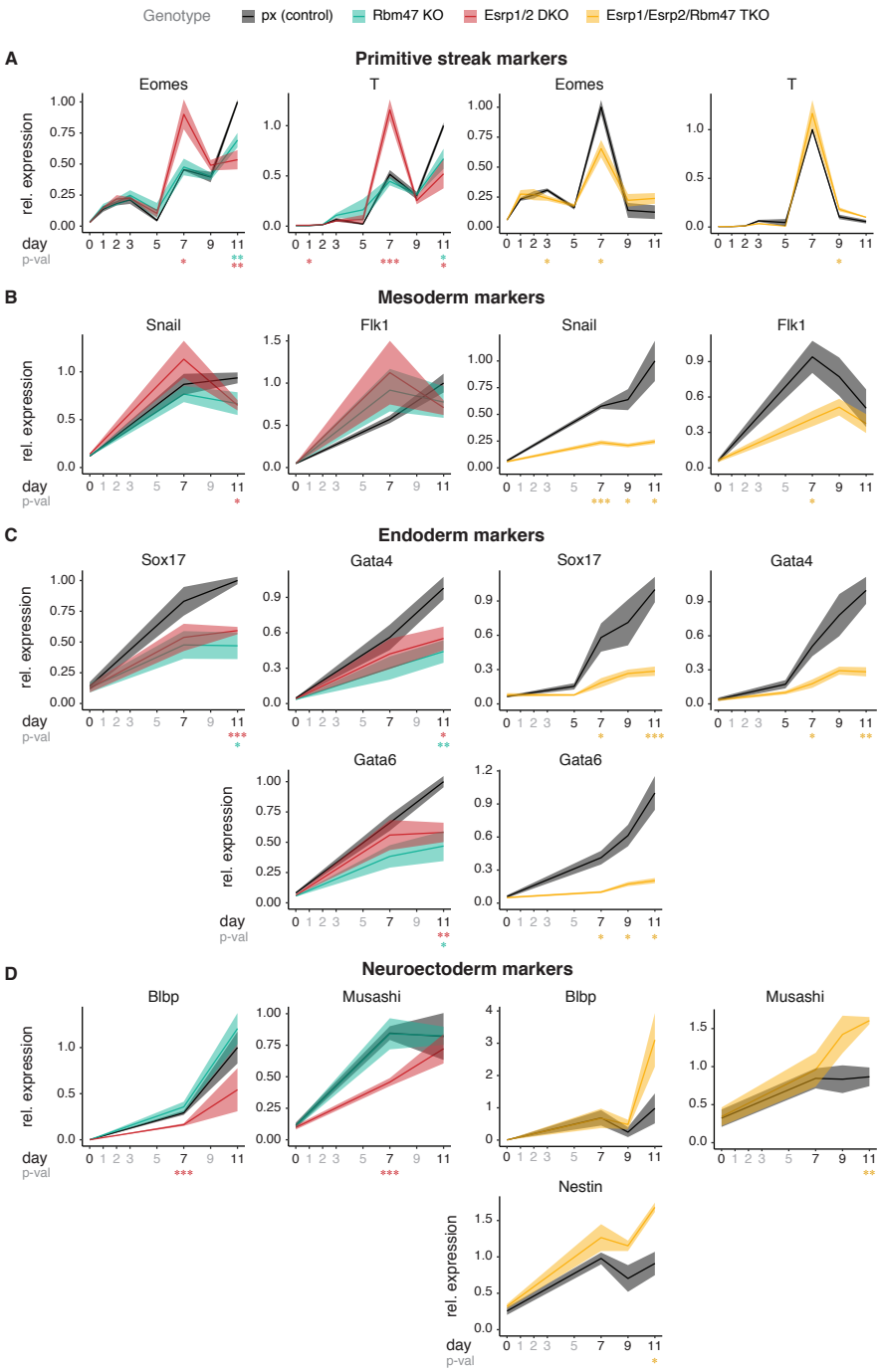
Over the course of EB differentiation, the expression dynamics of naïve and primed pluripotency markers in Rbm47 KOs are similar to controls (Figure 23B). On the other hand, Esrp1 Esrp2 DKOs show a very modest, but significant tendency for a slower downregulation of naïve pluripotency markers. Furthermore, they also show a slower downregulation of the primed pluripotency markers *Otx2* and *Fgf5* but not of *Oct6* or *Dnmt3b*. In fact, *Dnmt3b* expression levels in Esrp1 Esrp2 DKOs are lower than in controls at the maximum peak of its expression, 48hrs after LIF-withdrawal (Figure 23B). This might suggest a role for ESRP1 and/or ESRP2 in DNA methyltransferase regulation. All in all, this is indicative of an atypical primed pluripotency and suggested that the dynamics of entry and exit from primed pluripotency are slightly affected in Esrp1 Esrp2 DKOs but not in Rbm47 KOs.


A
Rbm47
Esrp1
Esrp2
rel. expression
day
p-val
Genotype
px (control)
Rbm47 KO
Esrp1/2 DKO
B
Genotype
px (control)
Rbm47 KO
Esrp1/2 DKO
Klf2
Klf4
Oct4
Rex1
Dnmt3b
Fgf5
Oct6
Otx2
C
Genotype
px (control)
Esrp1/Esrp2/Rbm47 TKO

On the other hand, in Esrp1 Esrp2 Rbm47 TKOs, the dynamics of naïve pluripotency marker expression levels match those in the controls (Figure 23C), suggesting that they do not show differences in the exit from naïve pluripotency. Furthermore, entry and exit dynamics from primed pluripotency are also not significantly affected as the timings of upregulation and downregulation of *Otx2*, *Oct6*, *Fgf5* and *Dnmt3b* gene expression are similar to controls (Figure 23D). However, we did find that *Dnmt3b* expression levels in Esrp1 Esrp2 Rbm47 TKOs tended to be lower than controls at the maximum peak of its expression and although this trend is not significant, it parallels what we saw for Esrp1 Esrp2 DKOs (Figure 23B). Together, these results confirm our previous observation that Esrp1 Esrp2 Rbm47 TKOs do not have a delayed exit from pluripotency.

During EB differentiation, Esrp1 Esrp2 DKOs display a higher peak in the gene expression profiles of primitive streak markers *Brachyury* (*T*) and *Eomes* compared to controls (Figure 24A left). Rbm47 KOs, on the other hand, do not show substantial differences. Furthermore, simultaneous depletion of *Esrp1*, *Esrp2* and *Rbm47* does not change *Brachyury* expression dynamics but does slightly decrease *Eomes* gene expression peak (Figure 24A right). Importantly, both *Brachyury* and *Eomes* are EMT drivers [435–438], therefore together these

Figure 23: Gene expression profiles of splicing factors and naïve and primed markers during embryoid body differentiation in cell lines depleted of the Rbm47/ Esrp regulatory axis. (A) Line charts showing the expression dynamics along the EB differentiation time course for the three candidate splicing factors: *Esrp1*, *Esrp2* and *Rbm47* in control and KO lines. **(B)** Top row, line charts depicting *Oct4* (general pluripotency marker), *Klf4* (early naïve pluripotency marker), *Klf2* and *Rex1* (naïve pluripotency markers) gene expression levels throughout EB differentiation as measured by RTqPCR show expression dynamics of Esrp1 Esrp2 KO, Rbm47 KO and control (px1, px4) cell lines. Bottom row, line charts depicting *Otx2*, *Oct6, Dnmt3b* (formative state, early primed pluripotency markers) and *Fgf5* (primed pluripotency marker) gene expression levels throughout EB differentiation as measured by RTqPCR show expression dynamics of Esrp1 Esrp2 DKO, Rbm47 KO and control cell lines. **(C)** Same as (A) but testing Esrp1 Esrp2 Rbm47 KO cell lines and controls in an independent set of experiments. Expression values are first normalized relative to the expression of Beta-actin (*Actb1*) and further normalized by the average maximum expression in the controls across each EB differentiation experiment. Central lines depict the mean value for two replicates and two clones per genotype and ribbons, the standard error of the mean. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated for each time point: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, coloured by genotype.


Genotype
px (control)
Rbm47 KO
Esrp1/2 DKO
Esrp1/Esrp2/Rbm47 TKO
A
Primitive streak markers
Eomes
T
Eomes
T
B
Mesoderm markers
Snail
Flk1
Snail
Flk1
C
Endoderm markers
Sox17
Gata4
Sox17
Gata4
Gata6
Gata6
D
Neuroectoderm markers
Blbp
Musashi
Blbp
Musashi
Nestin
rel. expression
day
p-val

results might suggest that the right balance between *Rbm47* and *Esrp1 Esrp2* expression is necessary for EMT induction. If *Esrp1* and *Esrp2* are depleted, it results in a stronger EMT induction but when *Rbm47* is depleted concomitantly with *Esrp1* and *Esrp2*, this stronger induction is diminished.

We studied the gene expression of mesodermal, endodermal and neuroectodermal markers to investigate whether depletion of the components of the Rbm47/Esrp axis provided any bias of differentiation induction into any specific lineage. Firstly, Esrp1 Esrp2 DKOs show a clear downregulation of endodermal and neuroectodermal lineage markers (Figure 24C, D left), yet the disparity of mesodermal marker expression across the two Esrp1 Esrp2 DKOs clones and experimental replicates impedes drawing conclusions for the possible impact of *Esrp1* and *Esrp2* depletion on mesodermal differentiation (Figure 24A, B left). Secondly, Rbm47 KOs seem to show a significant downregulation of endodermal induction (Figure 24C left). Finally, simultaneous depletion of *Esrp1*, *Esrp2* and *Rbm47* strongly inhibits mesendodermal induction (Figure 24B, C right) and promotes neuroectoderm differentiation (Figure 24D right). Taken together, this suggests that Rbm47 and Esrp1/Esrp2 coregulate key pathways involved in lineage differentiation.

Figure 24: Gene expression profiles of primitive streak, mesoderm, endoderm and neuroectoderm markers during embryoid body differentiation in cell lines depleted of the Rbm47/Esrp regulatory axis. (A) Line charts depicting primitive streak markers *Brachyury* (labelled as T) and *Eomes* expression levels throughout EB differentiation as measured by RTqPCR. **(B)** Line charts depicting mesodermal markers *Snail* and *Flk1* expression levels throughout EB differentiation as measured by RTqPCR. **(C)** Line charts depicting endodermal markers *Sox17*, *Gata4* and *Gata6* expression levels throughout EB differentiation as measured by RTqPCR. **(D)** Line charts depicting *Blbp* and *Musashi* expression levels throughout EB differentiation as measured by RTqPCR. Expression values are first normalized relative to the expression of Beta-actin (*Actb1*) and further normalized by the average maximum expression in the controls across each EB differentiation experiment. Central lines depict the mean value for two replicates and the two clones per genotype and ribbons, the standard error of the mean. P-values from two-sided Student's t-tests comparing each genotype with controls are indicated for each time point: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, coloured by genotype. Results shown are from two independent experiments performed in duplicate: one done using Esrp1 Esrp2 KO, Rbm47 KO and control cell lines (left) and the other using Esrp1 Esrp2 Rbm47 KO and control cell lines (right).

Part III: Large-scale search for alternative splicing regulators

To find additional regulatory networks that control naïve-to-primed alternative splicing landscape switch, we performed a Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing (SPAR-seq)[439] screen in collaboration with Dr. Hong Han and Dr. Ben Blencowe (University of Toronto). Briefly, SPAR-seq is a multiplex quantitative genomics screen that generates a sequencing output able to comprehensively test the effect of hundreds of trans-acting factors on dozens of alternative splicing events.

We examined a library of 1536 small interfering RNA (siRNA) single knockdowns and controls for their effect on 41 naïve-to-primed alternative splicing events. These comprised 38 cassette alternative exon events and three intron retention events (Figure 25A). These events were selected based not only on their differential splicing between naïve and primed samples but also on the possibility of obtaining robust results when combined in a multiplex PCR, what required several rounds of primer testing and adjustments. Additionally, three gene expression events corresponding to naïve, primed and housekeeping markers (*Klf4*, *Fgf5* and *Gapdh* respectively) were monitored as internal controls. All 44 events were simultaneously amplified by multiplex RT-PCR on RNA extracted from siRNA knockdowns in 96-well plates (Figure 25B). Then, to identify each well, a second PCR was performed to add dual-index barcodes. Finally, the barcoded amplicons were pooled and sequenced.

As discussed in previous sections, SL maintains cells in naïve state but remains permissive for primed pluripotency and occasional differentiation thus it presents an ideal platform for testing the effect of gene depletions on stem cell identity as it is sensitive enough to expose both positive and negative regulators of the naïve alternative splicing network. Accordingly, siRNA-knockdowns were performed on CGR8 mESCs grown in SL culture conditions. Specifically, 1416 genes were targeted and these included 654 splicing- and RNA- associated factors; 858 chromatin and transcription-associated factors; and 65 signalling and post-translational associated factors related to splicing regulation. Additionally, siRNAs targeting 32 positive controls (simultaneous knockdown of *Mbln1* and *Mbln2*) and 88 negative controls (including non-targeting siRNAs, mock controls and untreated samples)[439] were included.

Using a pipeline developed to quantify alternative splicing events from the SPAR-seq read output by employing a manually constructed exon-exon and exon-intron junction library, we were able to calculate the percentage of transcripts with the exon spliced in (PSI) and the percentage of transcripts with intron retained (PIR) values. These values were normalized to correct for variation among plates across the two biological replicates to ensure specificity and reproducibility (see materials and methods).

For each alternative splicing event, we identified candidate regulators that significantly shifted normalized PSI levels in the same direction in both biological replicates (Figure 26A). Then, for each valid event based on read coverage and PSI range (see Methods), we scored the regulators for which both replicates were among the top 50 induced changes in inclusion levels. We focused on those regulators that were identified in at least two or more naïve-to-primed events and that only changed events towards the isoforms found in one pluripotency state or the other. With this, we were able to identify two potential primed regulators whose knockdown drives the expression of 2iL ESCs isoforms and ten potential naïve regulators whose knockdown drives the expression of EpiSC isoforms (Figure 26B). To investigate if these factors could be regulating alternative splicing in the naïve-to-primed transition, we examined their expression levels across naïve and primed samples in three independent stem cell lines (Figure 26C). Interestingly, among all potential regulators, there was only one, Quaking (QKI), whose transcript abundance changes consistently between 2iL ESCs and EpiSCs. All other potential splicing factors do not substantially change the mRNA expression levels between 2iL ESCs and EpiSCs.

QKI is a STAR (signal transduction and activation of RNA) family RBP that has been shown to be key in oligodendrocyte differentiation[440–444], monocyte/macrophage differentiation[445], vascular development[446], myocyte cell function[447,448], EMT[425,449], and cell adhesion signalling[450]. It is involved mainly in alternative splicing regulation[441,443,445,447,451–453] but roles in circular RNA production[449], micro-RNA processing[454,455], mRNA stability[456] and translation[457] have also been described. In the case of alternative splicing regulation, QKI acts as a *trans*-acting factor by binding to QKI responsive elements (QRE), which are mainly found in introns[458]. As many other splicing regulators, it generally favours exon inclusion if bound downstream of the regulated exon, and exon skipping, if bound upstream of the regulated exon[447,450]. Since the SPAR-seq indicated us that QKI was regulating some of our tested events in

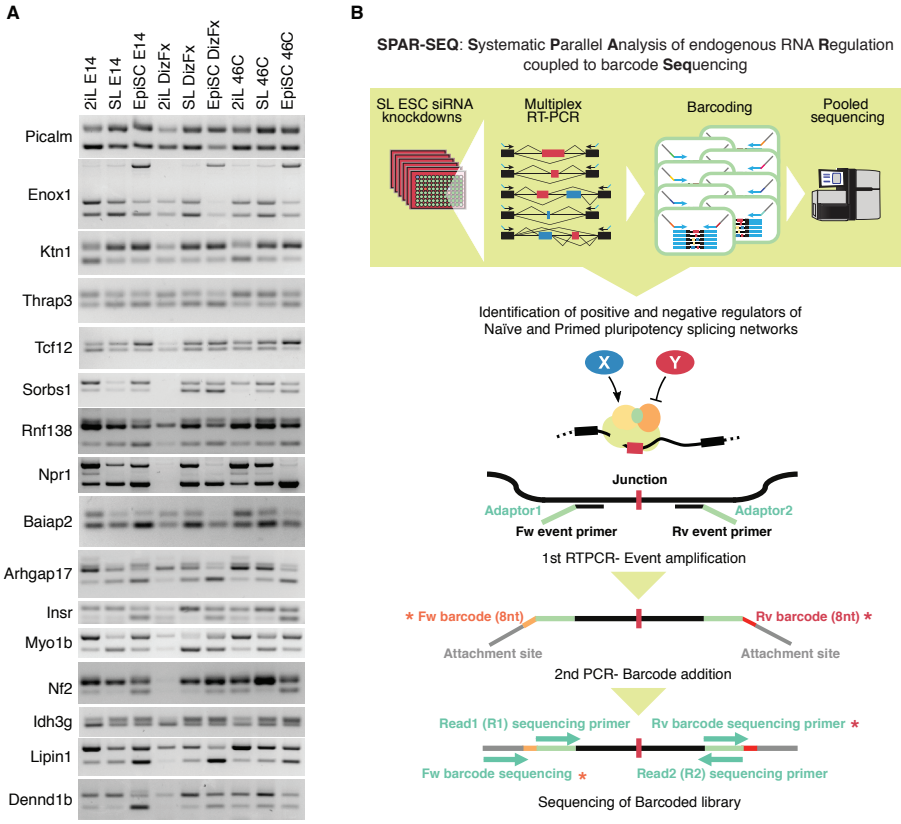


Figure 25: Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing (A) RT-PCR primer validations of naïve-to-primed alternative splicing events across three different cell lines (E14, DizFx and 46C) in three different culture conditions (2iL ESCs, SL ESCs, EpiSC). Please refer to materials and methods section for a complete list of the multiplex primers used in the screen. **(B)** SPAR-seq strategy. A total RNA pool was extracted from 1536 siRNA knockdown and control treatments that were performed with two biological replicates in SL ESCs (CGR8 cell line). In total, 32 96-well plates were used for SPAR-seq. For each well, a multiplex RT-PCR was applied to simultaneously amplify 44 transcript regions which assessed naïve-to-primed alternatively spliced exon inclusion, intron retention and gene expression. In this first RT-PCR, primers with universal adaptor sequences were used. Subsequently, unique dual-index barcodes were added by a second PCR for multiplex barcode sequencing. Barcoded sequencing libraries were pooled and sequenced using Illumina HiSeq2500. A total of four sequencing reads were produced: Read 1 (R1) and Read 2 (R2) were mapped to exon-exon junctions and exon-intron junctions and the two other reads were barcode sequencing reads to identify each unique siRNA sample. Adapted from Han et al. 2017[439].

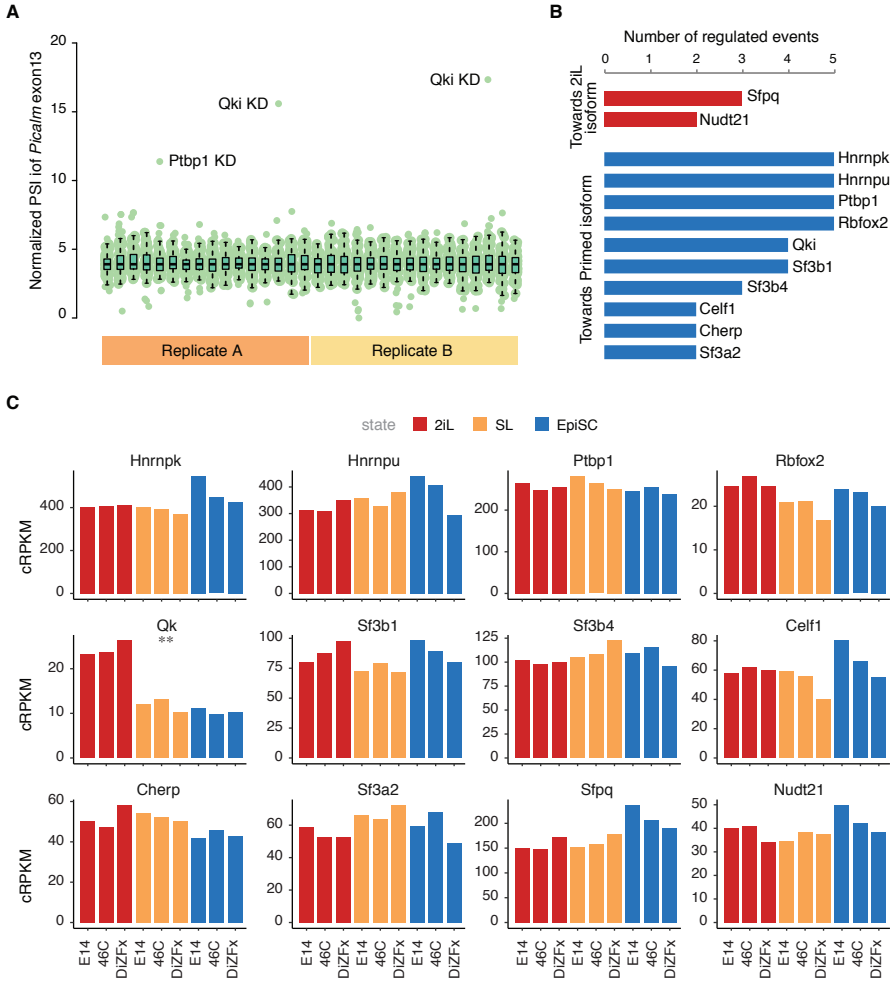


Figure 26: Identification of potential naïve-to-primed alternative splicing regulators using SPAR-seq output (A) Box plot depicting normalized PSI values for exon 13 in *Picalm*, one of the events monitored by SPAR-seq, as an example. Each dot in the plot represents one sample and each box plot corresponds to one 96 well plate. Both biological replicates are shown. The top three samples are marked with their respective siRNA knock-down target showing only *Qki* knock-down reproducibly increases the normalized PSI value of *Picalm* exon 13. KD: knock-down. **(B)** Bar plot of top candidate regulators identified by SPAR-seq depicts siRNA knock-down targets that reproducibly switched normalized PSI values of naïve-to-primed events towards their 2iL ESC (red) or EpiSC (blue) isoforms. For each candidate regulator, the number of regulated events is shown. **(C)** Transcript abundance levels of candidate regulators calculated using RNAseq across three different cell lines (E14, DizFx and 46C) in three

ESCs, we set out to explore the possible regulatory role of QKI in naïve and primed alternative splicing landscape maintenance.

Firstly, we analysed a formerly published *Quaking* knockdown RNA-seq dataset from Hayakawa-Yano et al. performed in primary neural stem cells[450] using *vast-tools* to identify alternative splicing switches. We cross-compared these alternative splicing exon switches with those occurring in the naïve-to-primed transition to find out whether other QKI-regulated exons were differentially regulated in the naïve-to-primed transition with the same direction of change. We found that co-occurring alternative exon switches in naïve-to-primed transition and in neural stem cells after diminishing *Quaking* levels substantially correlated (Figure 27A). Additionally, all types of naïve-to-primed events (primed specific, progressive and 2iL-excluding) had a high degree of consistency with the *Quaking* knockdown (Figure 27A bottom). This suggested a possible regulatory role for QKI in the naïve-to-primed transition. However, although *Quaking* had higher transcript levels in 2iL ESCs, they were very similar in SL ESCs compared to EpiSCs (Figure 26C).

If QKI has a regulatory role in the alternative splicing switch occurring in the naïve-to-primed transition, both naïve culture conditions would have to maintain higher levels of active QKI than in primed samples. At the transcriptional level, SL ESCs and EpiSCs had similar *Quaking* levels (Figure 26C). This hinted at a possible post-transcriptional or post-translational regulation of the QKI protein. Therefore, we first investigated the protein levels in SL ESCs and EpiSCs and found that both had similar QKI protein levels (Figure 27B). Additionally, we did not detect substantial QKI isoform changes in the naïve-to-primed transition. In terms of post-translational modifications, to our knowledge, just one QKI phosphorylation has been described to date. It takes place at QKI's highly conserved c-terminal tyrosine cluster[459]. This phosphorylation is carried out by Src family protein tyrosine kinases and decreases the ability of QKI to bind target RNA *in vivo* and *in vitro*[459]. This goes in line with the impact of tyrosine phosphorylation of other STAR proteins, in which tyrosine phosphorylation lowers the affinity to mRNA targets and

different conditions (2iL ESCs, SL ESCs and EpiSCs) shows *Quaking* (*Qki*) is the only factor that substantially changes among 2iL ESCs and EpiSCs (FDR < 0.01). cRPKMs: Corrected-for-mappability Reads per Kbp and Million mapped reads.

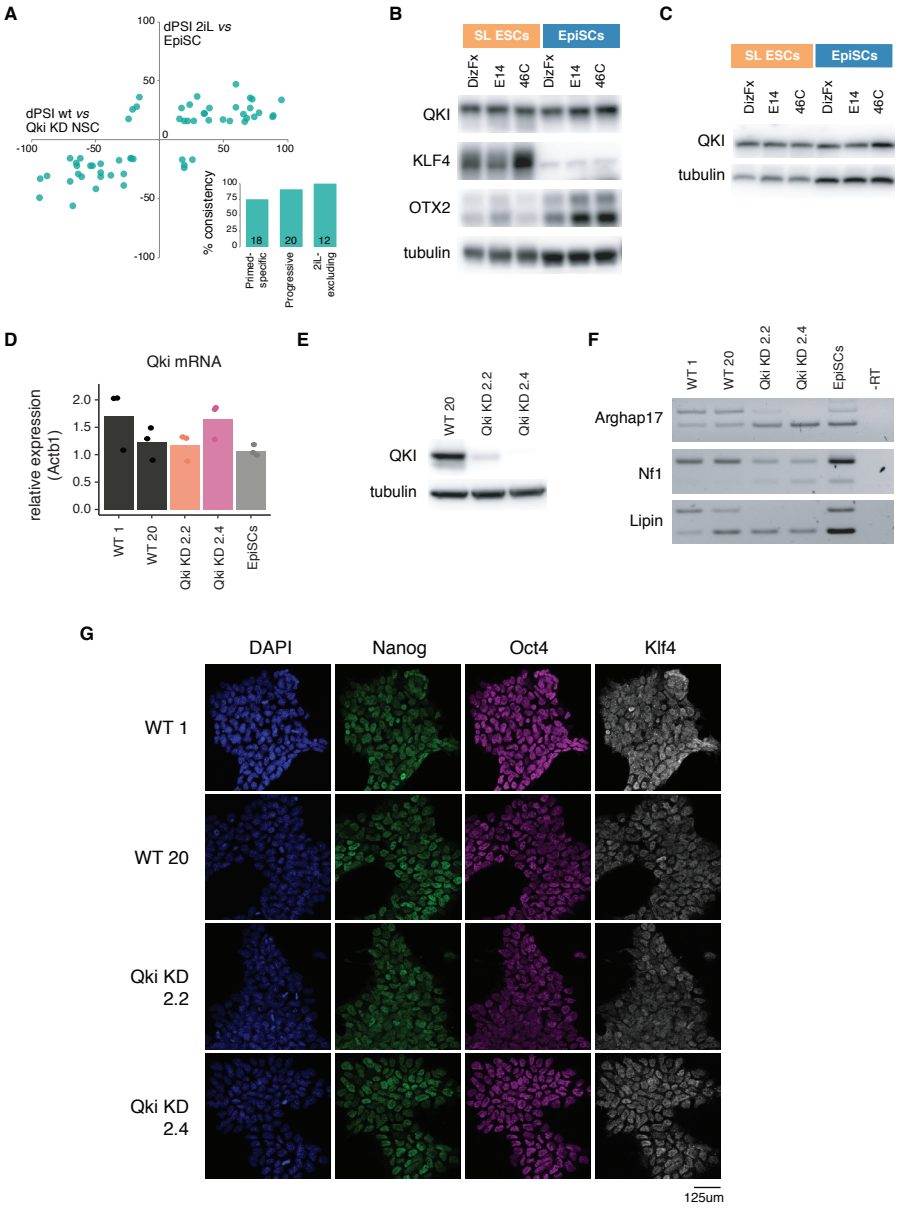


Figure 27: Regulatory role of QKI in naïve pluripotency (A) Scatter plot depicting the cross comparison of inclusion levels of differential naïve-to-primed alternative spliced exons changing upon Quaking shRNA knockdown (Qki KD) in primary neural stem cells from Hayakawa-Yano et al. 2017[450]. NSC: Neural stem cells. dPSI: difference in PSI. On the bottom right, bar plot illustrating the level of consistency of

abolishes the ability to regulate splice site selection[460,461]. Additionally, although it has not been described for QKI, other STAR proteins can have serine/threonine phosphorylations which in turn increase the RNA-binding affinity of the protein[462]. Therefore, we hypothesized that if the QKI phosphorylation levels were different in SL ESCs and EpiSCs, this would potentially explain the difference in the alternative splicing patterns of QKI targets. Using Phos-tag western blots we found that QKI phosphorylation levels are the same in SL ESCs compared to EpiSCs (Figure 27C).

Another possibility for the downregulation of QKI activity at the primed state of pluripotency would be the shortage of a co-factor that could be cooperating in the splice site selection. To our knowledge, only two factors have been proposed to have cooperative regulation of alternative splicing with QKI, these are PTBP1 and RBFOX2[425,447]. We found that the transcript levels of these two factors were not changing in the naïve-to-primed transition suggesting that they were not responsible for QKI's diminished activity in primed cells. Curiously, one study had suggested that QKI alternative splicing regulation opposes that of ESRP1 and RBM47[425]. This study, however, provided no evidence of shared targets between ESRP1 and/or RBM47 with QKI[425]. Considering first, that we found opposing regulatory roles between ESRP1 and RBM47 and secondly, that RBM47 is expressed in 2iL ESCs and increases expression levels in SL ESCs but is completely depleted in EpiSCs; we speculated that RBM47 could be cooperating with QKI in naïve-to-primed alternative splicing regulation.

---

differential naïve-to-primed cassette exon alternative splicing in neural stem cell Qki KD. The number at the base of each bar represents the number of shared differential alternative exon events. **(B)** Western blot showing similar QKI protein levels across SL ESCs and EpiSCs. KLF4 and OTX2 are shown as markers of naïve and primed pluripotency, respectively. **(C)** Phos-tag western blot showing no phosphorylation events occurring in QKI protein across SL ESCs and EpiSCs. **(D)** Bar plot showing *Qki* mRNA expression levels in wild type (WT) and Qki KD lines grown in SL conditions and of EpiSCs as a reference for primed pluripotency levels. Expression values are normalized relative to the expression of Beta-actin (*Actb1*). **(E)** Western blot showing QKI levels in the two *Quaking* knockdown (Qki KD) lines and wild type (WT) ESCs grown in SL condition. **(F)** Gel images showing RT-PCR amplification of differential naïve-to-primed events across across *Quaking* knockdowns (Qki KD), and wild type (WT) ESCs grown in SL condition and EpiSCs reveal strong isoform switches. -RT: minus reverse transcriptase control **(G)** Microscopy images of *Quaking* knockdown (Qki KD) and wild type (WT) ESCs grown in SL condition immunostained for naïve pluripotency markers show comparable heterogeneity levels among colonies.

In collaboration with Dr. Alessandro Dasti and Dr. Elias Bechara from the Tartaglia lab, who provided the mutant cell lines, we set out to explore the impact of QKI depletion on naïve-to-primed transition. We used two cell lines generated by CRISPR/Cas9-mediated targeting of *Quaking* in E14 ESCs: one, clone 2.2, showed a very strong reduction of QKI protein levels and the other, clone 2.4, had an almost complete depletion of QKI protein but none of them downregulated *Quaking* at the transcriptional level (Figure 27D, E). We found both cell lines show a strong switch in inclusion levels in a set of naïve-to-primed alternative exons when grown in SL conditions (Figure 27F). Specifically, the mutants mostly show a change in inclusion levels towards the primed form, however they usually showed more extreme inclusion or exclusion levels compared to EpiSCs (Figure 27F). We then investigated the impact that this had on naïve pluripotency. Interestingly, the *Quaking* knockdown (Qki KD) cell lines seem to have similar levels of naïve marker expression heterogeneity compared to wildtype controls (Figure 27F); however, unexpectedly they show higher naïve marker gene expression levels (Figure 28A).

We performed RNA-seq of these Qki KD ESC cell lines grown in parallel with wildtype ESCs in SL conditions to examine the effect of *Quaking* depletion on the general pluripotency state and alternative splicing landscape. Gene Set Enrichment Analysis (GSEA) shows that Qki KD ESCs are significantly more enriched in the differential gene expression profile of naïve pluripotency

---

comparing each genotype with controls are indicated: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$. **(B)** GSEA depicting enrichment in naïve gene expression profile of differentially expressed genes across *Quaking* depleted SL ESC lines compared to wildtype cell lines shows mutants upregulate naïve pluripotency gene expression profiles. **(C)** GSEA depicting enrichment in primed gene expression profile of differentially expressed genes across *Quaking* depleted SL ESC lines compared to wildtype cell lines shows mutants downregulate primed pluripotency gene expression profiles. Ranked lists of naïve and primed markers were made with the differential gene expression profile defined using paired comparisons of 2iL vs. EpiSCs across our three independent cell lines (E14, 46C, DizFx) and E4.5 vs. E6.5 epiblasts (see materials and methods for further details). **(D)** ASSEA of *Quaking* depleted SL ESC lines compared to wildtype cell lines shows enrichment in the exclusion of naïve highly included (higher in naïve) exons. **(E)** ASSEA of *Quaking* depleted SL ESC lines compared to wildtype cell lines shows enrichment in inclusion of primed highly included (higher in primed) exons. For ASSEA, we used ranked lists of differentially included exons made with the differential PSI profile defined using paired comparisons of 2iL vs. EpiSCs across our three independent cell lines (E14, 46C, DizFx) (see methods for further details). NES: Normalized enrichment score.

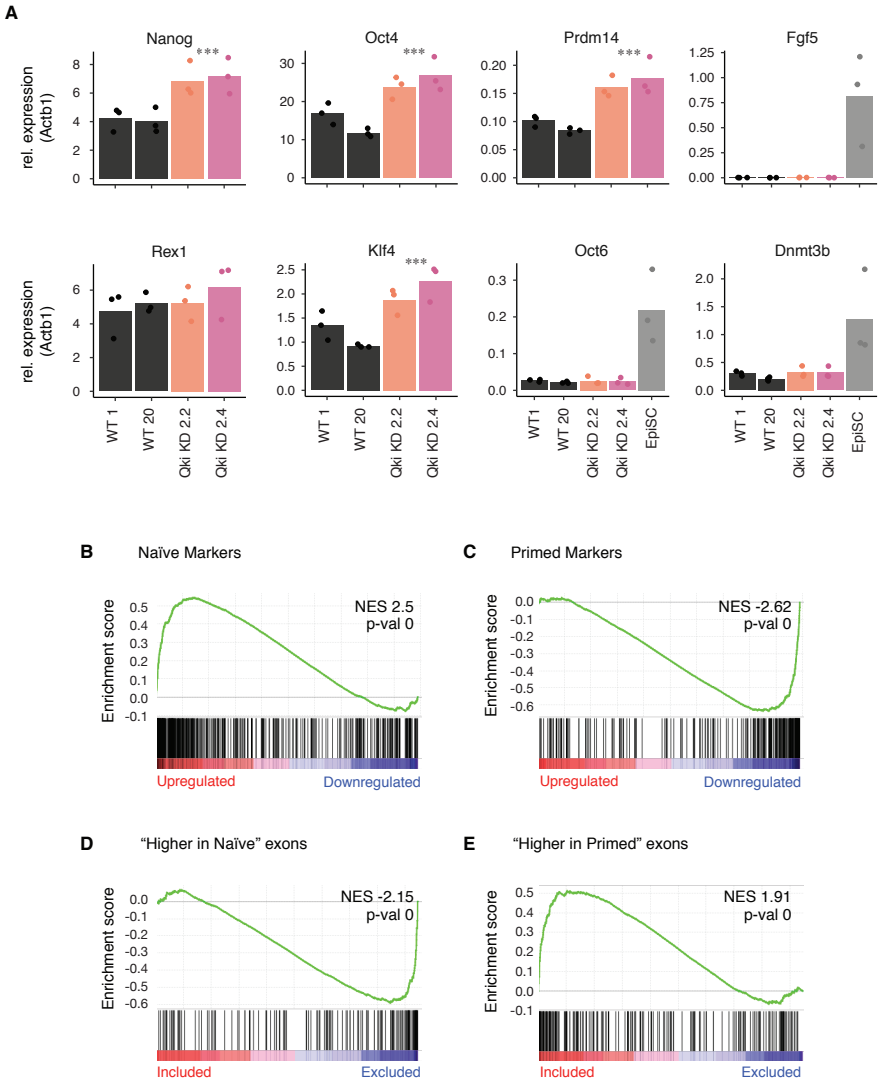


Figure 28: Pluripotent markers together with gene set and alternative splicing set enrichment analysis show that Qki KD upregulate naïve pluripotency and downregulate primed pluripotency gene expression profiles whereas they are significantly biased towards a more primed pluripotency alternative splicing landscape. (A) Bar plots of naïve and primed pluripotency markers depicting gene expression levels in *Quaking* knockdown (Qki KD) and wild type (WT) ESCs grown in SL condition as measured by RTqPCR. Expression values are normalized relative to the expression of Beta-actin (*Actb1*). P-values from two-sided Student's t-tests

compared to wildtype ESCs (Figure 28B). Additionally, they display a significant downregulation of primed pluripotency gene expression profile (Figure 28C). Contrarily, Alternative Splicing Set Enrichment Analysis (ASSEA) shows that Qki KD ESCs are significantly enriched in a more primed pluripotency alternative splicing landscape than wildtype ESCs with higher exclusion of naïve highly included exons (Figure 28D), and higher inclusion of primed highly included exons (Figure 28E).

Together, these results strongly suggested that QKI control over the regulatory layers of gene expression and alternative splicing is antagonistic. On one hand, at the level of gene expression, it supports naïve pluripotency, on the other hand, at the level of alternative splicing, it supports primed pluripotent alternative splicing landscape. This could imply a complex relationship between the alternative splicing profile that is controlled by QKI and the pluripotent state. Additionally, the strong depletion of *Quaking* levels could have artefactual consequences on the transcriptomic state of the *Quaking* mutant cells. Subsequently we wondered whether the misregulation of QKI and of QKI-regulated alternative splicing events had an effect on the exit from pluripotency. Firstly, to test the influence of *Quaking* depletion on cell differentiation we used clonal monolayer differentiation assays. Qki KD ESCs show a delayed exit from naïve pluripotency as after four days of LIF removal, *Quaking* depleted cells retain a greater proportion of mildly AP stained colonies compared to their wildtype counterparts (Figure 29). This goes in line with the enriched naïve pluripotency differential gene expression profile that we had previously observed (Figure 27I, 28A).

Secondly, to better pinpoint the dynamics of exit from naïve pluripotency, entry and exit from primed pluripotency as well as to identify any biases towards specific lineages upon *Quaking* depletion, we performed EB assays. Importantly, although we were able to perform EB differentiation assays of all clones, the Qki KD clone 2.2 notably starts having a high cell death rate by day five of differentiation. This is, roughly when primed pluripotency is exited. Curiously, we did not detect this for the other, more severe Qki KD clone.

We found that during EB differentiation *Quaking* depleted cells downregulate *Oct4* more rapidly than wildtype cells (Figure 31A). However, by day three of EB differentiation *Oct4* expression levels in mutants closely match those in wildtypes. Additionally, the gene expression profiles of naïve pluripotency markers *Klf4* and *Rex1* in mutants also closely match those in wildtypes during

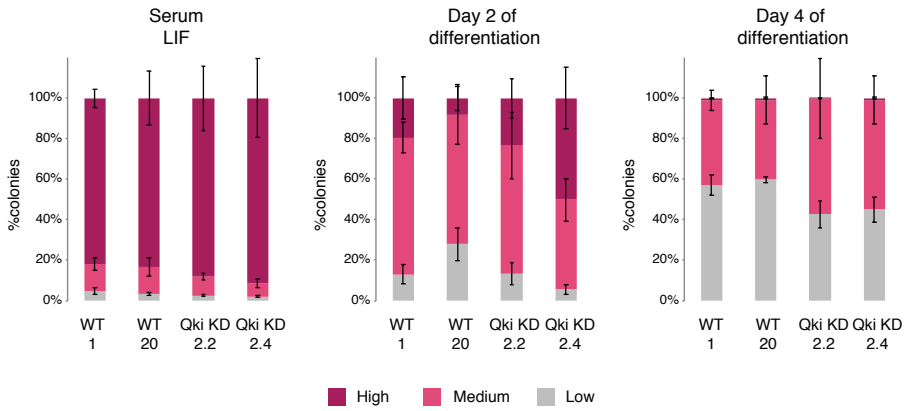
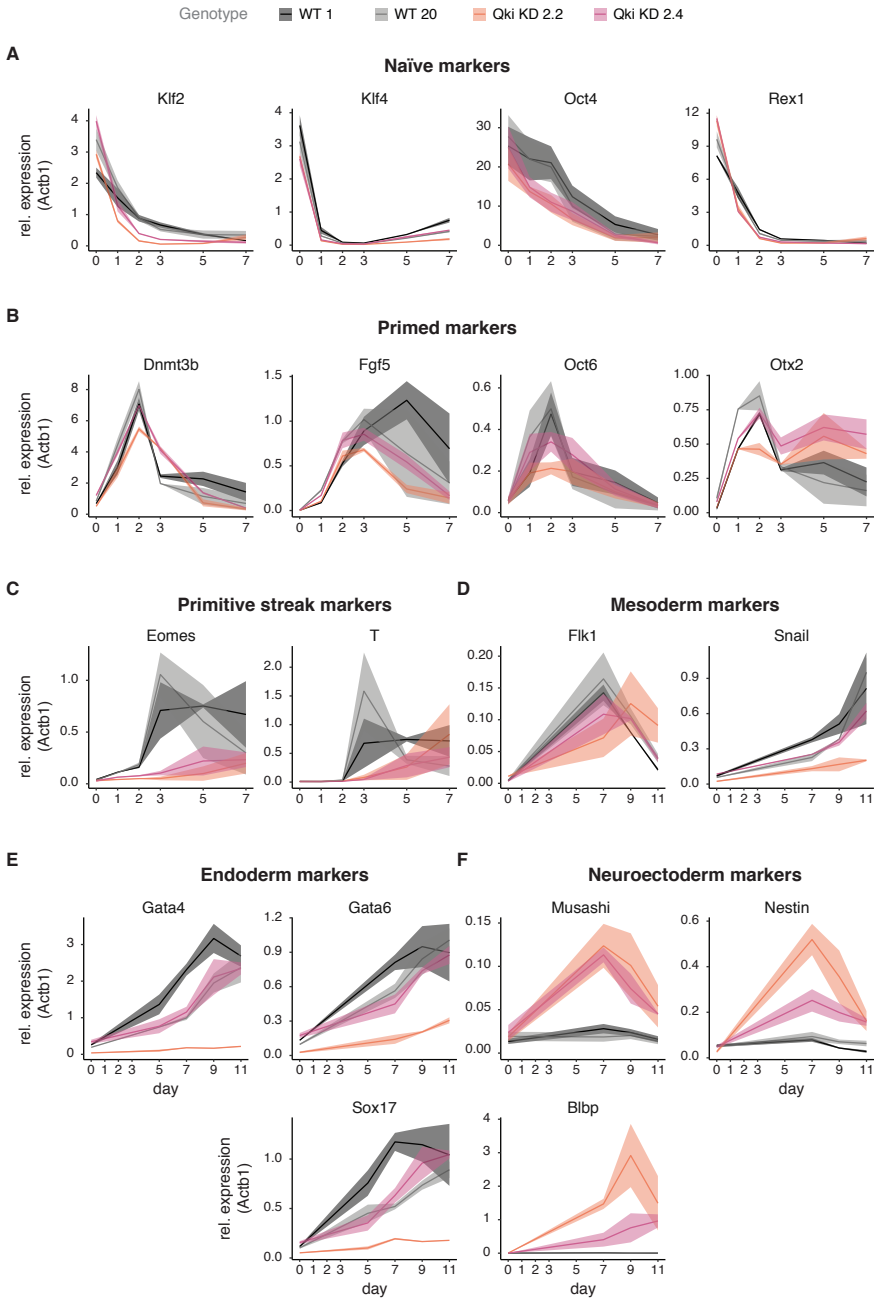


Figure 29: Pluripotency decrease assessed by AP staining during LIF removal clonal differentiation assay in ESCs depleted of Qki. ESCs were plated in SL at clonal density, LIF was removed for two or four days after which colonies were stained with AP staining. Colonies were counted and classified according to their level of AP staining (high, medium or low). Stacked plots depict the population distribution of stained colonies and shows a delay in loss of pluripotency for Qki KDs. A minimum of 120 colonies were counted per condition per clone. n=3.

the whole differentiation. Nonetheless, *Klf2* expression profiles show faster downregulation in *Quaking* depleted cells. Using primed pluripotency marker expression profiles, we found that *Quaking* depleted cells had similar entry dynamics into primed pluripotency compared to wildtype cell lines (Figure 31B). However, in the case of primed pluripotency exit dynamics it was less clear as *Otx2* and *Dnmt3b* showed a delay in *Quaking* depleted cells but *Oct6* and *Fgf5* did not. Additionally, *Otx2* expression levels did not decrease after day three. Although OTX2 has been shown to stabilize the EpiSC state[463,464], the fact that the other three primed markers do show downregulation of the primed pluripotency network at days five and seven suggests that *Otx2* expression stabilization, in this case, might have more to do with its regulatory role in neuroectoderm specification.

During EB differentiation, Qki KD cells display a prominent delay in the gene expression profile of primitive streak markers *Brachyury* (*T*) and *Eomes* compared to wildtype cells (Figure 31C). Additionally, this induction peak is also very mild in comparison. This observation could go in line with the alleged role of QKI in driving EMT processes[425,449]. Furthermore, we investigated


Genotype
WT 1
WT 20
Qki KD 2.2
Qki KD 2.4
A
Naïve markers
Klf2
Klf4
Oct4
Rex1
rel. expression (Actb1)
B
Primed markers
Dnmt3b
Fgf5
Oct6
Otx2
C
Primitive streak markers
Eomes
T
D
Mesoderm markers
Flk1
Snail
E
Endoderm markers
Gata4
Gata6
Sox17
F
Neuroectoderm markers
Musashi
Nestin
Blbp
day

whether *Quaking* depletion provided any bias of differentiation induction into any specific lineage. Curiously, we found Qki KD clones show different gene expression profiles among them, with the Qki KD clone 2.2 always showing a more severe phenotype.

On one hand, Qki KD clone 2.2 shows an extremely modest induction of *Snail* gene expression compared to wildtype cells while Qki KD clone 2.4 show an expression profile close to wildtype cells (Figure 31D). This could suggest a reduction in lateral plate mesoderm differentiation, in Qki clone 2.2. Furthermore, another lateral plate mesoderm marker, *Flk1*, showed a similar trend, albeit less extreme than *Snail*. Also, in this case, Qki KD clone 2.4 has similar *Flk1* expression profiles to wildtype cells while Qki KD clone 2.2 showed a milder induction. Secondly, endoderm differentiation is clearly inhibited in Qki clone 2.2 while this is not the case for Qki KD clone 2.4 that shows an induction pattern similar to wildtype cells (Figure 31E). Finally, both Qki KD clones show a notable induction of the neuroectoderm lineage (Figure 31F). However, again the phenotype is significantly more extreme in Qki KD clone 2.2. Taken together, these results might suggest not only that QKI is a key regulator involved in lineage differentiation, as was previously known, but also that these regulatory functions are tightly dependent on QKI expression levels. However, we must be cautious in our interpretation as these results might be

Figure 30: Gene expression profiles of naïve, primed, primitive streak, mesoderm, endoderm and neuroectoderm markers during embryoid body differentiation in Qki depleted cell lines. (A) Line charts depicting *Oct4* (general pluripotency marker), *Klf4* (early naïve pluripotency marker), *Klf2* and *Rex1* (naïve pluripotency markers) gene expression levels throughout EB differentiation as measured by RTqPCR. **(B)** Line charts depicting *Otx2*, *Oct6*, *Dnmt3b* (formative state, early primed pluripotency markers) and *Fgf5* (primed pluripotency marker) gene expression levels throughout EB differentiation as measured by RTqPCR **(C)** Line charts depicting primitive streak markers *Brachyury* (T) and *Eomes* expression levels throughout EB differentiation as measured by RTqPCR. **(D)** Line charts depicting mesodermal markers *Snail* and *Flk1* expression levels throughout EB differentiation as measured by RTqPCR. **(E)** Line charts depicting endodermal markers *Sox17*, *Gata4* and *Gata6* expression levels throughout EB differentiation as measured by RTqPCR. **(F)** Line charts depicting *Blbp*, *Nestin* and *Musashi* expression levels throughout EB differentiation as measured by RTqPCR show differential expression dynamics of Qki KD and wildtype cell lines. Experiments performed in duplicate. Expression values are normalized relative to the expression of Beta-actin (*Actb1*). Central lines indicate the average of two experiments and the ribbons mark the values for each individual replicate.

affected by the high level of cell death upon EB differentiation that the Qki KD clone 2.2 cell line possess. Additionally, further experiments would be required to elucidate whether the differences in differentiation biases we see here are solely to do with QKI levels as opposed to other genetic variations in the two Qki KD clones.

Discussion

Regulation of naïve and primed pluripotency states by Rbmx family proteins

The complex family of Rbmx in mammals

In mouse, the Rbmx family is composed of four members: RBMX, RBMY, RBMXL1 and RBMXL2. *Rbmx* and *Rbmy* are thought to be amongst the earliest sex chromosomes genes, deriving around 130 million years ago from the ancient X-Y proto pair (before the divergence of eutherian and metatherian)[465]. While *Rbmx* is described as being ubiquitously expressed and subject to X-inactivation[466], *Rbmy* remains testis-specific implying an important evolutionarily conserved role in male fertility[467–469] (it is estimated that only 3% of ancestral genes survived decay in the Y-chromosome[470]). Mouse RBMX and RBMY show 44% overall identity (58% similarity), this increments to 60% identity (and 80% similarity) within the RNA recognition motif (RRM). The other two family members, *RbmxL1* and *RbmxL2*, are retrocopies of *Rbmx*. *RbmxL1* and *RbmxL2* have 96% and 75% amino acid identity with *Rbmx*, respectively (97% and 83% similarity). Just within the RRM, identity increments to 99% and 87% (100% and 92% similarity). We have shown that despite the high similarity amongst RBMX, RBMXL1 and RBMXL2, these family members have contrasting effects in the definition of naïve and primed pluripotency identities.

No study to date has ever reported divergent functionality amongst *Rbmx* and its retrocopies however it is particularly compelling to ponder on the evolutionary aspects of the shared and divergent functionality in Rbmx family members as retrogene conservation is indicative of function[471]. Most retrocopies are "dead on arrival" and degenerate rapidly without functional selection, however, retrogenes can be conserved if they are advantageous. Importantly, while *RbmxL1* seems to have generated from a more recent retrotransposition, *RbmxL2* evolved from a retrotransposition that took place about 65 million years ago and *RbmxL2* is found in all placental mammals[423]. This conservation highlights functional importance; however, it was not until very recently that the function of RBMXL2 was assessed in a study. In this study, Ehrmann et al found that *RbmxL2* deletion blocks spermatogenesis, where it is essential for meiosis[423].

It is widely known that testis possess a vast transcriptomic complexity mainly generated by high levels of transcription, relaxed chromatin and altered expression of RNA-binding proteins[312,316,472–478]. Up to 90% of all protein coding genes are

expressed in a single sperm cell by the end of its maturation[478] so maintaining a robust differentiation program through a strong post-transcriptional regulation is essential for fertility. RBMXL2 was suggested to play an important role towards it by protecting the meiotic transcriptome from cryptic splice site poisoning and aberrant splice site usage[423]. *RbmxL2* has been repeatedly described as being exclusively expressed in testis, specifically during the meiotic and post-meiotic phase[417,423,479–481] albeit we have shown here that this is not the case as it is also expressed during naïve pluripotency. Indeed, several autosomal retrogenes originate from X-linked progenitor genes and are expressed in testis[482]. It has been commonly hypothesized that conservation of these testis-specific autosomal retrogenes is driven by the functional compensation they might accomplish during silencing of the X chromosome in male meiosis. In the case of the RBMX family of proteins, *Rbmx* and *Rbmy* transcription is inactivated during male meiosis due to the segregation of sex chromosomes to the "XY body"[482] and *RbmxL2* expression was proposed to provide a like-for-like replacement[417,418,423,479,483]. The results presented in this thesis suggest that independently from being important in spermatogenesis, *RbmxL2* expression might have been co-opted for the regulation of naïve pluripotency. Secondly, we show that the expression of *Rbmx* and *RbmxL2* can have antagonistic effects on pluripotency states, therefore we suggest that RBMXL2 does not simply substitute RBMX as has been repeatedly implied.

Finally, it is important to add here that *Rbmx* has been found to be essential in human by three independent studies that screened for essential genes[484–486], however, this is not the case in mouse[487]. Also, in mouse, *RbmxL2* is not essential but leads to male infertility[423]. In line with this, we found that all the ESC knock-out cell lines we generated were viable although *Rbmx RbmxL1* depletion led to high cell death upon differentiation. Assessing the viability and impact on differentiation of single depletion of *Rbmx* or *RbmxL1* independently, would shed light on the roles of each family component. In any case, our results suggest that in mouse RBMX and its retrocopies may compensate for depletion of each other. Three evidences are in line with this hypothesis. First, that upon *Rbmx* depletion, expression of RbmxL2 is enhanced and vice versa. Secondly, that *Rbmx RbmxL2* depleted mutants have a more severe phenotype than *RbmxL2* depleted mutants. Thirdly, that despite our efforts and clear evidence of good CRISPR-Cas9 gRNA efficiency for *Rbmx* and *RbmxL1* co-occurring deletion, we were not able to generate a *Rbmx RbmxL1 RbmxL2* triple-knockout cell line, hinting that it is not viable.

Opposing and complementary roles for Rbmx and RbmxL2 in the Naïve-to-primed transition

We found that RBMXL2 is the splicing factor with the biggest change in expression levels in the naïve-to-primed pluripotency transition. This, together with the observation that RBMX expression levels go in the opposite direction, we hypothesized that RBMX functionally replaced RBMXL2. Here, we report that *RbmxL2* depletion pushes cells towards a more primed pluripotent state. These cells not only exit naïve pluripotency faster but also exit primed pluripotency sooner and enter differentiation before their wild type counterparts. This suggests that *RbmxL2* might be playing a role in safeguarding the naïve pluripotency network from differentiation cues. Additionally, in differentiation assays of *RbmxL2* depleted cells, we observed that the wave of activation of primitive streak markers had a bigger amplitude and less duration than control cells. This could hint at a larger degree of synchronicity among *RbmxL2* depleted cells at the onset of differentiation.

RbmxL2 depletion causes a switch toward a more primed alternative splicing network consisting on both exclusion of primed pluripotency excluded exons and inclusion of primed included exons. This suggests that the molecular function of RBMXL2 in naïve pluripotency might be regulation of alternative splicing. Furthermore, it could be that the general transcriptional change towards primed pluripotency in *RbmxL2* depleted cells is kindled by this alternative splicing landscape switch towards primed pluripotency isoforms. However, a note of caution is due here since in this thesis we did not provide evidence for a direct function of RBMXL2 in alternative splicing regulation, and thus, this output could be product of an indirect effect. Despite all known functions of RBMXL2 are alternative splicing-related[417,423], RBMX also has non-splicing related functions that could indirectly impact the alternative splicing landscape such as involvement in DNA-damage response[488], transcriptional regulation[489,490] and heterochromatin formation[491]. Considering the high sequence similarity that these two proteins share, it is possible that RBMXL2 might share some of these functions or have other yet unknown functions that could also explain the alternative splicing switch that we detect upon *RbmxL2* depletion towards a primed pluripotency network.

We found that there are two RBMX isoforms whose expression is differentially regulated in the naïve-to-primed pluripotency state transition in both human and mouse. Specifically, these isoforms differ in the 3'end of the *Rbmx* transcript due to the alternative splicing inclusion of an 1253bp exon that codes for the last 102 amino acids of the protein and contains the 3'UTR of the transcript. The alternative transcript

termination contains a stop codon and an alternative 3'UTR. Therefore, this alternative transcript could potentially generate a shorter protein isoform and have a different post-transcriptional regulation due to its contrasting 3'UTR sequence. However, this alternative splicing event has only been reported in one study, which was in fact when *Rbmx* was first mapped to the genome[492]. It was largely disregarded ever since and the effect of inclusion or exclusion of this exon and its function have never been investigated. Interestingly, this region of RBMX is a low complexity region that contains an RGG box, is highly conserved, is necessary and sufficient to bind RNA[493], can bind m^6A-methylated RNAs[421] and is responsible for the interaction of RBMX with Transformer 2 beta homolog (TRA2B1)[419] and the phosphorylated C-terminal of RNA polymerase II[422]. Additionally, a frameshift mutation in this terminal region of RBMX causes an intellectual disability syndrome[494]. All this suggests an enormous potential for a regulatory role in the alternative inclusion of this exon.

We found that the two *Rbmx* transcript isoforms were expressed in naïve pluripotency and this switched to almost exclusively expressing the long isoform in primed pluripotency. Depletion of this long isoform in SL ESCs led to an enriched naïve state expression profile and downregulated primed state gene expression compared to controls. However, in terms of alternative splicing landscape, depletion of the long isoform prompted inclusion of primed highly included exons. This is contrary to what we expected and could suggest that the main role of the c-terminal region of Rbmx in pluripotency state regulation might not be through its function as an alternative splicing regulator as we will discuss later. The alternative splicing profile of *Rbmx* exon mutants rendered them to cluster separately from whole *Rbmx* gene deletion mutants and of other family members (*RbmxL1* and *RbmxL2*), hinting that the alternative splicing regulatory role of the long isoform might be different than that of the short isoform. On the other hand, *Rbmx* exon depletions and all *Rbmx* depletions showed an enrichment towards more inclusion of primed highly included exons compared to controls, suggesting that RBMX mainly promotes primed highly included exon exclusion.

We detected that as cells exit naïve pluripotency and differentiation progresses the *Rbmx* terminal exon is more included. Interestingly, this coincides with an increase in *Rbmx* transcript levels. The fact that we didn't detect that increase in *Rbmx* exon mutants could suggest that the inclusion of this exon has a regulatory role on *Rbmx* expression. One possibility could be that the alternative 3'UTR of the alternative termination bears sequences such as miRNA target sequences that render less stability to the transcript. Another possibility is that the inclusion of the short isoform triggers nonsense mediated decay of the transcript. The stop codon in this transcript is 84bp

away from the exon-exon junction which complies with the nonsense mediated decay 55nt-rule that states that when the distance between stop codons and the exon-exon junction closest 3' end is over 55nts in length, this transcript triggers the nonsense mediated decay pathway and is degraded[495].

One unanticipated finding was that *RbmxL2* depletion induced the inclusion of *Rbmx* alternative exon. Indeed, it could be that this effect is indirect and just a consequence of the tendency towards primed pluripotency that these mutants have. However, it could also suggest that RBMXL2 promotes the skipping of this event as the cross-regulation of splicing regulators through alternative splicing of transcripts encoding for other splicing regulators is a common theme in alternative splicing networks[496]. This finding would be consistent with the fact that upon *RbmxL2* depletion in ESCs, *Rbmx* expression levels are higher. Also, upon *Rbmx RbmxL1* depletion, *RbmxL2* expression increases. Together these results suggest a cross-regulation among Rbmx family members in naïve and primed pluripotency.

Human and mouse RBMXL1 and other *Rbmx* retrocopies have been formerly described as conceptual proteins as the open reading frame was maintained and their translation to protein was possible but not confirmed[466]. The predicted amino-acid sequence of RBMXL1 mainly differs in three amino acids, these are one arginine and two glycines in the first RGG box of RBMX which are deleted in RBMXL1. Thus, antibody specificity discriminating the two proteins is quite dubious and, to our knowledge, has not been achieved to date. Despite this, most functional studies of mouse RBMX completely disregard the possible expression of *RbmxL1*. We scanned the literature for evidence for translation of *RbmxL1* transcripts in pluripotency and found a proteomic study where RBMXL1 unique peptides were identified[497]. Using data from this study, we observed that RBMXL1 was expressed in a constant manner in naïve pluripotent stem cells and during a 2-day differentiation to primed pluripotency time course. In this thesis, we tested the combined depletion of *Rbmx* and *RbmxL1* in ESCs and found that it resulted in the downregulation of both naïve and primed gene expression patterns, suggesting that cells acquired another fate. *Rbmx RbmxL1* depleted mutants also had increased cell death and clear defects in endomesodermal gene expression during differentiation.

Interestingly, we found that the phenotypes among the two *Rbmx RbmxL1* depletion mutants differed. More specifically, the flawed differentiation phenotype is milder in the mutant with a smaller deletion in *RbmxL1* (1bp insertion plus a 3bp deletion in the coding sequence vs. a 566bp deletion, a 4bp deletion and 2 substitutions in the coding sequence). In this line, we also observed that the two *Rbmx RbmxL2* depletion mutants

also differed in the severity of their phenotypes in the same direction, albeit in a more modest manner. Again, the mutant with a smaller deletion, this time in the *Rbmx* gene (2905bp vs. 4299bp), had a milder phenotype towards primed pluripotency gene expression. This, of course, could be due to arbitrary causes like intrinsic clone variability or off-target effects but it could also suggest genetic compensation through transcriptional adaptation[498], a process by which a mutation enhances the expression of related genes that can compensate for the expression loss of the mutated gene. In line with this, a recent study showed that premature termination codon-containing mRNAs can lead to transcriptional adaptation in a process dependent on mutant mRNA decay[499]. This finding might therefore imply that bigger deletions may trigger less compensatory effects in our mutant cell lines. To validate this, a bigger number of mutant clones should be analysed in the context of naïve and primed transition and entry into differentiation. Additionally, single gene deletion mutants would be necessary to further study genetic compensation among the Rbmx family, therefore *Rbmx* knockout cell lines and *RbmxL1* knockout cell lines would still need to be generated.

We found that *RbmxL2* depletion had antagonistic effects on naïve and primed gene expression profiles compared to *Rbmx* terminal exon depletion. This confirmed our hypothesis for a functional replacement of RBMXL2 by RBMX in the naïve-to-primed transition. However, the double *Rbmx RbmxL2* depletion mutants slightly augmented the enrichment in primed pluripotency gene expression profile and increased the downregulation of naïve pluripotency gene expression that we detected in *RbmxL2* depletion mutants. This finding suggests that RBMX can slightly compensate for RBMXL2 deficiency and implies that RBMXL2 and RBMX share some roles in naïve pluripotency maintenance. *RbmxL2* single depletion mutants and *Rbmx RbmxL2* depletion mutants show similar naïve-to-primed transition dynamics, pluripotency exit dynamics and differentiation dynamics overall. However, we found that *RbmxL2* single depletion mutants had a strong bias towards endomesodermal fate during differentiation and that this bias was not as strong in *Rbmx RbmxL2* depletion mutants. This could indicate that a certain equilibrium of RBMX-RBMXL2 is needed for balanced differentiation. When this equilibrium is broken by RBMXL2 depletion, cells are biased to an endomesodermal fate, but when both players, RBMX and RBMXL2, are absent then the fate equilibrium is reached again. This phenomenon is particularly intriguing as we saw that *RbmxL2* is not expressed at the onset of differentiation.

Proteins in the Rbmx family are highly modular proteins with multifunctional roles. One of their most significant features is that they contain a combination of multiple RNA binding domains. This modular domain structure is frequently seen

in RNA binding proteins and can contribute to RNA binding affinity and specificity, organization of RNA topology, spacing between binding sites, promote protein-protein interactions or binding to several RNA molecules[500]. So far, the role that each domain plays in alternative splicing regulation has been characterized using only a handful of minigenes[417,419,420,423,424,482,501,502] and more work still needs to be done specially to clarify divergence among the different family members. Precisely, it would be interesting to understand how *Rbmx* and *RbmxL2* can have such contrasting effects in pluripotency while sharing a high degree of identity. Still, we can try to interpret the origin of these differences by focusing on previously published observations and by pinpointing potentially relevant amino acid substitutions or deletions among the domains. Importantly, RBMX, RBMXL2 and RBMXL1 all have the same modular structure. They have an N-terminus with an RNA recognition motif (RRM) containing two conserved ribonucleoprotein domains (RNP1 and RNP2), and a long low complexity C-terminus containing a nascent transcript targeting domain (NTD) and two RGG boxes.

The RRM motif renders RBMX specificity to two consecutive adenines in single stranded RNA[424] but also to 5'-CCC-3' and 5'-CCA-3' motifs[418] albeit with less affinity. RBMX RRM does not show any specific binding preference to hairpin structures[418]. Also, the RRM of RBMXL2 was reported to preferentially interact with 5'-GUU-3' containing single stranded RNAs, but this binding preference was tested using a limited set of sequences and could potentially be biased[417]. The sequences of RNP1 and RNP2 domains are identical in RBMX and RBMXL2 and both predictably maintain the canonical RRM $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold structure but their RRMs also contain various differences that could account for the differing binding preferences of the two proteins. While the structure of RBMX's RRM bound to RNA has been resolved by NMR, we still lack the structural insight of RBMXL2's RRM. Specifically for RBMX, the RNA lies as a single strand on its β-sheet surface while the C-terminal region of the RRM interacts directly with the two consecutive adenines in the RNA[424]. Notably, in this C-terminal region there two key amino acid differences among RBMX and RBMXL2 that may explain their RNA binding affinities, these are in a glutamic acid and a serine at positions 82 and 88, respectively. In RBMX, the first consecutive adenine in the bound RNA is specifically recognized by three hydrogen bonds formed with the side chains of Lys80 and Glu82 and the backbone amide of Thr85[424]. In RBMXL2, this Glu82 was substituted by an alanine and the hydrogen bond with the adenine cannot form. Another amino acid substitution in RBMXL2 predictably stops a hydrogen bond from forming, in this case with the second consecutive adenine in the bound RNA. When RBMX RRM binds, a hydrogen bond is formed between the phosphate group of the second adenine and the backbone amide of Ser88[424]. Importantly, this contact is not involved in the recognition of the adenine as the previous case, this is recognized

by Lys9 and Thr85 which are conserved, but instead stabilizes the RNA-protein contact which likely contributes to enhancing the binding affinity of RBMX[424]. Predictably this might not occur in RBMXL2 as this hydrophilic serine, is substituted by a hydrophobic valine which likely reduces the binding affinity of RBMXL2.

Motif analysis of transcriptome-wide RBMX binding sites by crosslinking and immunoprecipitation (CLIP) of in vivo RNA did not render the same motifs as those in the studies mentioned above which evaluated binding affinities specifically for the RRM. Actually using CLIP, AGAAC/AGGAC was the most enriched binding motif[421]. Another study using in-vitro RNA-protein binding assays concluded that RBMX could bind very efficiently but non-specifically to RNA[419], however, it could be that it was tested on biased sequences and these experiments should be repeated with appropriate controls. Notably, this disparity in reported binding motifs also occurs for RBMXL2 as a recent study using CLIP reported RBMXL2 binds AA dinucleotides[423]. This goes in line with the notion that the RRM is not the only domain in RBMX that binds RNA, indeed the last 57 amino acids in the c-terminus can also bind RNA[421,493]. This, however, has never been tested for RBMXL2. Additionally, even if it is eventually shown that the RRM is not necessary for RBMXL2 to bind RNA either, RRM binding would predictably help stabilize possible protein-RNA interactions. Thus, understanding disparities between the RRM binding affinities of RBMXL2 and RBMX would be key to understand how these proteins modulate interactions with distinct targets in naïve and primed pluripotency. Moreover, a recent study has found that the RRM of RBMX is important in maintaining a balanced expression of genes related to epithelial cell development[422] which hints at a possible role in primed pluripotency. An important factor to note here is that the RRM of both RBMX and RBMXL2 have been shown to be superfluous for RBMX and RBMXL2 splicing activity in a number of splicing events[417,418,420,423,424,482,502] but not for others[422]. This is relevant first because it is an indication that the binding and recognition of target sequences by the RRM might not be necessary for splice-site selection while other domains might be essential and secondly because it indicates that RBMX and RBMXL2 might affect splicing indirectly and not by binding with RNA.

Interestingly, a few studies have reported the association of RBMX with nascent transcripts[422,493,503]. RBMX associates with most transcription units in the oocytes of the newt *Pleurodeles waltl*[503] and in *Xenopus laevis*[493] and this recruitment could be dependent on active transcription as RBMX re-localizes to nucleolar caps after inhibition of transcription[501]. The nascent transcript targeting domain (NTD), a 50 amino acid span in the middle of RBMX's sequence, is autonomously responsible for targeting RBMX to nascent transcripts in *Xenopus oocytes*[493]. It does not contain

the consensus nuclear localization signal (NLS), the RRM is not implicated in this recruitment[418,493] and the NTD lacks RNA affinity[493] suggesting that probably it is recruited through protein-protein interactions. Together, these studies paint a picture where NTD associates to nascent transcripts indirectly first, maybe fulfilling a more general role and then RBMX engages in direct binding to RNA when appropriate cis-acting signals appear from the emerging RNA. Importantly, this domain has not been studied in RBMXL2, neither has RBMXL2 ever been reported to associate with nascent transcripts. However, the homologous region has 74% identity with RBMX (86% similarity). It is difficult to speculate whether this degree of similarity would be enough for the recruitment of RBMXL2 to active transcription sites, whether it would be recruited with the same affinity as RBMX and if that recruitment affects their respective regulatory roles on splicing. Investigating it would be important for a deep understanding on the molecular mechanisms behind RBMXL2 and RBMX alternative splicing regulation in naïve and primed pluripotency.

Another relevant feature of RBMX and its retrocopies is the presence of two arginine-glycine-glycine (RGG) boxes. RGG boxes are highly conserved motifs rich in arginines and glycines, often as arginine-glycine-glycine repeats, that add high functional versatility to the proteins harbouring them. They grant key roles influencing physiological processes such as transcription, alternative splicing, DNA-damage response, translation and apoptosis by conferring crucial biochemical properties like RNA-binding, protein-protein interaction and protein localization[504,505]. The RGG motif is the second most common RNA binding domain[506] but its specificity has been described as degenerate for features that are widely shared, thus granting promiscuous binding[507]. However, they are not entirely indiscriminate as most studies agree in a preference for GC-rich sequences and complex RNA structures[507]. This fits in well with the pattern we saw for primed included exons which were precisely high in GC-content. In line with this, a recent study examining target nucleotide composition bias across different splicing factors reported that RBMX-dependent exons are highly GC-rich and predicted that they are embedded in stable secondary structures[508]. Considering that the reported preference of RBMX RRM (discussed above) is not high GC content in complex RNA structures, it is possible that this propensity might be given by RGG.

The sequences in RGG boxes in RBMXL1 and RBMXL2 slightly diverged from RBMX. RBMXL1 has a deletion of a whole RGG repeat on the first RGG box (RGG1) and RBMXL2 has two relevant substitutions on the first and the second RGG box (RGG1 and RGG2, respectively). More specifically, RBMXL2 substituted a glycine in one of the three RGG repeats of the RGG1 box to a threonine. This substitution locates precisely in the boundary between two consecutive RGG repeats and might affect the

functionality of both RGG repeats in the RGG1 box. The second possibly influential substitution in RBMXL2 lies in the second RGG repeat of the RRG2 box, where a glycine is substituted by a valine. Both valine and threonine are branched-chain amino acids which implies more bulkiness near the protein backbone whereas glycine, the amino acid they substitute, is the smallest amino acid as it instead contains a hydrogen as its side chain[509]. This could mean that these two key substitutions could reduce the flexibility and add rigidity to the RGG1 and RGG2 domains. The arginine in RGG repeats is frequently the residue establishing hydrogen bonding and amino-aromatic interactions, thus, it is often responsible for mediating the protein's interactions to nucleic acids and proteins. Therefore, since RBMXL2 conserves all arginines in both RGG boxes it might be that RBMXL2 is still capable to establish interactions albeit with restriction compared to RBMX. Importantly, RGG boxes are intrinsically disordered, with no stable structure and no compacted form given their low hydrophobicity. Accordingly, the peptide chain is solvent, exposed and with a large interaction surface area[507]. This guarantees plasticity and freedom to sample different conformational states in a fast and efficient way. In agreement with this, RBMX has been shown to bind both structured and unstructured RNAs[421,422], this however, has not been investigated for RBMXL1 or RBMXL2.

Although the relationship of RGG domains with RRMs remains unexplored, the fact that RGG motifs seem to have greater binding affinity might suggest that binding might be largely driven by RGG and then fine-tuned by the RRM. The RRM might then add structural stability to the interaction and thus increase the affinity to the target. It is also tempting to think that the long and disordered structure between both RGG boxes allows RBMX family of proteins to bind with more than one RNA or with distant motifs of the same RNA. The later would bridge distant RNAs close together and could potentially aid in splice site selection by generating RNA looping as PTB does[375]. We propose here that the RGG boxes of RBMXL2 and RBMXL1 might have a reduced conformational flexibility compared to RBMX and that this might reduce RNA sequence recognition or shift RNA preference and account for the differing impact of protein depletion in the context of pluripotency.

Various yeast two-hybrid studies have shown multiple protein-protein interactions of RBMX including RNA processing factors such as hnRNPC, SAM68, YT521B, SAFB, TRA2B and hnRNPK[418,419,467,510–513]. Also, RBMX seems to interact with itself and with RBMXL2 and this interaction is not nucleic acid-mediated[418]. Together, it seems that RBMX might build protein-protein interaction hubs to grant regulation of protein-RNA interactions and RNA metabolism. In line with this, one of the most compelling interactions of RBMX has been very recently discovered, this is, with the

phosphorylated C-terminal of RNA polymerase II[422]. Together with binding to nascent RNA, this renders RBMX the ability to associate co-transcriptionally, modulate RNA polymerase II occupancy in regulated exons and fine-tune alternative splicing transcriptome wide. Interestingly, the RGG boxes of RBMX are responsible for this binding[422]. Thus, given their subtle differences in RGG boxes, RBMXL2 and RBMXL1 might bind RNA polymerase II with less affinity compared to RBMX. In this line, Zhou et al tested the impact on splicing regulation of RBMX mutations in the RRM, RGG1 and RGG2[422]. Curiously, RGG1 mutations affected the inclusion of much less targets compared to mutations in RGG2 or RRM, which had overlapping effects. This may hint that RBMXL1 might still retain most of RBMX splicing targets, and the effect of losing an RGG repeat in RGG1 might not be so severe on the alternative splicing regulatory function. Additionally, as discussed previously, we found that the short isoform of RBMX does not contain the RGG2 box. Interestingly, rescue of RBMX knockdown by overexpressing an RBMX RGG2 mutant has a more severe phenotype on alternative splicing landscape than non-rescued mutants[422]. This is suggestive of a dominant negative effect implying that while RBMX RGG2 mutant loses RNA binding capabilities and maintains the ability to interact with RNA polymerase II, it blocks other proteins from binding. Thus, our observations together with results from Zhou et al[422], allows to speculate that the short isoform of RBMX might have a similar effect.

Protein-protein interactions of RBMX (and potentially, RBMXL1 and RBMXL2) with other RNA binding proteins can be either synergistic or antagonistic with respect to the functional role of the bound protein. These effects may be independent from RBMX RNA binding, either because RBMX might aid in the interaction with other proteins or, in the case of the antagonistic effect, because RBMX might sequester that protein and inhibit its binding to target RNAs. One relevant interaction is with the splicing factor TRA2B as RBMX has been shown to have both synergistic[419] and antagonistic[417,502] effects depending on the splicing event being regulated. Specifically, in the case of *Lysine Demethylase 4* (*Kdm4)* and *Tra2b* exons, RBMX has an antagonistic effect because it either competes for binding sites with TRA2B1 or because it sequesters TRA2B1[417,423,502]. In the case of *Survival motor neuron gene 2* (*Smn2*) exon, RBMX has a synergistic effect as it is recruited by TRA2B1 bound to RNA, binds upstream and enhances exon inclusion[424]. Interestingly, those levels of inclusion cannot be reached if either TRA2B1 or RBMX are expressed separately, indicating an additive effect probably in recruiting spliceosomal components[419]. Interestingly, the domain in RBMX responsible for TRA2B binding was mapped to the c-terminus, spanning just a few amino acids more than the region coded by the alternative exon[419]. A mutation in this region didn't just abolish RBMX-TRA2B binding but it also abrogated binding to the synergistically regulated RNA target *Smn2* exon 7 transcript[419]. Additionally, the RRM domain was shown not to be necessary for

the regulation of this exon[418,424], possibly indicating that RBMX RNA binding is not necessary. Precisely *Drosophila melanogaster* transformer locus (tra) protein, one of the main players in sex determination, regulates splicing without binding to RNA. Tra binds to Tra2, which does have a RNA binding domain, forming a complex in female flies that regulates the splicing of target transcripts and determines the sex of the fly[514].

On the other hand, only antagonistic effects for RBMXL2 with TRA2B have been reported so far [417,423]. RBMXL2 amino acid sequence diverges from RBMX more in this c-terminal alternative exon compared to the whole protein (70% identity and 76% similarity vs. 75% identity and 83% similarity). This, together with our results indicative of contrasting roles in pluripotency, might hint that the binding to TRA2B by RBMXL2 might be diminished or at least different to that of RBMX. However, no study to date has investigated the role of TRA2B in pluripotency. Curiously, by looking into our data and VastDB, we have seen that TRA2B is equally highly expressed in naïve and primed pluripotency (125cRPKMs) and it is an early gene that is upregulated upon reprogramming of mouse fibroblast to iPSCs (by day 4). It would certainly be interesting to study the cross-talk, if any, between RBMX, RBMXL2 and TRA2B1 in pluripotency as it could potentially pinpoint the molecular mechanism behind specific regulatory differences among RBMX and RBMXL2.

RBMX has been recently identified as a N^6-Methyladenosine (m^6A) reader[421]. m^6A is the most abundant internal modification of mRNA in eukaryotes, it is reversible, deposited co-transcriptionally and involved in a number of RNA processes such as splicing, RNA stability and translation[515–518]. Most m^6A modifications occur in exonic regions near splice sites[516,517] and can regulate splicing by recruiting m^6A readers that can either recruit splicing factors or be splicing factors themselves[421,519,520]. Additionally, high levels of m^6A in introns is associated with alternative splicing events[517]. RBMX recognises m^6A-methylated RNA but not through direct recognition of the m^6A, instead, RBMX binds a purine-rich motif that becomes accessible upon m^6A modification of the RNA structure[421]. This suggests that m^6A aids in the recognition and binding of RBMX to part of its targets. RBMX-bound m^6A sites are highly enriched in RNA polymerase II occupancy and are largely located in splice sites of RBMX-regulated exons[422]. This suggests that the inclusion of a part of RBMX-regulated exons is m^6A-dependent.

Interestingly, m^6A weakens mRNA stability in mESCs, and is not required for stem cell maintenance or preimplantation embryo development but its depletion promotes naïve pluripotency and blocks naïve-to-primed transition leading to early post-implantation lethality[521,522]. Depletion of m^6A in EpiSCs leads to faster differentiation, disruption of the primed state and cell death[521]. With this, it would be interesting to

understand how m^{6}A regulation and splicing converge to functionally regulate naïve and primed pluripotency. Specifically, it would be very useful to know whether m^{6}A-marks are enriched in exons regulated in the naïve-to-primed transition. Given our observation of high cell death upon differentiation of *Rbmx RbmxL1* depleted cells, one possibility is that RBMX and RBMXL1 might be playing a role in the exit from pluripotency by regulating m^{6}A-marked exons. Additionally, RBMX binds m^{6}A-methylated RNA through its RGG2 box[421]. The RGG repeats at RGG2 are essential for the recognition[421]. We propose that since RBMX short isoform does not contain this RGG2 box, it might not recognize m^{6}A-methylated exposed motifs, and thus in the naïve state, stem cells don't regulate those exons. Also, RBMXL2 might have a lower capacity to bind to the m^{6}A-methylated exposed motifs given its substitution in one of the RGG repeats (discussed above). Therefore, in naïve stem cells m^{6}A-methylated RNA might be unrecognized by the most part of the expressed RBMX family whereas in primed pluripotency, m^{6}A-methylation might be playing a role in fine-tuning alternative splicing through RBMX. In contrast with other m^{6}A readers, RBMX binds directly to the m^{6}A-methylated site which might also allow to control RNA fate by modulating the binding of other m^{6}A readers.

Non-splicing roles of Rbmx family members

It is also possible that the main role of the Rbmx family of proteins in the naïve-to-primed transition is not related to their role in alternative splicing regulation. In the upcoming section I will give a brief overview of the additional possible roles RBMX, RBMXL1 and RBMXL2 might have over naïve and primed pluripotency and speculate on their impact on cell state determination.

Waisman et al reported that stalls in DNA replication, as the ones generated by the DNA-damage response, inhibit the naïve-to-primed pluripotency transition[523]. Also, upon DNA damage, RBMX is recruited at the site of damage and associates with double strand breaks in a Poly(ADP-ribose)-dependent manner[488]. *Rbmx* depletion sensitizes cells to DNA-damaging agents[488]. The mechanism by which RBMX contributes to the DNA-damage response remains unknown and intriguing as RBMX localization to the damaged sites is not necessary for repair[488]. One possibility is that RBMX is post-translationally modified at the sites of DNA damage, then released to perform an adjusted alternative splicing regulation. Additionally, *Rbmx* depletion reduces the expression levels of *Brca2*, a key protein in homologous recombination[488]. Thus, the effect on DNA-damage response could be indirect through regulation of BRCA2 levels. It has been suggested that this regulation is through alternative splicing but it has never been tested. Interestingly, depletion of *RbmxL2* induces the utilization of a

cryptic splice site of the *Braca2* transcript, generating an aberrant splice isoform[423]. This again suggests that the Rbmx family might be regulating the alternative splicing of *Braca2*. Another possible explanation is that given its association to nascent RNA[422], RBMX responds to DNA damage (or RNA damage) by changing RNA processing or transcription to ensure correct gene expression. In any case, we believe it is quite unlikely that the RBMX family might be regulating the naïve-to-primed transition through its function on DNA-damage as the exit from naïve pluripotency is not a DNA-damaging process.

On a different note, RBMX binds to H3K9me3-marked heterochromatin and maintains the sonication-resistance of heterochromatin presumably to inhibit transcription factor activity and cell identity conversion[491,524]. In human cells, double-knockdown of *Rbmx* and *RbmxL1* loosens sonication-resistant heterochromatin but does not upregulate many genes in absence of ectopic relevant transcription factors[491]. Also, it does not reduce sonication resistance of non-heterochromatic sites[491]. One possibility is that this is mediated by the RGG domains of RBMX and RBMXL1 as they could be involved in phase separation which is key for heterochromatin formation[525,526]. Curiously, double-knockdown of *Rbmx* and *RbmxL1* on fibroblasts also renders a higher reprogramming efficiency in induced hepatocyte reprogramming[491]. Mouse and human stem cells broadly show open chromatin and decompaction in both naïve and primed pluripotency and this clearly separates them from somatic cells[185], even if it is not as open as earlier stages of preimplantation embryo development[527]. There is a gradual increase of H3K9me3 foci when directly comparing 2iL ESCs, SL ESCs and EpiSCs[528]. In EpiSCs, these H3K9me3 foci are essential for silencing the major satellite repeats at pericentromeric heterochromatin[528]. Additionally, the heterochromatin architecture also changes from naïve to primed pluripotency from widely disperse 10nm fibres to peripheral heterochromatin[186,529]. It could be that RBMX recruitment to H3K9me3 foci is important to establish the heterochromatic architecture of EpiSCs. Importantly, H3K9me3 removal in EpiSCs does not change the primed identity[528] suggesting that the effect of RBMX on pluripotency state through heterochromatic architectural change might not be enough to explain the shift we see towards naïve state in *Rbmx* exon KOs.

RBMX has also been directly implicated in transcriptional regulation in two independent studies that reported either transcriptional repression or transcriptional enhancement controlled by *Rbmx* expression. Zhao et al reported RBMX binding to the promoter of *Gonadotropin-releasing hormone 1* (*GnRH1*) gene and showed that overexpression of *Rbmx* downregulates its expression[489]. In the other hand, Takemoto et al reported RBMX binding to the promoter of the *Sterol regulatory element (SRE)-*

binding protein 1c (*Srebp1c*) gene and showed that RBMX acts as an enhancer of its expression[490]. We cannot rule out that RBMX might be acting as a transcriptional regulator in the naïve-to-primed transition and that part of the phenotypes that we observe in our knock-out mutants might be due to transcriptional regulation as well as post-transcriptional regulation.

Regulation of naïve and primed pluripotency states by the Esrp/Rbm47 regulatory axis

On the role of ESRP1 and ESRP2 in pluripotency state definition and in differentiation

ESRP1 and ESRP2 are two paralogous RNA-binding proteins that coordinate an epithelial-associated splicing program mainly by regulating splicing events in genes involved in actin cytoskeleton organization, cell-cell adhesion, cell polarity and cell migration[433,434]. To date, only two studies have touched upon ESRP functions in stem cells[389,530]. A first study investigated the role of ESRP1 in mESC pluripotency and found that knockdown of *Esrp1* induces an increase in expression of the pluripotency factors *Oct4*, *Sox2* and *Nanog* which in turn enhance self-renewal and impair early differentiation *in vitro*[530]. The authors described this fact as "paradoxical" considering that *Esrp1* expression is downregulated simultaneously with pluripotency factors during pluripotency exit. Importantly, they completely overlooked the distinction between naïve and primed pluripotency. Mechanistically, this study proposed that ESRP1 controls pluripotency through the binding to the mRNA of the above-mentioned pluripotency factors, sequestering them and reducing their polysomal loading[530]. However, the study lacked sufficient data to support such a claim as the researchers based their argument on an RNA immunoprecipitation of ESRP1 in ESCs, which lacked proper controls. A later study stated that ESRP1 and ESRP2 were not required for pluripotency maintenance[389]. However, this study did not present any evidence towards it as data was not shown, and based the argument on the observation that Esrp1 Esrp2 DKO ESCs could be maintained indefinitely in stem cell culture without pluripotency marker expression loss[389]. A more thorough examination is needed to fully understand the functions of *Esrp1* and *Esrp2* on distinct pluripotency states, pluripotency exit and lineage determination.

The more general role of the ESRP regulatory programme in development has previously been studied[390,391]. Importantly, deficiency of the programme does not lead to early embryonic lethality in mice. Instead, the depletion of *Esrp1* in mice results in neonatal lethality at postnatal day 0 due to deficient craniofacial development with fully penetrant cleft lip and cleft palate[390]. On the other hand, *Esrp2* depleted mice are viable and fertile but have an abnormal liver development with higher expression of foetal markers and low mature hepatocyte characteristics[379,390]. Furthermore, the combined depletion of both *Esrp1* and *Esrp2* in mice results in marked organogenesis defects associated to epithelial-cell specific deficiencies[390]. More specifically, Esrp1 Esrp2

DKO mice are smaller and have more profound craniofacial phenotypes, malformed forelimbs, forepaw syndactyly, lung and salivary gland agenesis and thin transparent skin with less hair follicles[390]. The severity of Esrp1 Esrp2 DKO mice phenotypes compared to single Esrp1 or Esrp2 KOs clearly implies functional redundancy. This functional redundancy has also been observed at the level of alternative splicing regulation[390,434].

In this thesis, we show that *Esrp1* and *Esrp2* are differentially expressed in 2iL ESCs, SL ESCs and EpiSCs, indicating regulated expression levels in the naïve-to-primed pluripotency transition. Specifically, *Esrp1* expression is completely absent in 2iL ESCs but present in both SL ESCs and EpiSCs. Conversely, *Esrp2* expression is detectable in 2iL ESCs and SL ESCs but downregulated in EpiSCs. Additionally, *Esrp1* expression is much higher than *Esrp2* expression in both SL ESCs and EpiSCs. This goes in line with reports showing that in most tissues where both genes are expressed, *Esrp1* is more highly expressed relative to *Esrp2*[390,531]. We observed that *Esrp1* depletion in SL ESCs does not induce a significant change in *Esrp2* expression while *Esrp2* depletion in SL ESCs induced an increase of *Esrp1* expression in one of the Esrp2 KO clones but not in the other. A previous study using full thickness epidermis from E18.5 KO mice also examined the cross regulation of *Esrp1* and *Esrp2* expression upon depletion[390]. It showed a trend to increase *Esrp1* expression upon *Esrp2* depletion, in line with what we have observed in our KO ESC lines. Also, in line with our results, in this study they did not observe a decrease of *Esrp2* expression upon *Esrp1* depletion. Nonetheless, further research is needed to fully uncover the cross regulation between these two paralogues specially to find whether there are any cell type specific compensatory networks in the ESRP1 ESRP2 regulatory axis.

In contrast to previously published results[530], we found that *Esrp1* depletion in ESCs does not substantially affect pluripotency exit dynamics. More specifically, Fagoonee et al reported that *Esrp1* knockdown cells had higher AP staining levels compared to controls both in SL ESC conditions and upon differentiation[530] whereas we found that the distribution of AP stained colonies was not significantly different in Esrp1 KOs in either of the two conditions. Importantly, Fagoonee used short hairpin RNA (shRNA)-mediated knockdowns of *Esrp1* whereas we made complete gene knockouts of *Esrp1*. Thus, the discrepancy in the results might be due to off-target effects or genetic compensation through transcriptional adaptation.

In the case of *Esrp2* depletion, however, we found that although both Esrp2 KO clones have similar AP staining levels in SL ESC conditions, they have differing levels of AP staining upon differentiation. Specifically, one Esrp2 KO clone (clone 75.1) showed

similar AP staining levels compared to controls while the other (clone 75.2), showed higher levels of AP staining compared to controls. This is indicative of a slower exit from pluripotency of Esrp2 KO clone 75.2 but not of Esrp2 KO clone 75.1. It might be that Esrp2 KO clone 75.2 takes longer to exit pluripotency because it expresses some components of the pluripotency network at higher levels in comparison to controls causing it to take longer to completely dismantle the pluripotent regulatory system, however, we did not find such a difference. If that were the case, we would expect that Esrp2 KO clone 75.2 also showed higher levels of AP staining in SL conditions, which is not the case. Thus, a deeper analysis would be needed to fully understand the disparity between the two Esrp2 KO clones. Curiously, one of the most striking differences between both clones is that Esrp2 KO clone 75.2 displays much higher ESRP1 protein levels compared to Esrp2 KO clone 75.1. The impact that this may have on pluripotency exit, however, remains to be explored.

The simultaneous depletion of both *Esrp1* and *Esrp2* expression in SL ESCs leads to upregulation of naïve pluripotency differential expression profile and downregulation of primed pluripotency differential expression profile. Additionally, depletion of *Esrp1* and *Esrp2* expression also generates a more naïve differential alternative splicing profile with both more inclusion of naïve highly included exons and more exclusion of primed highly included exons. This tendency towards a naïve pluripotency transcriptional landscape in Esrp1 Esrp2 DKO cells might be the cause for the subtle yet significant changes in the exit dynamics from naïve pluripotency. Specifically, Esrp1 Esrp2 DKOs show a slight delay in the downregulation of naïve pluripotency marker expression upon EB differentiation. Additionally, Esrp1 Esrp2 DKOs show a slight delay in the downregulation of the expression of two primed pluripotency markers (*Fgf5* and *Otx2*), but not of other primed markers such as *Oct6* or *Dnmt3b*. This might indicate a subtle influence of the alternative splicing landscape switch regulated by ESRP1 and ESRP2 on primed pluripotency exit. However, much more work would still be needed in this direction to understand precisely which alternative splicing event switches are relevant in naïve and primed pluripotency exit.

ESRP1 and ESRP2 are epithelial-associated master regulators of alternative splicing events involved in epithelial-to-mesenchymal transitions (EMT), among other processes [532]. EMT is the morphogenetic process by which polarized epithelial cells become mesenchymal cells (see introduction). ESRP1 and ESRP2 negatively regulate EMT by maintaining an epithelial-specific alternative splicing programme[432]. Together with ESRP1 and ESRP2 downregulation, the EMT program is driven by a series of transcription factors that concomitantly inhibit the expression of cell-cell adhesion proteins, trigger mesenchymal morphology and increase cellular migratory

capabilities[532,533]. We found that Esrp1 Esrp2 DKO cells have higher induction levels of EMT drivers such as *Brachyury* and *Eomes*[435–438] during EB differentiation. This was expected because EMT is promoted by depletion of *Esrp1* and *Esrp2*; however, it somewhat clashes with Fagoonee's previously published study where the single knockdown of *Esrp1* was shown to downregulate the expression of *Brachyury* during EB differentiation[530]. This suggests that ESRP1 depletion has a different effect on EB differentiation potential than simultaneous and complete depletion of both ESRP1 and ESRP2. It is nonetheless counterintuitive to think that the effect would be antagonistic as ESRP1 and ESRP2 are considered to play functionally redundant roles[532]. With the work I present here, we cannot rule out that possibility given that we have not performed EB differentiation experiments with Esrp1 KO clones. However, as discussed above, this study also claimed slower exit from pluripotency of Esrp1 KD ESCs, which we did not observe in our Esrp1 KO ESCs. Again, this discrepancy might be due to off-target effects or genetic compensation through transcriptional adaptation and encourages a deeper understanding of the potential role of ESRP1 and ESRP2 in differentiation.

During EB differentiation of *Esrp1* and *Esrp2* depleted cells, after a boost of primitive streak marker induction, we would expect a differentiation bias enriched towards mesenchymal lineages. What we found is that Esrp1 Esrp2 DKOs precisely downregulate neuroectodermal lineage induction. However, this is also accompanied by a downregulation of endodermal lineage induction, and while there is a trend towards an enhanced expression of mesodermal markers, this is not significant due to clone disparity. These differences could be explained by different genetic compensatory mechanisms among clones, yet they remain to be explored. Curiously, Fagoonee et al reported that teratomas derived from Esrp1 KD ESCs show greater neuroepithelium content compared to controls albeit only when ESCs were injected under non-saturating conditions[530]. This finding could go against our results, as we did not see a bias towards neuroectoderm lineage in Esrp1 Esrp2 KO ESCs. Again, to clarify this discrepancy, it would be key to explore the differentiation bias of our Esrp1 KO clones.

Finally, it is important to add that independently from the role of ESRP1 and ESRP2 on alternative splicing regulation there could be other roles that might impact on naïve and primed pluripotency such as translational regulation. In line with this, *Esrp1* expression was reported to affect polysomal loading of a number of RNAs in ESC and cancer cell lines[530,534]. Furthermore, *Esrp1* expression negatively affects the translation efficiency of reporter mRNAs mediated by several 5'UTR sequences[530,534]. The intensity of translational inhibition by ESRP1 depends on the complexity of the secondary structure of the mRNA 5'UTR[534]. Interestingly, the 5'UTR of *c-Myc* was found among the most strongly regulated 5'UTR sequences in cancer cells[534]. MYC protein levels

correlate with pluripotency state: naïve pluripotent stem cells have high MYC levels, while primed pluripotent stem cells express lower levels[535]. Importantly, differing MYC levels trigger competitive interactions among stem cells to ensure embryos preserve the naïveness of the stem cell pool before gastrulation[535]. Cells containing high levels of MYC eliminate those with lower levels by cell competition (by triggering apoptosis after repeated contact)[535]. It is possible that ESRP1 might be playing a role in this by downregulating MYC protein levels in stem cells and more research remains to be done in this direction to assess the extent of ESRP1's translational regulatory role in pluripotency.

On the role of RBM47 in pluripotency state definition and in differentiation

RBM47 has previously been involved in developmental processes. Its knockdown in zebrafish results in complete loss or reduced head development[536]. In contrast, *Rbm47* in mice is critical for embryonic viability and postnatal growth[537]. Most Rbm47 KO mice die during gestation from E10-11 onward[537]. The few that are born, are consistently smaller in size compared to controls and do not close that size gap throughout their whole life[537]. Furthermore, restoring *Rbm47* expression in the embryo proper and not in extraembryonic tissues in Rbm47 KO mutants results in phenotypically normal rescued mutants with growth and viability similar to controls[537], indicating a critical role for *Rbm47* expression specifically in the embryo proper.

Examining the role of RBM47 in the regulation of naïve and primed pluripotency alternative splicing landscapes was compelling in many ways. First, *Rbm47* is differentially expressed in naïve-to-primed transition. Second, it was previously described that RBM47 mostly promotes epithelial splicing patterns with a significant overlap of coregulated EMT-associated targets with ESRP[425], yet *Rbm47* expression in 2iL ESCs and EpiSCs showed an inverse pattern compared to *Esrp1* expression. Curiously, it had been previously shown using a cancer cell line that 39% of the RBM47-ESRP overlapping targets were regulated in a different or opposite direction[425]. Although the effects of this different regulation were overlooked, they are likely crucial to decisively uncover the true degree of functional compensation of each factor.

We found that *Rbm47* depletion in ESCs leads to a downregulation of naïve pluripotency differential expression profile and an upregulation of primed pluripotency differential expression profile. Additionally, it leads to higher exclusion of naïve highly included exons. Despite this tendency towards a primed pluripotency transcriptional landscape, Rbm47 KO ESCs do not show substantial changes in the dynamics of exit from naïve

pluripotency or from primed pluripotency. However, we did find a differentiation bias of Rbm47 KO ESCs towards endoderm induction downregulation. This goes in line with previous findings showing that *Rbm47* is expressed in the definitive endoderm of E8.5 mouse embryos and afterwards restricted to endoderm-derived organs[538]. Together with our results, this might indicate that *Rbm47* expression is important for endoderm specification, but more work is needed in this direction.

RBM47 has been previously reported to play a role in alternative splicing regulation[389,539]. However, other roles have also been reported and it is thus possible that the phenotypes we and others have observed upon *Rbm47* depletion might be due to the sum of different molecular roles of RBM47. For example, recently, RBM47 was shown to interact with microRNA (miRNA) precursors in a large-scale screen for RBP-pre-miRNA interactions[540], suggesting that RBM47 might be involved in the biogenesis of specific miRNAs. Naïve and primed pluripotent stem cells have different miRNA signatures[278,541–543] and miRNAs play a key role in the naïve-to-primed transition as shown using miRNA-deficient Dgcr8 KO ESCs[541,544]. This is the case of the miRNA miR-302 cluster, which is more highly expressed in primed pluripotency[541,542,545] and its increase in expression levels promotes exit from naïve pluripotency by enhancing the MEK pathway[278]. It would be interesting to know whether RBM47 plays any role in shaping the naïve miRNA landscape in addition to its impact on the alternative splicing landscape. Furthermore, RBM47 was reported to bind to NANOG[546] however, the effect of this binding remains unexplored. NANOG is a key transcription factor in naïve pluripotency maintenance, its expression fluctuates and is finally downregulated in primed pluripotency[209,547,548] (see introduction). Its binding to RBM47 might have a relevant effect on naïve pluripotency stabilization and would be an interesting aspect to explore. Finally, RBM47 is an essential co-factor of APOBEC1-mediated Cytidine-to-Uridine (C-to-U) RNA editing[538]. Although C-to-U RNA editing has never been examined in stem cells, it would be interesting to know if it occurs and whether it has an important impact on pluripotency. Curiously, the only other known cofactor of the APOBEC deaminase, is A1CF and it is not expressed in naïve or primed stem cells. This could suggest a complete dependency of C-to-U RNA editing on *Rbm47* expression in the context of pluripotency.

Regulatory crosstalk and co-regulation of Rbm47 and Esrp in pluripotency and differentiation

We found that *Rbm47* depletion in ESCs leads to an upregulation of *Esrp1* expression levels. This upregulation in ESRP1 levels might be contributing to the downregulation of naïve pluripotency that we observed in Rbm47 KOs, as Fagoonee et al reported that overexpression of *Esrp1* in ESCs leads to the downregulation of *Oct4*, *Sox2* and *Nanog* expression levels[530]. Furthermore, Esrp1 Esrp2 DKO clones showed decreased levels of *Rbm47* expression compared to controls. Despite the disparity in the different molecular contexts, this agrees with the results from Yang et al in cancer cell lines[425], as they reported that *Esrp1* and *Esrp2* depletion reduced the levels of *Rbm47* expression. Together, these results suggest a complex crosstalk among these three molecular players that still remains to be elucidated.

Interestingly, Esrp1 Rbm47 DKOs show a strong delay in the exit from pluripotency, but the simultaneous depletion of *Esrp2,* together with *Esrp1* and *Rbm47*, abrogates this phenotype. This indicates that ESRP2 must play a role in delaying pluripotency exit in the absence of ESRP1 and RBM47. We did not find a significant change in *Esrp2* expression upon *Esrp1* and *Rbm47* depletion, however, we did not test ESRP2 at the protein level or during pluripotency exit. In any case, it would be important to further explore the role of ESRP2 in the context of RBM47 and ESRP1 depletion. Particularly, whether ESRP2 carries specific functions that are masked upon RBM47 or ESRP1 expression. In this line, previous reports using cancer cell lines and immortalized epithelial cell lines have shown that ESRP2 is a weaker splicing regulator compared to ESRP1 and that they share many of their targets[432,549,550]. Target preference, however, is likely context-dependent and can be affected by expression levels of the target mRNAs and by posttranslational modifications of the regulators. Importantly, to date, no study has examined the functional redundancy of ESRP1 and ESRP2 in ESCs but the results I present in this thesis clearly point that it would be a very interesting path to explore. One possibility could be that ESRP2 has specific alternative splicing targets which have an impact on pluripotency exit only in the context of *Esrp1* and *Rbm47* depletion. Another possibility is that ESRP2 has splicing-independent roles in this context. In line with this, previous studies have suggested that ESRP2 has specific splicing-independent roles[550] but to date, there is no direct evidence that corroborates that hypothesis in ESCs.

We found that Esrp1 Esrp2 Rbm47 TKO ESCs downregulate both naïve and primed pluripotency differential expression profiles. This could suggest that the right balance

of either ESRP or RBM47 is needed in pluripotency and when both are not present, the cells acquire an identity that is neither naïve nor primed.

Regarding alternative splicing, Esrp1 Esrp2 Rbm47 TKO ESCs show higher exclusion of naïve highly included exons. This is comparable to the phenotype that Rbm47 KOs showed and in contrast to Esrp1 Esrp2 DKOs. Additionally, the inclusion levels of alternative splicing changes in Esrp1 Esrp2 Rbm47 TKO ESCs are closer to those found in Rbm47 KOs compared to those found in Esrp1 Esrp2 DKOs, which in turn, cluster separately. These results are in disagreement with Yang et al who previously proposed that the role of ESRP1 predominates over RBM47 when they have opposing alternative splicing functions[425]. However, this apparent disagreement over the dominance of ESRP1 vs. RBM47 could be due to the different systems studied as Yang et al used EMT in cancer cells.

We observed that Esrp1 Esrp2 DKO ESCs have a weaker induction of DNA Methyltransferase 3 Beta (*Dnmt3b*) expression during EB differentiation compared to controls. Additionally, Esrp1 Esrp2 Rbm47 TKOs had a slight trend in this direction but did not show such substantial *Dnmt3b* expression differences compared to controls. Thus, in this case, it seems that RBM47 might not be functionally compensating for ESRP loss, on the contrary, it might be enhancing the effect of *Esrp1 Esrp2* depletion on *Dnmt3b* downregulation. However, the difference is too small to draw any solid conclusions. Together, this might indicate that ESRP1, ESRP2 and RBM47 cooperate to regulate *Dnmt3b* expression. DNMT3B is a *de novo* DNA methyltransferase whose expression is induced during primed pluripotency[551,552] (see introduction), thus it might be interesting to examine whether *Esrp1*, *Esrp2* and *Rbm47* depletion affects the epigenetic profile of these differentiating cells. Specifically, it would be compelling to look for the genes with differential gene body methylation in DKOs and TKOs compared to controls as this differential methylation could also help explain the differentiation bias in the mutants.

Previous to this thesis, Yang et al had proposed that reduction of RBM47 levels together with reduction of ESRP levels favours the induction of EMT[425]. We found that during EB differentiation Esrp1 Esrp2 Rbm47 TKO ESCs show similar induction levels of *Brachyury* compared to controls but lower induction of other EMT drivers such as *Eomes* and *Snail*. This came as a surprise since it partially disagreed with what was previously predicted: if the downregulation of both ESRPs and RBM47 was important for EMT induction, we would have seen a higher induction of EMT drivers upon depletion, not lower. Furthermore, there is a clear downregulation of endomesodermal lineage induction and an upregulation of neural lineage induction in Esrp1 Esrp2

Rbm47 TKOs upon differentiation. Thus, they display a strong epithelial bias, not mesenchymal. Together, these data suggest that during stem cell differentiation there is a regulatory programme co-regulated by ESRP1, ESRP2 and RBM47 which must be shut down for neuroectodermal differentiation to occur, not for EMT (as it was shown for cancer cells). Indeed, neural tissues show very low expression levels of *Esrp1*, *Esrp2* and *Rbm47* and it would be very interesting to explore if and how neural progenitors downregulate these three factors and the impact this downregulation may have on cell function and physiology.

Regulation of naïve and primed pluripotency states by Quaking

One underlying question in this thesis is whether the alternative splicing landscape switch in naïve-to-primed transition can actually drive the transition from the naïve to the primed cell state. Although with this thesis we have only scratched the surface of a possible answer, one of our most meaningful observations towards it came by studying the possible regulatory role of QKI in the naïve-to-primed transition. QKI has emerged as a crucial factor of many developmental transitions and differentiation processes including oligodendrocyte differentiation[440–444], monocyte/macrophage differentiation[445], vascular development[446] and EMT[425,449,553]. Here, we describe that QKI regulates many naïve-to-primed alternative spliced cassette exon events and that severe or complete depletion of QKI results in a more primed alternative splicing profile.

Notably, 2iL ESCs maintain higher mRNA expression levels of QKI than SL ESCs and EpiSCs, thus more naïveness entails higher expression levels of QKI, however, we also found that QKI depletion promotes the upregulation of a naïve differential gene expression profile. The later goes in line with recent findings showing that QKI loss maintains stemness and self-renewal of neural stem cells and glioma stem cells by modulating RNA stability and alternative splicing[554]. Mainly, QKI loss downregulates the endolysosome-dependent degradation of receptor-ligand complexes. This desensitizes cells to ligand-receptor complex-mediated signalling and as a result, cells are not poised to receive new fate determination signals[555]. It is possible that this is also the case in naïve pluripotency and more work should be done in this direction.

Taken together our work showed that QKI depletion promotes a naïve differential gene expression profile and downregulates a primed differential gene expression profile, but also promotes a more primed differential alternative splicing inclusion. This disconnection between the gene expression network and the alternative splicing landscape occurring upon QKI depletion in ESCs could have various interpretations. One possibility is that QKI functionality in the cell strongly depends on its protein load. In this model, in the complete absence of QKI, the cell gains a naïve gene expression identity and a primed differential alternative splicing network; then upon gradual increase of QKI protein levels, gene expression would change towards primed pluripotency; finally, at high QKI protein levels, the cell regains a naïve gene expression identity. The fact that inclusion/exclusion levels in QKI KDs are more extreme than in primed pluripotency, supports this hypothesis. Additionally, we detected disparity

in the differentiation potential of both QKI KD clones, which also supports this hypothesis. To test it, a key experiment would be to gradually decrease QKI levels in naïve stem cells and assess the gene expression profile. It is possible that the QKI KDs that we used were too extreme for cells to acquire a primed pluripotency gene expression, it might have surpassed that point and acquired a naïve pluripotency gene expression instead. Be that as it may, we found that the naïve-to-primed differential alternative splicing landscape can, in fact, be uncoupled from the naïve-to-primed gene expression network.

One crucial question that emerges from our study of QKI in naïve-to-primed transition is: how do primed stem cells downregulate QKI activity? We know that QKI is expressed at the same levels in SL ESCs and EpiSCs but that its activity is probably partly impaired in EpiSCs. This goes in line with another study that reported that QKI promotes EMT but does not change its expression levels during the process[425] (a different study involving QKI in EMT did see differences in QKI expression upon EMT albeit using other EMT model[449]). We will discuss a few possibilities that could explain how QKI activity could be regulated during the naïve-to-primed transition in the upcoming lines.

One possibility could be that QKI activity in pluripotency is regulated by a co-factor or a co-activator that could be interacting and enhancing target recognition only in naïve pluripotency. RBM47 could be one possible co-factor given its expression across ESCs downregulation in EpiSCs, however, we did not test RBM47's implication in QKI regulatory role. For this, it would be key to identify whether they share RNA targets and, if they do, it would be key to know how these RNA targets are recognized. Secondly, it would be important to find out if QKI and RBM47 physically interact, and if they do, how do both RBPs bind onto the same RNA transcript at different *cis*-elements, form a higher-order RNP complex and co-regulate splicing? What are the kinetics of binding? Is QKI binding first and then recruiting RBM47 or the other way around? Or perhaps both form a heterodimer before binding to the RNA targets? How is the secondary structure of RNA targets affected upon dual binding? Answering these and many other questions regarding this regulatory axis would certainly be crucial to further understand naïve-to-primed alternative splicing regulation.

Another possibility is that, in primed pluripotency, another splicing factor (or set of splicing factors) competes for *cis*-binding in QKI target RNAs. This has been previously suggested during skeletal muscle differentiation with poly-pyrimidine tract-binding protein (PTBP1)[447]. PTBP1 coregulates 42% of QKI-regulated alternative splicing events during muscle cell differentiation, and either cooperates or competes for

binding to regulate them[447]. In the case of exon 9 of Capping Actin Protein of Muscle Z-Line Subunit Beta (*Capzb*), for example, the inclusion is promoted in myotubes-to-myoblasts differentiation due to an increase in QKI and a decrease in PTBP1 levels[447]. In this case, exon inclusion is activated by QKI and repressed by PTBP1 and the binding motifs of these two splicing factors are in close physical proximity to one another on the downstream intron of the regulated exon suggesting a competition among these two splicing factors for binding to regulate the event. Curiously, PTBP1 also came up in our SPAR-seq screen as a possible regulator of naïve-to-primed alternative splicing events. Equally to QKI, its downregulation promoted primed differential alternative splicing inclusion. However, unlike QKI, we did not find a change in *Ptbp1* expression at the transcriptional level. PTBP2, the paralogue of PTBP1 however, does seem to have an increase in expression in primed pluripotency. Specifically, *Ptbp2* gene expression increases over five-fold in EpiSCs compared to 2iL ESCs (and over 4.6-fold compared to SL ESCs). Although PTBP2 and PTBP1 share 74% sequence identity and display similar RNA binding across the transcriptome, some of their target exons overlap and others are unique[556–565]. It could be that in primed pluripotency PTBP2 is competing for binding with QKI to regulate alternative spliced exons, yet this hypothesis remains to be tested.

Additionally, another possibility could be that QKI's activity is modulated through a post-translational modification that may be occurring either in naïve or primed pluripotency. We did not find differential phosphorylation of QKI in naïve and primed pluripotency, however, several other posttranslational modifications could be regulating QKI function. For example, QKI has been reported to be arginine methylated by PRMT1, although its effect on QKI activity is not known[566]. Furthermore, other STAR family members have been shown to increase their RNA binding capabilities upon lysine acetylation[567]. Thus, the possible post-transcriptional modifications of QKI upon naïve-to-primed transition continue to be an interesting subject-matter to explore.

Furthermore, it is important to point out that the regulation of alternative splicing is just one role of QKI, but might not be the leading role in naïve and primed stem cells. Indeed, roles in circular RNA production[449], micro-RNA processing[454,455], mRNA stability[456] and translation[457] have been described in other cell states and their specific impact on naïve and primed pluripotency maintenance should be studied for a complete understanding of QKI's role in naïve-and-primed pluripotency.

Finally, despite there being a clear difference in QKI protein levels between the two QKI KD clones used in this thesis, we found no substantial differences in naïve-to-

primed transition and exit from pluripotency among them. Importantly, we did find a discrepancy in their differentiation potential. More specifically, we found that QKI KD clone 2.2 (with higher QKI levels) had a strong deficiency in mesendodermal induction in EB differentiation while this was not apparent for the other clone. Also, we observed high cell death rates upon EB differentiation only in this QKI KD clone 2.2. These results suggested a milder phenotype in the cell line with almost undetectable levels of QKI. Firstly, these results should be validated using other clones as it could be an artefact resulting from an off-target CRISPR/Cas9 editing effect. If confirmed with other clones, it would be hard to explain by genetic compensation through transcriptional adaptation[498,499] given that both clones have a similar deletion (a short deletion in the first coding exon of the *Qki* loci). This would point instead to a compensatory mechanism at the protein level. To our knowledge, there are no reports showing this QKI compensation to date. However, QKI null mice have been reported to die at E10.5 due to defective vascular remodelling and abnormal vitelline vessels in the yolk sac which is attributed to the lack of supporting smooth muscle cells[446,568]. This implies that QKI full KO mutants are capable of developing passed the naïve-to-primed transition and points at a posterior defect leading to lethality.

On the binary lens through which we look at the naïve-to-primed transition

Cellular differentiation is a continuous process. This poses a fundamental challenge in biology as it defies the categorical labels that are given to the different stages during differentiation. The prevailing view of cell types can be reduced to static snapshots of a continuous process. In the case of embryonic pluripotency, the distinction between two types of pluripotency was a conceptual landmark in the field, but it is essentially limited by the culture media that stabilizes the two cell states. Thus, there might be other intermediate cellular states that remain uncategorized simply because they are not stable *in vitro* or the proper conditions to maintain them are still unknown. Recently, a number of studies have defined naïve-to-primed intermediate cell states mostly based on *in vitro* experiments. These include the "naïve-primed intermediate" state[569], the "poised" state[544], the "formative" state[570] and the "rosette" state[571]. Importantly, there is no agreement on whether some of these states are the same or how much they diverge from one another. Additionally, their definition mainly relies on transcriptional differences despite there being many other cellular layers of complexity. As the field moves forward, researchers must find a way to unify a classification of these cell states and do so beyond their differences in gene expression. Here, we have provided a new layer of knowledge in the definition of naïve and primed states that paints a new picture of transcriptional complexity that was so far unexplored. However, much work still needs to be done to understand how the series of events that turn the naïve alternative splicing landscape into primed are temporally controlled and whether there are more cell types, and thus, more alternative splicing landscapes in between.

For all that, a pending question remains for future studies, much broader than naïve and primed pluripotency states, this is, how do we define a cell type?

Conclusions

Conclusions

1. The alternative splicing landscape is remodelled during the naïve-to-primed pluripotency transition.

2. Distinct genomic features characterize the exons alternatively spliced in the naïve-to-primed transition in mice as well as humans.

3. Depletion of the splicing regulator RBMXL2 in mESCs leads to the downregulation of the naïve pluripotency gene expression and alternative splicing profiles, the upregulation of primed pluripotency profiles, and causes a faster exit from both pluripotency states. On the contrary, simultaneous depletion of RBMX and RBMXL1 in mESCs leads to high cell death upon differentiation.

4. Loss of the splicing regulators ESRP1 and ESRP2 shifts the gene expression profile of mESCs towards a naïve pluripotency state whereas loss of RBM47 causes the opposite trend. However, neither the depletion of ESRP family members nor RBM47 affect pluripotency exit dynamics.

5. The splicing regulator QKI regulates the alternative splicing of a subset of the naïve-to-primed differential cassette exons.

Materials and methods

ESC maintenance

ESCs were grown on 0.1% gelatine-coated (Millipore, ES-006-B) plates (Thermo Scientific Nunc Cell-Culture Treated Multidishes, 140675) either in Serum-LIF (SL) media or 2-inhibitors-LIF (2iL) media. Gelatine coating was left for a minimum of 15 minutes to set.

SL media consisted on Glasgow's Minimum Essential Medium (GMEM) BHK-21 (Gibco, 21710-025) supplemented with 10% batch-tested foetal bovine serum (Seralab, A1060013 EU-000-H), 0.5mM Sodium Pyruvate (Gibco, 11360-070), 1mM L-Glutamine (Gibco, 25030-024), Minimum Essential Medium non-essential amino acids solution (Gibco,11140-050), 0.1mM 2-Mercaptoethanol (Millipore, ES-007-E) and 1000U/ml ESGRO recombinant mouse LIF protein (Millipore, ESG1107). SL ESCs were routinely dissociated into a single-cell suspension using 1x TrypLE Express Enzyme (Gibco, 12605028), split and replated after quenching with media and centrifuging at 300g for 3mins. For long term storage, ESCs were cryopreserved in FBS (Gibco, 10270106) and 10% Dimethyl sulfoxide Hybri-Max (Sigma Aldrich, D2650).

2iL media consisted on 50% Neurobasal medium (Thermo Fisher Scientific, 21103-049), 50% Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (Applied Biosystems, 21331-020), 0.5x N-2 supplement (Applied Biosystems, 17502-048), 0.5x B-27 supplement without vitamin A (Life Technologies, 12587-010), 1mM L-Glutamine (Gibco, 25030-024), 0.1mM 2-Mercaptoethanol (Millipore, ES-007-E), 1000U/ml ESGRO recombinant mouse LIF protein (Millipore, ESG1107), 1uM MEK inhibitor PD0325901 (Selleckchem, S1036) and 3uM GSK3-beta inhibitor CHIR99021 (Selleckchem, S1263). 2iL ESCs were routinely dissociated into a single-cell suspension with Accutase (Biolegend, 423201), split and replated after quenching with DPBS 1x (Gibco, 14190) and centrifuging at 300g for 3mins. For long term storage, 2iL ESCs were cryopreserved in N2B27 media (basal 2iL media, without inhibitors or LIF) and 10% Dimethyl sulfoxide Hybri-Max (Sigma Aldrich, D2650).

All 2iL ESCs in this study were propagated from SL ESCs for 5 or 6 passages.

EpiSC maintenance

EpiSCs were routinely grown on FBS (Gibco, 10270106) coated plates (Thermo Scientific Nunc Cell-Culture Treated Multidishes, 140675). For the first 10 passages of derivation from SL ESCs FBS coating is not used, instead, we use Human Plasma Fibronectin Purified Protein 10ug/ml (Millipore, FC010) in DPBS 1x (Gibco, 14190). Both plate coatings were left to set for a minimum of 1 hour at 37ºC and 5% CO_2.

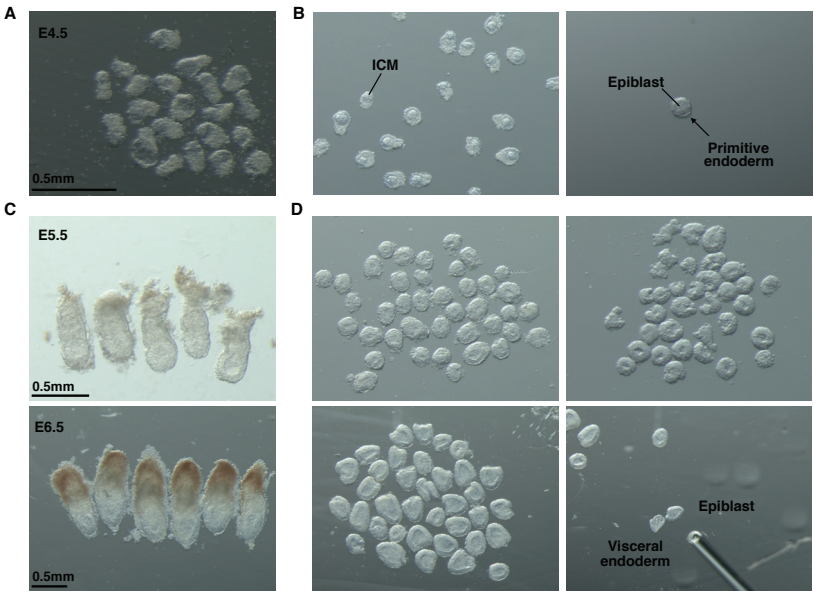
EpiSC media consisted on 50% Neurobasal medium (Thermo Fisher Scientific, 21103-049), 50% Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (Applied Biosystems, 21331-020), 0.5x N-2 supplement (Applied Biosystems, 17502-048), 0.5x B-27 supplement without vitamin A (Life Technologies, 12587-010), 1mM L-Glutamine (Gibco, 25030-024), 0.1mM 2-Mercaptoethanol (Millipore, ES-007-E), 20ng/ml of Recombinant Human/Mouse/Rat Activin A (R&D Systems, RYD-338-AC-050) and 12ng/ml Recombinant Mouse bFGF (R&D Systems, RYD-3139-FB-025). For passaging, EpiSC colonies were manually dissociated with fresh media into smaller clusters by scraping with a plastic pipette. Previous to dissociation EpiSCs are incubated for 3 minutes with either PBS1x, in case of routine passaging, or with 1mg/mL Collagenase Type IV (Stemcell Technologies, 7909), in the case of first 8 passages of derivation from ESCs.

For long term storage, EpiSCs were cryopreserved in N2B27 media (basal EpiSC media, without factors) and 10% Dimethyl sulfoxide Hybri-Max (Sigma Aldrich, D2650).

In all cases, EpiSCs were derived from SL ESCs by passaging over 16 times in EpiSC conditions.

Immunosurgery of E4.5 ICMs

Blastocyst embryos were collected at E4.5dpc, at this stage, the epiblast and primitive endoderm are fully segregated and thus ICM isolation by immunosurgery followed by epiblast dissection is feasible but laborious. We incubated whole E4.5 embryos at 37ºC in equilibrated 20% anti-mouse serum (Sigma, M5774) in embryo handling medium (N2B27- see above) for 1 hour. This was followed by incubation at 37ºC and 5% CO_2 in 20% complement sera from guinea pig (Merck, 234395) in embryo handling medium for 15



MM Figure 1: Procedure for *in-vivo* sample collection (A) E4.5 blastocysts **(B)** Two stages of the immunosurgery protocol for epiblast dissection. To the left, embryos after incubation with complement sera followed by incubation in culture media. Note the lysed cells surrounding the ICM. To the right, ICM after 4-hour incubation in hanging drops. Note how primitive endoderm cells grow around a central mass of epiblast cells. **(C)** E5.5 embryos (top) and E6.5 embryos (bottom). **(D)** Two stages of epiblast separation from the visceral endoderm in post-implantation embryos. E5.5 embryos (top) and E6.5 embryos (bottom).

minutes. Whole embryos were then transferred to a pre-equilibrated culture medium and incubated at 37°C and 5% CO_2 for 30 minutes, until the lysed cells separated from the ICM. At this point, residual trophectoderm was thoroughly removed by repetitive manual pipetting using a pulled Pasteur pipette just bigger than the ICM. ICMs were then incubated at 37°C and 5% CO_2 for 4 hours in hanging drops of equilibrated embryo handling medium. Finally, pulled Pasteur pipettes of smaller inner diameter were used to gently separate the primitive endoderm from the epiblast (MM Figure 1A, B). We obtained around 40– 50 cells per embryo and these were frozen immediately in RLT buffer for posterior RNA extraction.

Epiblast dissection of E5.5 and E6.5 embryos

Postimplantation epiblast samples were isolated by manual dissection in M2 medium (Sigma, M7167). All the embryos used had developed a cavity and the epiblast had formed an epithelium. Extraembryonic ectoderm was first removed with a wide mouth-controlled pipette. Then, the visceral endoderm was mechanically separated by drawing the embryo portion through a second pulled Pasteur pipette with a narrower aperture (MM Figure 1C, D). Postimplantation epiblast samples were then frozen immediately in RLT buffer for posterior RNA extraction. Visceral endoderm samples were also collected for E6.5 embryos and frozen immediately in RLT buffer for posterior RNA extraction.

CRISPR/Cas9 gene-editing

We followed the Ran et al [572] protocol to design, clone, test and employ all gRNAs used to generate each KO cell line in this thesis. For every final KO, we used two gRNAs that targeted the sequence of interest. For example, in the case of exon-specific deletions, each gRNA targeted one of the two flanking introns (see Table for further details on targeted sites). For every KO generation, the best gRNAs pair was chosen after testing the editing efficiency of six independent gRNAs in all possible combinations. For the generation of Rbm47/ESRP axis KOs, we employed the multiplex CRISPR/Cas9 assembly system[573]. This allowed us to construct all-in-one CRISPR/Cas9 vectors expressing multiple gRNAs for multiple gene KOs.

Lipofectamine 2000 (Invitrogen, 11668019) was used for transfection of 1ug of each CRISPR/Cas9 vector into ESCs plated on 6-well plates with ~70-80% confluency following manufacturer's instructions. Importantly, we always transfected empty vectors as appropriate controls for each experiment. Media was changed 5 hours after transfection. Selection with 1ug/ml puromycin (Sigma, P8833) started 24 hours after transfection and lasted ~5-6 days (conditioned by the extent of cell death in un-transfected controls) with daily media-replacement. After selection, cells were dissociated using 1x TrypLE Express Enzyme (Gibco, 12605028) and plated either on 96-well plates at a 0.6 cells/well concentration, or on 100mm dishes at 1000cells/dish concentration.

If plated on 96-well plates, media was changed after 6 days of re-plating. At day 12, cells in wells that had grown colonies were dissociated using 1x TrypLE

Express Enzyme and plated in parallel on to 48-well plates (for screening for positive clones by genotyping) and 24-well plates (for expanding the cell line if positive).

If plated on 100mm dishes, media was changed 4 days after re-plating. At day 7-9, ESC colonies were picked and replated. Briefly, media was replaced by DPBS 1x (Gibco, 14190), individual colonies were manually picked using a p20 pipette, placed onto drops of TrypLE Express Enzyme and disaggregated by pipetting. Individual disaggregated colonies were then plated on 96-well plates. Media was changed the next day and every other day. After 3-6 days, wells with grown colonies were dissociated using 1x TrypLE Express Enzyme and plated in parallel on to 48-well plates (for screening for positive clones by genotyping) and 24-well plates (for expanding the cell line if positive).

Efficient gene-editing was identified among ESCs clones by genotyping screening and positive clones were karyotyped, expanded and cryopreserved.

Genotyping screening for positive CRISPR/Cas9 gene-edited clones

Confluent ESCs in 48-well plates were dissociated using 1x TrypLE Express Enzyme (Gibco, 12605028), washed with DPBS 1x (Gibco, 14190) and pelleted by centrifuging at 300g for 3mins. Cell pellets were resuspended in 75ul of lysis buffer containing 50mM Tris-HCl pH 8-8.5 (in-house), 1mM EDTA (in-house), 0.5% Tween-20 (Merck, P9416) and 8ug/ml Proteinase K (Thermo Scientific, EO0491) and incubated at 55ºC for a minimum of 5 hours or overnight. Samples were then heated at 95ºC for 10 minutes for proteinase K inactivation and stored at 4ºC (or at -20ºC for long-term storage). Cell lysis was then employed to genotype each clone by PCR using primers flanking gRNA binding sites.

GoTaq G2 Flexi DNA Polymerase (Promega, M7806) was used for standard genotyping PCR reactions using 3ul of lysed cell solution and a total volume of 25ul for each reaction. PCRs were performed in a thermal-cycler (Applied Biosystems, 2720) using the following touch-down run-program: Initial denaturation step at 94ºC for 3 minutes; 10 cycles of 30 seconds at 94ºC, 30 seconds annealing at temperatures from 63ºC decreasing 0.5ºC each cycle and 2 minutes elongation at 72ºC; 25 cycles of 30 seconds at 94ºC, 30 seconds annealing at 58ºC and 2 minutes elongation at 72ºC; and a final elongation step of 5 minutes at 72ºC. The resulting PCR product was size-fractioned by 1%

ultrapure agarose (Invitrogen, 16500500) gel electrophoresis containing SYBR Safe DNA Gel Stain (Invitrogen, S33102) and visualized under blue light.

Depending on the resulting PCR products, gel bands were then extracted from the gel using GeneJET Gel Extraction Kit (Thermo, K0691) or PCR products were directly purified using QIAquick PCR Purification Kit (Qiagen, 28104), following manufacturer's instructions. The resulting DNA fragments were sequenced by Sanger sequencing using BigDye Direct Cycle Sequencing Kit (Applied Biosystems, 4458689) to confirm deletions.

Karyotyping of ESC lines

ESCs of ~70% confluency were passaged to a 6-well plate with a splitting ratio of 1:2 to 1:4. The next day, the media was changed and 2 hours later 0.2ug/ml of KaryoMax Colcemid Solution (Gibco, 15212012) was directly added to the media. After a 1-hour incubation at 37°C and 5% CO_2, cells were dissociated using 1x TrypLE Express Enzyme (Gibco, 12605028), washed with DPBS 1x (Gibco, 14190) and pelleted by centrifuging at 300g for 3mins. Pelleted cells were resuspended gently in 5ml of ice-cold hypotonic 0.56% KCl solution, by inverting once, and were kept at room temperature for 6 minutes followed by cell-pelleting by centrifuging at 300g for 3mins. Cell pellets were then resuspended in a drop of supernatant and then by dropwise adding 2ml of ice-cold freshly-made fixative consisting of Methanol: Glacial Acetic Acid (3:1) (Merck, 1060182500) and mixing continuously with a vortex. Cells were pelleted by centrifuging at 300g for 4mins and the fixation step was repeated once more. Cell pellets were then resuspended in 1ml of fixative. Metaphase chromosome spreads were made by pipetting drops of this cell suspension from a height of 1 meter onto fixative-washed slides. Slides were air-dried and coverslips were mounted using Fluoroshield with DAPI (Sigma, F6057). Slides were examined under an oil immersion microscope at 100x magnification. Images were taken and chromosomes were counted of at least 20 spreads per clone.

Clonal Monolayer differentiation assay

Each ESC clone was plated on three 0.1% gelatine-coated (Millipore, ES-006-B) 12-well plates at an 800 cell/well concentration on SL media and grown at 37°C and 5% CO_2 for a total of six days. Two days after plating, the media of

one of the wells was changed to EB media. Four days after plating, the media of another well was changed to EB media. Six days after plating, plates were collected, and cells were stained with Alkaline phosphatase staining.

Alkaline Phosphatase staining

Alkaline phosphatase staining was carried out using the Leukocyte Alkaline Phosphatase Kit (Merck, 86R). Briefly, ESCs grown on 0.1% gelatine-coated (Millipore, ES-006-B) plates (Thermo Scientific Nunc Cell-Culture Treated Multidishes, 140675) were fixed with citrate-acetone-formaldehyde fixative at room temperature for 30 seconds. Plates were then washed with deionized water for 45 seconds and incubated with an alkaline-dye solution (sodium nitrite solution, FRV-alkaline solution and Naphthol AS-BI alkaline solution at 1:1:1 and diluted in 45 volumes of distilled water) for 20 minutes at room temperature and protected from light. Plates were washed with distilled water for 2 minutes and air-dried.

Pictures of colonies were taken using an Olympus stereo microscope (DP73) with oblique illumination.

Embryoid body formation assays

ESCs were dissociated using 1x TrypLE Express Enzyme (Gibco, 12605028), and counted after quenching with SL media with no LIF and centrifuging at 300g for 3mins. ESCs were then seeded on 90mm non-gelatinised petri dishes (Fisher, 11309283) at a density of 300 thousand cells per plate in EB media. EB media consisted on Glasgow's Minimum Essential Medium (GMEM) BHK-21 (Gibco, 21710-025) supplemented with 10% FBS (Gibco, 10270106), 0.5mM Sodium Pyruvate (Gibco, 11360-070), 1mM L-Glutamine (Gibco, 25030-024), Minimum Essential Medium non-essential amino acids solution (Gibco,11140-050) and 0.1mM 2-Mercaptoethanol (Millipore, ES-007-E). In each experiment, 6 plates for each clone was plated. ESC samples were taken on the day of plating (day 0) and then every day until day 3. After that, samples were taken every other day until day 11. On day 5 the media was changed, briefly, the suspension cultures were collected onto 50 ml falcons and embryoid bodies were left to settle on the bottom of the falcon for 20 minutes at room temperature. The supernatant media was then removed and replaced with fresh media. On day 9, only half of the volume contained in the

plate is collected for sampling and replaced with fresh media. The remaining embryoid bodies are collected on day 11.

For all sample collection, suspension cultures were collected, embryoid bodies were pelleted by centrifuging at 300g for 3mins, washed once in DPBS 1x (Gibco, 14190) and pelleted by centrifuging at 300g for 3mins. Cell pellets were frozen and stored at -80ºC until RNA or protein extraction.

RNA extraction

Total RNA was extracted with different kits according to different downstream applications, importantly, all samples from the same experiments and comparable experiments were extracted with a single kit. In the case of RNA extracted from *in-vitro* samples for basic cDNA synthesis and downstream RT-PCR or RTqPCR analysis, we used the NZY Total RNA Isolation kit (Nzytech, MB13402). In the case of RNA extracted from *in-vivo* samples for basic cDNA synthesis and downstream RT-PCR or RTqPCR analysis, we used Arcturus PicoPure RNA Isolation Kit (Applied Biosystems, KIT0204). For *in-vitro* RNA-seq samples, we used the RNeasy Plus Mini kit (Qiagen, 74136) whereas for *in-vivo* RNA-seq samples we used the RNeasy Plus Micro kit (Qiagen, 74034).

cDNA synthesis

Total RNA was reverse transcribed using SuperScript III Reverse transcriptase (Invitrogen, 18080044). For 0.5ug of total RNA, 5uM of anchored oligo$(dT)_{20}$ (Sigma, own design) and 1mM dNTP mix (Invitrogen, 10297018) were mixed in RNase-free water to a final volume of 10ul. After 5 minutes at 65ºC, samples were put in ice and a mix composed of 4ul of RT 5x first-strand buffer, 1ul 0.1M DTT, 20 U Ribolock RNase inhibitor (Thermo, EO0381), 4 ul of RNase-free water and 100 U SuperScript III Reverse transcriptase was added and mixed by pipetting gently. We set one reaction mixture without SuperScript III Reverse transcriptase as a negative control for every round of cDNA synthesis. The mixtures were incubated at 50ºC for 50 minutes and 70ºC for 15 minutes. The cDNA samples were then diluted 1:10 in RNase-free water and stored at -20ºC.

RT-PCR

GoTaq G2 Flexi DNA Polymerase (Promega, M7806) was used according to manufacturer's instructions, with 1ul of diluted cDNA and a total volume of 25ul for each reaction. PCRs were performed in a thermal-cycler (Applied Biosystems, 2720). For each validation of AS events, RT-PCR conditions could vary strongly depending on primers, length of amplicon and amount of template, however, here we state the standard RT-PCR conditions we used (with varying annealing times, depending on primers; and varying elongation times, depending on the amplicon size): Initial denaturation step at 94ºC for 3 minutes; 10 cycles of 30 seconds at 94ºC, 30 seconds annealing at temperatures from 63ºC decreasing 0.5ºC each cycle and 1-minute elongation at 72ºC; 25 cycles of 30 seconds at 94ºC, 30 seconds annealing at 58ºC and 1-minute elongation at 72ºC; and a final elongation step of 5 minutes at 72ºC.

The resulting PCR product was size-fractioned by 2.5% ultrapure agarose (Invitrogen, 16500500) gel electrophoresis containing SYBR Safe DNA Gel Stain (Invitrogen, S33102). However, the percentage of agarose used for the gel varied depending on the DNA sizes to separate. Although 2.5% was the standard, if the size to separate was smaller than 30bp, we used 3% agarose gels. Gels were run in 1x TBE (89mM Tris, 89mM boric acid, 2mM EDTA) at ~80-100V for approximately 40 minutes and visualized under blue light in a GelDoc transilluminator (Biorad). Band intensities were subsequently quantified using Image J.

RTqPCR

Quantitative real-time PCRs were performed using a LightCycler 480 (Roche) and 384-well plates. All primers were designed to anneal to adjacent exons where possible to amplify only mature mRNA. Additionally, all generated amplicons of 100 – 200nts in length and all were tested for correct amplification. Reactions were done in a total volume of 10ul and contained 2x NZYSpeedy qPCR Green Master Mix (NZYtech, MB22403), 750nM primers and 1ul of diluted cDNA template. The run-programme used was as follows: Initial denaturation step at 95ºC for 10 minutes; 45 amplification cycles of 15 seconds at 95ºC, 20 seconds at 60ºC and 10 seconds at 72ºC with a single acquisition; melting of 30 seconds at 95ºC, 1 minute at 68ºC and increase to 98ºC with continuous acquisition followed by 30 seconds cooling at 40ºC. Each reaction was performed in duplicate. The output Ct values were normalized relative to

the expression of Beta-actin (Actb) as housekeeping gene and calculated using the delta-Ct method. All RTqPCRs performed were done in parallel using negative controls where cDNAs reactions lacked reverse transcriptase.

Stranded RNA-seq library preparation and sequencing

The *in-vivo* RNA sequenced samples were obtained in various days from a total of 22 pregnant mice. Each stage was then pooled together before extracting RNA. Overall, we isolated 58 E4.5dpc epiblasts, 73 E5.5dpc epiblasts and 60 E6.5dpc epiblasts. All these *in-vivo* RNA sequenced samples were obtained in collaboration with Professor Jennifer Nichols and all embryos were dissected at the Cambridge Stem Cell Institute.

RNA-seq library preparations were performed by the CRG Genomics Core Facility using the TruSeq Stranded mRNA Sample Prep Kit v2 (Illumina, RS-122-2101/2) according to the manufacturer's instructions. Briefly, 500ng of total RNA was used for poly(A)-mRNA selection using streptavidin-coated magnetic beads and were subsequently fragmented to approximately 300bp. cDNA was synthesized using random primers and SuperScript II reverse transcriptase (Invitrogen, 18064-014). Instead of dTTP, the second strand of cDNA incorporated dUTP. Double-stranded DNA was then A-tailed and barcoded Truseq adapters were ligated. AMPure XP beads were used for subsequent purification steps. Finally, libraries were evaluated for quantity and size distribution using Agilent DNA 1000 chip and quantified by RTqPCR with KAPA Library Quantification Kit (KapaBiosystems, KK4835) before amplification with Illumina's cBot. 20pM of each library was loaded onto a flow cell and was sequenced at 4 samples per lane following the 2 x 125bp paired-end protocol on Illumina's HiSeq 2500.

SPAR-seq

The Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing (SPAR-seq) protocol followed in this thesis is detailed by Han et. al [439]. These experiments were done in collaboration with the Blencowe lab (University of Toronto) and the siRNA knockdown library used in this thesis is the same one used in Han et al [439]. Briefly, a total RNA pool was extracted from 1536 siRNA knockdown and control treatments that were performed with two biological replicates in SL ESCs (CGR8 line), at the

Lunenfeld-Tanenbaum Research Institute (LTRI) SMART robotics facility. Specifically, SMART pool siRNAs (Dharmacon) targeted 1416 genes including 654 splicing- and RNA- associated factors (consisting of all known splicing factors, spliceosome-associated proteins and known and predicted RBPs); 858 chromatin and transcription-associated factors (including previously reported factors involved in splicing, chromatin associated proteins, proteins with known or predicted chromatin binding/remodelling/modifying domains); and 65 signalling and post-translational associated factors related to splicing regulation. Additionally, siRNAs targeted 32 positive controls (simultaneous knockdown of Mbln1 and Mbln2) and 88 negative controls (including non-targeting siRNAs, mock controls and untreated samples). Please refer to Han et al [439] for a complete list of all targeted genes and knock-down efficiencies. Cell plating, siRNA transfection and RNA purification were all automated. In short, SL ESCs were seeded in 96-well plates at 3000 cells per well, 2 days after, they were transfected with siRNAs, and 2 days after transfecting, total RNA was purified using RNeasy Plus 96 Kit (Qiagen).

In total, 32 96-well plates were used for SPAR-seq. For each well, a multiplex RT-PCR was applied to simultaneously amplify 44 transcript regions spanning various exons which assessed naïve-to-primed alternatively spliced exon inclusion, intron retention and gene expression. In the case of assessing exon inclusion and intron retention, primers were designed near splice site junctions to monitor specific splice variants. In some, cases more than one AS event was monitored within the same amplified region. The event-specific primer pool composition was optimized using several rounds of MiSeq (Illumina) with RNA extracted from 2iL ESCs, SL ESCs and EpiSCs. All primers contained a 5' universal adaptor sequence. In total, a pool of 88 primers was added to each reaction at a final concentration of 0.025uM for each primer. RT-PCRs were performed using the OneStep RT-PCR kit (Qiagen, 509210212) with ~40-60ng of RNA per 20ul reaction in 96-well Thermal Cyclers (Applied Biosystems) with the following program: 30 minutes at 50°C, 15 minutes at 95°C, 30 cycles of 40 seconds at 94°C, 1 minute at 58°C, 3 minutes at 72°C and 10 minutes at 72°C.

Unique dual-index barcodes were used for multiplex barcode sequencing: unique 8nt reverse barcodes were used for each sample and forward 8nt barcodes were used to mark each half of each 96-well plate. These forward and reverse barcodes were added after universal adaptor sequences to the RT-PCR amplicons during a second PCR using Phusion High-Fidelity DNA polymerase

(Thermo Scientific, F530) following manufacturer's instructions and employing 1ul of the multiplex RT-PCR product per 20ul reaction. Barcoded sequencing libraries were then pooled and sequenced at the Donnelly Sequencing Centre using an Illumina HiSeq 2500 with 4 lanes at 768 samples per lane following the 2 x 125bp paired-end protocol.

For the analysis and identification of candidate regulators of naïve-to-primed events, we first mapped, profiled and compared the alternative splicing events across all samples using manually annotated exon-exon and exon-intron junctions. We were able to detect 44 splicing events occurring across our samples. To correct for variation among plates, we normalized PSIs differently depending on the value of median PSI of the event in all replicates and the value of the median PSI of that event in each plate.

For a given event in each sample, if the value of the median PSI across all plates was greater than the value of the median PSI in that plate, we calculated the normalized PSI with the following formula:

$$PSI_{Normalized} = \frac{\Big(\big(100 \times \big(Median_{all} - Median_{plate} + PSI\big)\big) - \big(PSI \times Median_{all}\big)\Big)}{\big(100 - Median_{plate}\big)}$$

If on the other hand, the value of the median PSI in that plate was greater than the value of the median PSI across all plates, we calculated the normalized PSI with the following formula:

$$PSI_{Normalized} = \frac{PSI \times Median_{all}}{Median_{plate}}$$

Candidate regulators were identified using normalized PSI levels of each event by selecting the top 50 samples that most changed inclusion levels and top 50 samples that most changed exclusion levels. Candidates were filtered to guarantee that both replicates were consistent, only candidates that had both replicates in the top 50 were selected. We selected those that appeared to be regulating at least 2 or more naïve-to-primed events and that only changed events towards the isoforms found in one pluripotency state or the other.

Western Blot

Total cell extracts were collected directly from plates on ice by scraping using m-PER Mammalian Protein extraction reagent (Thermo Scientific, 78501) supplemented with 1 tablet per 10ml of Complete Mini EDTA-free Protease Inhibitor Cocktail tablets (Sigma, 11836170001) and 1 tablet per 10ml of Phostop phosphatase inhibitor tablets (Sigma, 4906845001) after washing with cold DPBS 1x (Gibco, 14190). Protein extracts were quantified using the Pierce BCA Protein Assay Kit (Thermo, 23227) and adjusted for equal amounts of total protein concentration in each sample. Laemmli buffer 4x containing 40% Glycerol (Merck, G5516), 250mM Tris-HCl pH6.8 (in-house), 8% SDS (in-house), 0.1% Bromophenol blue (Sigma, B0126) and 10% 2-Mercaptoethanol (Sigma, M6250) was then added to all samples.

After denaturation of samples at 95ºC for 10 minutes, 9-15ug of total protein extracts were loaded and separated in 4–15% Criterion TGX Precast Midi Protein Gels (Biorad, 5671085) at 80-120V by gel electrophoresis in running buffer (in-house) containing 25mM Tris, 192mM glycine and 0.1% SDS. Proteins were transferred into an Immun-Blot PVDF membrane (Biorad, 1620177) using wet transfer at 100V for 90 minutes kept at 4ºC in transfer buffer (in-house) containing 25mM Tris, 192mM glycine and 20% Methanol. Membranes were stained with Ponceau solution to confirm uniform transfer. Subsequently, they were blocked in 5% skimmed milk (Sigma, 70166) diluted in TBS-T (10mM Tris-HCl pH7.4, 100mM NaCl, 0.1% Tween-20) for a minimum of 30 minutes, before incubation with primary antibodies overnight at 4ºC. Later, membranes were washed with TBS-T three times before incubating them with horseradish peroxidase (HRP)-conjugated secondary antibodies for one hour at room temperature. Membranes were washed in TBS-T three times and chemiluminescence was assayed using Immobilon Forte Western HRP substrate (Merck, WBLUF0100) and an Amersham 600 Imager (GE Healthcare Life Sciences). All antibodies were diluted in 5%Milk TBS-T.

When needed for re-blotting, membranes were mildly stripped using a mild stripping buffer at pH 2.2 composed of 1,5% glycine (Sigma, G8898), 0.1% SDS (Sigma, L3771) and 1% Tween-20 (Merck, P9416). Membranes were incubated with stripping buffer for 10 minutes at room temperature, stripping buffer was then replaced and membranes were incubated further for 5 minutes. Membranes were then washed in PBS twice for ten minutes and in TBS-T twice for 5 minutes. Then membranes were ready to block and re-blot.

For Phos-tag western blots, we used SuperSep Phos-tag 12.5% Precast Gels (Fujifilm, 195-17991). Gel electrophoresis were run at constant 30mA for 45 minutes in running buffer (see above). The gels were then washed in 0.01mM EDTA transfer buffer three times for 20 minutes at room temperature for the elimination of metal ions, and once with transfer buffer without EDTA for 10 minutes more. Then proteins were transferred to an Immun-Blot PVDF membrane using wet transfer at 100V for 1 hour kept at 4°C in transfer buffer (see above) and western blot developing protocol was followed as normally (see above).

Immunofluorescence

ESCs were grown on gelatine-coated glass coverslips of 13mm diameter (VWR, 631-1578) placed on 22mm plates, and EpiSCs, on 10ug/ml fibronectin-coated glass coverslips (see cell maintenance section above for further details). Plates were washed once with PBS (in-house) and fixed in 4% Paraformaldehyde in PBS (Merck, 158127) for 10 minutes. Cells were washed again with PBS and then permeabilized using 0.4% Triton X-100 in PBS (Sigma, X100) for 5 minutes. Cells were washed again with PBS and blocked in 10% BSA (Sigma, A2153) 0.1% Triton X-100 in PBS for 30 minutes at room temperature. Coverslips were then withdrawn from plates and incubated overnight at 4°C in a wet chamber with primary antibodies in 1% BSA, 0.1% Triton X-100 in PBS. The following day, coverslips were washed twice in PBS and incubated with secondary antibodies in 1% BSA, 0.1% Triton X-100 for 45 minutes in a wet chamber in the dark. Subsequently, coverslips were washed twice in PBS, excess PBS was removed and coverslips were mounted using Fluoroshield with DAPI (Sigma, F6057). Samples were visualized using a Leica TCS SP8 with 40x oil immersion objective. For each plane, 10 z-stacks were taken with a step size of 2um. Images were analysed using ImageJ.

Sample-to-sample correlation distance matrices and principal component analysis

Read alignment and gene-level quantification were done using *vast-tools* (based on bowtie[574]) for our starting dataset: 3 cell lines 46C, E14 and DizFx in 3 culture conditions 2iL, SL and AF. Gene levels counts were processed using edgeR[575] for filtering and normalization following the standard pipeline. Principal component analysis (PCA) was done using the top 1000 most variable genes and plotted using the R package "factoextra". Spearman correlation was

calculated pairwise for our 9 samples using gene expression values and plotted as a heatmap, using "1-correlation" as the distance measure for clustering samples.

For global alternative splicing (AS) analyses an equivalent procedure was performed. Briefly, AS quantification was done using *vast-tools* and events were filtered by a minimum level of variability across samples of 10% PSI. We performed PCA and sample-to-sample correlation clustering using the same approach as above independently for AS exons and alternatively retained (AR) introns, with roughly 3000 events for both AS exons and AR introns.

Differential Gene Expression (DGE) analysis

Gene counts were used as input to edgeR for differential gene expression analysis. DGE was calculated for the 2iL vs EpiSC comparison using the quasi-likelihood (QL) F-test and a design matrix taking into consideration cell culture conditions and the cell line (~ culture condition + line). P-values were corrected using Benjamini & Hochberg method to generate FDR (false discovery rate) values. We used a 0.01 FDR cut-off to call differentially expressed genes.

Naïve and primed differential alternative splicing analysis using VAST-tools

Read alignment, alternative splicing profiling and comparison was made using Vertebrate Alternative Splicing and Transcription Tools (*vast-tools*)[311,380] version 2.3.0. Strand-specific mappings to mm10 or hg19 libraries were performed. Actual reads, meaning reads before mappability correction, were considered for the profiling of each alternative splicing event. For the definition of alternatively spliced cassette exons, we established that a minimum of either 10 actual reads all mapping to exclusion splice junctions or 10 reads mapping to one of the two inclusion splice junctions and 5 to the other, were necessary for the events to be considered (VLOW quality score). For retained introns defined with our dataset, we considered only those with a minimum of either 10 actual reads mapping to skipping splice junctions or 10 actual reads mapping to one of the inclusion exon-intron junctions and 5 actual reads mapping to the other (VLOW quality score). For the definition of alternative splice site usage (alternative 3' splice site or alternative 5' splice site), we established that a minimum of 15 actual reads mapping to the sum of all the splice junctions

were necessary for the events to be considered (VLOW quality score). Percent spliced in (PSI), percent of intron retention (PIR) and percent splice site usage (PSU) for each sample were calculated.

We used the analysis of our starting dataset (3 cell lines 46C, E14 and DizFx in 3 culture conditions 2iL, SL and AF) for establishing the differential naïve-to-primed alternatively spliced event set and for subsequent analyses. Samples were paired according to the cell line. Unless stated otherwise in the text, determination of differential alternatively spliced event sets was done by comparing non-overlapping PSI distributions across samples with a minimum difference of 15% PSI between paired samples, if possible, and a minimum range of 5% PSI considering all samples in each condition. Differentially retained introns were defined using binomial testing across samples[576]. Finally, for the definition of differential alternative splice site usage, only alternative splice sites of host exons with a minimum 25% PSI were considered.

Protein impact prediction of alternative splicing event usage

The impact of the inclusion or exclusion of differential alternatively spliced exons was predicted according to their position in either the coding sequence (CDS), untranslated regions (either 5' or 3' UTRs) or non-coding RNA. Exons located in the coding sequences were then predicted to be disruptive or CDS-preserving. Alternative exons were considered disruptive if either their exclusion or inclusion supposed a frameshift in the open reading frame (ORF); if the included exon contained a premature stop codon predicted to undergo non-sense mediated decay (NMD) meaning it was located over 50nt upstream an exon-exon junction; or if the included exon contained a premature stop codon that supposed a truncation of the protein by more than 300 amino acids. Alternative exons were considered CDS-preserving if none of the above cases was met and either the number of nucleotides in the exon was multiple of three, thus preserving the open reading frame (ORF); if the exon contained a start codon; if the exon contained a stop codon; or if the exon contained a premature stop codon not predicted to undergo non-sense mediated decay (NMD).

Transcriptome-wide alternative exon feature scores

To assess the genomic features associated with exons regulated in the naïve-to-primed transition, we grouped exons genome-wide into 5 groups. The first two groups correspond to exons classified as differentially spliced in the 2iL-vs-EpiSC comparison, divided by their direction of change: "higher in 2iL ESCs" or "higher in EpiSCs". We compared these exons to three reference groups of exons: alternative exons with a PSI value between 10% and 90% in at least one sample and that do not change (dPSI < 3%) in the 2iL-vs-EpiSC comparison; "cryptic" exons with a maximum PSI value across our samples lower than 10%, and "constitutive" exons with a minimum PSI of 90% or higher across all compared samples. For the latter two categories, one 1000 random exons were selected for genomic feature analyses.

For the splicing analysis of all alternative exons grouped by their genomic features, we used all alternative exons genome-wide as defined above, without any dPSI filtering. For each genomic feature we generated six equally sized groups or sextiles and quantified the dPSI between EpiSCs and 2iL ESCs and plotted the distribution of dPSI values.

For summarizing the effect of each genomic feature on the dPSI between EpiSCs and 2iL ESCs, we computed a single score per feature corresponding to the dPSI difference (ddPSI) between the first and last quantiles of each feature. For exon length scoring, the median ddPSI between exons in the 250nt to 200nt group and the shorter-than-50nts group was calculated. For intron length scoring, the difference in dPSI between exons with average flanking introns longer than 10kb and those with introns shorter than 200nts was calculated. For exonic GC content scoring, the ddPSI between exons with GC content over 58% and exons with GC lower than 40% was estimated, and for intron GC content, the limits were set at 50% and 33% respectively.

Exon-intron definition score was calculated for each exon considering exon GC-content together with the difference in intron and exon lengths. More specifically, we estimated relative values for each feature by making use of the categories defined above for GC-content and exon length. For exon GC content relative score (GCC_{exon}), exons in the "over 58%" category scored 1 point; exons in the "58% to 50%" category scored 2 points; exons in the "50% to 45%" category scored 3 points, exons in the "45% to 40%" category scored 4 points and exons in the "less than 40%" category scored 5 points. For exon

length relative score ($Length_{exon}$), exons in the "350nt to 250nt" and "250nt to 200nt" categories scored one point, the rest, scored 2 points. For exon-intron difference length relative score ($Length_{diff}$), first we calculated a value for the ratio of the exon length and the average length of the flanking upstream and downstream introns ($ExonIntron_{length\ diff}$) with the following formula:

$$ExonIntron_{length\ diff} = Log_{10}\left(\frac{exon_{length}}{average\ flanking\ intron_{length}}\right)$$

Subsequently, the length relative score ($Length_{diff}$) was calculated for each exon using the exon-intron differential length value ($ExonIntron_{length\ diff}$): if it was over -0.25, $Length_{diff}$ was 1 point; if it was between -0.25 and -0.75, $Length_{diff}$ was 2 points; if it was between -0.75 and -1.25, $Length_{diff}$ was 3 points; if it was between -1.25 and -1.5, $Length_{diff}$ was 4 points; if it was between -1.5 and -1.75, $Length_{diff}$ was 5 points; and if it was below -1.75, $Length_{diff}$ was 6 points. Finally, the exon-intron definition score was calculated following the formula:

$$Exon\ intron\ definition\ score = GCC_{exon} + Length_{exon} + Length_{diff} - 3$$

Gene set enrichment analysis (GSEA)

GSEA was performed using GSEA 4.0.0 software[426]. The gene set used to test enrichment in naïve markers and primed markers was established using our RNA-seq dataset. Specifically, lists of naïve and primed markers were generated using the differential gene expression profile in paired comparisons of 2iL vs. EpiSCs across our three independent cell lines (E14, 46C, DizFx) and E4.5 vs. E6.5 epiblasts. Differentially expressed genes were identified using VAST-tools. Briefly, the fold change in normalized cRPKMs (corrected-for-mappability Reads per Kbp and Million mapped reads[577]) between naïve and primed samples was calculated and 263 genes were established as naïve markers and 207, as primed markers. These markers had a minimum of 5-fold change of the average cRPKMs between the two groups and a minimum difference between each of the individual replicates of 2-fold. Additionally, markers had a minimum cRPKM of 5 in at least one of the compared groups.

Subsequently, we tested the enrichment of these markers within lists of differentially expressed genes in each of our comparisons (control vs. KO)

ranked by cRPKM fold change. Enrichment maps were generated to visualize GSEA results and normalized enrichment scores (NES) were calculated for each dataset comparison.

Alternative splicing set enrichment analysis (ASSEA)

ASSEA was performed using GSEA 4.0.0 software[426]. We used the differential naïve-to-primed alternatively spliced event set (see the section above on naïve and primed differential alternative splicing analysis using VAST-tools) to test for enrichment in primed included events and primed excluded events. We tested the enrichment of this alternative splicing exon switches within ranked lists of all alternative splicing exons that had coverage across all compared samples (KO and control) and which had PSI values between 10% and 90% in at least one of the samples. The exons in these ranked lists were ordered according to their difference in average PSI between samples (control vs. KO). Enrichment maps were generated to visualize GSEA results and normalized enrichment scores (NES) were calculated for each dataset comparison.

Consistency tests

Consistency tests evaluated the similarity among comparisons either of differential gene expression or differential alternative splicing events, as stated in the text. The values for each comparison was plotted on a scatter-plot (see Figure 6A) and the percentage of events occurring in the quadrants I and III were calculated in relation to the total number of events. We term this percentage as percentage of consistency.

For differential gene expression consistency tests comparing mouse and human naïve and primed datasets, differentially expressed genes were identified using VAST-tools. The fold change in normalized cRPKMs between mouse naïve and primed *in-vitro* paired samples was calculated. Differentially expressed genes were identified using a minimum of 5-fold change of the average cRPKMs between the two groups and a minimum difference between each of the individual replicates of 2-fold. Subsequently, the fold change in normalized cRPKMs between human naïve and primed samples was calculated and differentially expressed genes were identified using a minimum 2-fold change. Finally, human differentially expressed genes that were also differentially expressed in mouse were subjected to consistency tests.

In the case of alternative splicing consistency tests, differentially spliced events were identified in each group with a minimum difference of PSI of 15%, unless stated otherwise. Subsequently, consistency tests were done using events that were differentially spliced in both comparisons.

Gene Ontology term enrichment analysis

We used GOrilla software[578] to analyse the gene ontology (GO) enrichment of genes harbouring 2iL-vs-EpiSCs alternatively spliced exons. Specifically, we used the "two unranked lists of genes" module. As backgroup set, we used multiexonic genes meeting similar read coverage criteria as the genes with differentially spliced events. Statistical significance of GO-term enrichment was defined using p-values. Similar GO terms were joined together using semantic similarity as implemented by REViGO[579].

Orthologue exon identification

Human orthologue exon identification of mouse naïve-to-primed alternatively spliced cassette exons was done using the alternative splicing event-level orthology from VastDB[311]. Briefly, coordinate conversions were performed using Galaxy Liftover (from hg19 to mm9). Mouse exons with a LiftOver hit in human are evaluated for the presence of at least a canonical splicing acceptor site or donor site in the flanking introns (allowing for non-conserved 3' or 5' splice sites), if positive, these then were considered to be genome conserved. These exons were then examined for alternativity as defined in VastDB (events having PSI levels between 10% and 90% in at least 10% of tissue samples with coverage or with a 25% PSI range).

External human pluripotent samples, mouse *in-vivo* samples and differentiated tissue datasets

Human naïve and primed pluripotency RNA-seq samples were paired within each study, all datasets used and the corresponding samples are specified bellow. Mouse *in-vivo* published RNA-seq datasets were taken from Boroviak et al[248]. Differentiated tissue RNA-seq datasets were taken from VastDB[311].

Target protein	Usage	Host	Brand & reference
KLF4	1:200	Rabbit	Sigma, HPA002926
NANOG	1:500	Rat	eBioscience, 14-5761-80
OCT4	1:100	Mouse	Santa Cruz, sc5279
Rabbit IgG (Alexa Fluor 647)	1:300	Goat	Thermo, A32733
Rat IgG (Alexa Fluor 488)	1:300	Goat	Thermo, A11006
Mouse IgG (Alexa Fluor 594)	1:300	Donkey	Thermo, A32744

MM Table 1: Antibodies used for Immunofluorescence experiments

Target protein	Usage	Host	Brand & reference
KLF4	1:1000	Rabbit	Sigma, HPA002926
NANOG	1:250	Rat	eBioscience, 14-5761-80
OCT4	1:250	Mouse	Santa Cruz, sc5279
RBM47	1:500	Rabbit	Novus, NBP1-83664
ESRP1	1:500	Rabbit	Sigma, HPA023720
QKI	1:1000	Rabbit	Abcam, ab126742
OTX2	1:1000	Goat	R&D Systems, AF1979-SP
TUBULIN	1:2000	Mouse	Cell Signaling, 3873S
E-CADHERIN	1:500	Rat	Sigma, U3254
TIA1	1:500	Goat	Santa Cruz, sc-1751
RBFOX2	1:300	Rabbit	Bethyl, A300-864A
Rabbit IgG (HRP)	1:10000	Goat	Cell Signalling, 7074S
Mouse IgG (HRP)	1:10000	Goat	Abcam, ab205719
Rat IgG (HRP)	1:5000	Chicken	Santa Cruz, sc-2956
Goat IgG (HRP)	1:5000	Mouse	Santa Cruz, sc-2354

MM Table 2: Antibodies used for Westernblot experiments

SRA code	Cell line & Culture	State	Study
SRR1736688	WIS2_NHSM4i	naïve	Irie 2015
SRR1736689	WIS2	primed	Irie 2015
SRR1736690	WIS2	primed	Irie 2015
SRR2297451	H1_NHSM4i	naïve	Sperber 2015
SRR2297450	H1_NHSM4i	naïve	Sperber 2015
SRR1561941	H1	primed	Sperber 2015
SRR7460677	RH5_2a2iL	naive	Taei 2018 (a)
SRR7460678	RH5_2a2iL	naive	Taei 2018 (a)
SRR7460680	RH5	primed	Taei 2018 (a)
SRR7460675	RH6_2a2iL	naive	Taei 2018 (b)
SRR7460676	RH6_2a2iL	naive	Taei 2018 (b)
SRR7460679	RH6	primed	Taei 2018 (b)
SRR5151102	NK2_5iLA	naive	Collier 2017
SRR5151103	NK2_5iLA	naive	Collier 2017
SRR5151104	NK2_5iLA	naive	Collier 2017
SRR5151105	NK2	primed	Collier 2017
SRR5151106	NK2	primed	Collier 2017
SRR5151107	NK2	primed	Collier 2017
SRR1660342	ELF1_3iL	naive	Grow 2015
SRR1660343	ELF1_3iL	naive	Grow 2015
SRR1660344	ELF1_3iL	naive	Grow 2015
SRR1660345	ELF1	primed	Grow 2015
SRR1660346	ELF1	primed	Grow 2015
SRR1660347	ELF1	primed	Grow 2015

MM Table 3: Human RNAseq datasets

Lines	Clone	Deletion		Karyotype
Empty vector	px3			80%
	px5			70%
RbmxL2 KOs	2.1	1048bp deletion in CDS		33%
	2.8	1047bp deletion in CDS		60%
	2.5	998bp deletion + 6 substitutions in CDS		62%
	G1			60%
	H1			81%
Rbmx exon KOs	E1	1619 bp deletion in exon 9 & flanking introns		55%
	B2	1604 bp deletion in exon 9 & flanking introns		53%
Rbmx RbmxL1 DKOs	H3	Rbmx	4307bp deletion in exons & introns	75%
		RbmxL1	1bp insertion + 3bp deletion in CDS	
	A2	Rbmx	4296bp deletion in exons & introns	70%
		RbmxL1	566bp & 4bp deletions in CDS + 2 substitution	
Rbmx RbmxL2 DKOs	B1	Rbmx	4299bp deletion in exons & introns	72%
		RbmxL2	998bp deletion + 6 substitutions in CDS	
	D1	Rbmx	2905bp deletion in exons & introns	67%
		RbmxL2	998bp deletion + 6 substitutions in CDS	

MM Table 4: CRISPR-Cas9 mESC knockout clones generated for Rbmx/RbmxL2/RbmxL1 Project

Lines	Clone	Deletion		Karyotype
Empty vector	px1			62%
	px4			60%
Rbm47 KOs	73.1	754bp in CDS		59%
	73.7	753bp in CDS		55%
Esrp1 KOs	74.1	570bp deletion in CDS & introns		81%
	74.3	572bp deletion in CDS & introns		63%
Esrp2 KOs	75.1	918bp deletion in CDS & introns		61%
	75.2	918bp deletion in CDS & introns		65%
Rbm47 Esrp1 DKOs	A26	Rbm47	752bp in CDS	70%
		Esrp1	572bp deletion in CDS & introns	
	A28	Rbm47	752bp in CDS	70%
		Esrp1	554bp deletion in CDS & introns	
Esrp1 Esrp2 DKOs	B3	Esrp1	554bp deletion in CDS & introns	62%
			556bp deletion in CDS & introns	
		Esrp2	909bp deletion in CDS & introns	
	B11	Esrp1	572bp deletion in CDS & introns	67%
		Esrp2	906bp deletion in CDS & introns	
			907bp deletion in CDS & introns	
Esrp1 Esrp2 Rbm47 TKOs	C11	Rbm47	752bp in CDS	59%
			758bp in CDS	
		Esrp1	1297bp deletion in CDS & introns	
		Esrp2	2747bp deletion in CDS & introns	
	C14	Rbm47	753bp in CDS	55%
		Esrp1	1149bp deletion in CDS & introns	
		Esrp2	909bp deletion in CDS & introns	

MM Table 5: CRISPR-Cas9 mESC knockout clones generated for Esrp1/Esrp2/Rbm47 Project

Plasmid	Primer #	Primer name	Sequence	Back-bone
bVR2	45	RbmxEx2_2F	caccgCACCAATAAAGAGCTTTCCT	px459
	46	RbmxEx2_2R	aaacAGGAAAGCTCTTTATTGGTGc	
bVR8	57	RbmxEx9_2F	caccgCTTCCACCATAAGATGGCGG	px459
	58	RbmxEx9_2R	aaacCCGCCATCTTATGGTGGAAGc	
bVR12	65	RbmxInt8_3F	caccgTTCCCTCTTATGGCCAGCAA	px459
	66	RbmxInt8_3R	aaacTTGCTGGCCATAAGAGGGAAc	
bVR15	71	RbmxInt9_3F	caccgGACACTTAGAGGCATTAATG	px459
	72	RbmxInt9_3R	aaacCATTAATGCCTCTAAGTGTCc	
bVR 17	3	RmbxL2_2F	caccgCAACCTCGAGACCGACGAGA	px459
	4	RmbxL2_2R	aaacTCTCGTCGGTCTCGAGGTTGc	
bVR 22	13	RmbxL2_7F	caccgCTGTAGGAATCCCGCGGCGC	px459
	14	RmbxL2_7R	aaacGCGCCGCGGGATTCCTACAGc	
bVR44	89	Rbm47_1F	caccgGAAGACCTCACAGCCACGCT	px330A-1x6
	90	Rbm47_1R	aaacAGCGTGGCTGTGAGGTCTTCc	
bVR47	95	Rbm47_4F	caccgGCTGGCCAAGCCCGTGGACA	px330S2
	96	Rbm47_4R	aaacTGTCCACGGGCTTGGCCAGCc	
bVR52	105	Esrp1_3F	caccgGTTCCATCTTGCTGCTGTGA	px330S3
	106	Esrp1_3R	aaacTCACAGCAGCAAGATGGAACc	
bVR53	107	Esrp1_4F	caccgGTTCTTTCCTAGTGAAATGT	px330S4
	108	Esrp1_4R	aaacACATTTCACTAGGAAAGAACc	
bVR58	117	Esrp2_3F	caccgGGCCTTTTACTTCCCAGGCA	px330S5
	118	Esrp2_3R	aaacTGCCTGGGAAGTAAAAGGCCc	
bVR59	119	Esrp2_4F	caccgGAAGGTCACAAGTCAGCAAG	px330S6
	120	Esrp2_4R	aaacCTTGCTGACTTGTGACCTTCc	
bVR66	89	Rbm47_1F	caccgGAAGACCTCACAGCCACGCT	px330A-1x2
	90	Rbm47_1R	aaacAGCGTGGCTGTGAGGTCTTCc	
bVR67	105	Esrp1_3F	caccgGTTCCATCTTGCTGCTGTGA	px330A-1x2
	106	Esrp1_3R	aaacTCACAGCAGCAAGATGGAACc	
bVR68	117	Esrp2_3F	caccgGGCCTTTTACTTCCCAGGCA	px330A-1x2
	118	Esrp2_3R	aaacTGCCTGGGAAGTAAAAGGCCc	
bVR69	107	Esrp1_4F	caccgGTTCTTTCCTAGTGAAATGT	px330S2
	108	Esrp1_4R	aaacACATTTCACTAGGAAAGAACc	
bVR70	119	Esrp2_4F	caccgGAAGGTCACAAGTCAGCAAG	px330S2
	120	Esrp2_4R	aaacCTTGCTGACTTGTGACCTTCc	

Plasmid	Primer #	Primer name	Sequence	Back-bone
bVR71	117	Esrp2_3F	caccgGGCCTTTTACTTCCCAGGCA	px330A-1x4
	118	Esrp2_3R	aaacTGCCTGGGAAGTAAAAGGCCc	
bVR72	89	Rbm47_1F	caccgGAAGACCTCACAGCCACGCT	px330A-1x4
	90	Rbm47_1R	aaacAGCGTGGCTGTGAGGTCTTCc	
bVR73	Rbm47 gRNAs: Ligation into bVR66 of insert in bVR47			px330A-1x2
bVR74	Esrp1 gRNAs: Ligation into bVR67 of insert in bVR69			px330A-1x2
bVR75	Esrp2 gRNAs: Ligation into bVR68 of insert in bVR70			px330A-1x2
bVR76	Rbm47 + Esrp1 gRNAs: Ligation into bVR72 of inserts in bVR47, bVR52, bVR53			px330A-1x4
bVR77	Esrp1 + Esrp2 gRNAs: Ligation into bVR71 of inserts in bVR52, bVR53, bVR70			px330A-1x4
bVR78	Rbm47 + Esrp1 + Esrp2 gRNAs: Ligation into bVR44 of inserts in bVR47, bVR52, bVR53, bVR58, bVR59			px330A-1x6

MM Table 6: Plasmids generated for CRISPR-Cas9 mediated gene targeting

Gene	Sequence	ID
Actb F	ACGCACGATTTCCCTCTCAGC	29
Actb R	GGCCCAGAGCAAGAGAGGTAT	30
Blbp F	AACCTGGAAGCTGACAGACAGT	199
Blbp R	TCACAGTTGGTTTGGTCACG	200
Dnmt3b F	CGAGAACAAAAGTCGAAGACG	BP1
Dnmt3b R	GGGTTCTTCTTTCCACAGGAC	BP2
Eomes F	TTCACCTTCTCAGAGACACAGTTCAT	7
Eomes R	GAGTTAACCTGTCATTTTCTGAAGCC	8
Esrp1 F	CTGTCTCCTCCCTCCTACACA	89
Esrp1 R	CTATTAGGCGAACCTGGGGG	90
Esrp2 F1	CTACTTGAAGGGTCCAATGGTCA	183
Esrp2 R1	GGCCTGTCTCGTATTTCTGCT	184
Fgf5 F	CCTTGCGACCCAGGAGCTTA	17
Fgf5 R	CCGTCTGTGGTTTCTGTTGAGG	18
Flk1 F	AGAACATTTGTCCGAGTTCACA	BP5
Flk1 R	CGGACTTGACTGCCCACT	BP6
Gata4 F	TCAACCGGCCCCTCATTAAG	13
Gata4 R	CACCCTCGGCATTACGACG	14
Gata6 F	GACTCCTACTTCCTCTTCTTCTAATTCAGA	BP7
Gata6 R	ACCTGAATACTTGAGGTCACTGTTCTC	BP8
Klf2 F	AAGAGCTCGCACCTAAAGGC	149
Klf2 R	CGCATCCTTCCCAGTTGC	150
Klf4 F	ACCTATACCAAGAGTTCTCATC	15
Klf4 R	TCTGGCACTGAAAGGGCCGG	16
Musashi1 F	ACGACCCCTGCAAGATGTTC	203
Musashi1 R	GACACTCTTTCACCTCCCCG	204
Nanog F	CTTACAAGGGTCTGCTACTGAGATGC	3
Nanog R	TGCTTCCTGGCAAGGACCTT	4
Nestin F	GAGGTGACCCTTGGGTTAGA	202
Nestin R	TAGCCCTACCACTTCCTGCT	201
Otx2 F	AATCAACTTGCCAGAATCCAGGG	153
Otx2 R	GCTGTTGGCGGCACTTAGC	154
Pou3f1 F	TTTCTCAAGTGTCCCAAGCC	155
Pou3f1 R	ACCACCTCCTTCTCCAGTTG	156
Pou5f1 F	CGTGGAGACTTTGCAGCCTG	1

Gene	Sequence	ID
Pou5f1 R	GCTTGGCAAACTGTTCTAGCTCCT	2
Qki F	TCAAGCTGAAGAGAGCGGTT	197
Qki R	AGGCAAGGGCTGGTGATTT	198
Rbfox2 F	ATGGAGTGCCACACACTCAG	187
Rbfox2 R	TGTGCTCCACCTTCTGTCTG	188
Rbm47 F	CACATGATCAGCCCCATTGC	193
Rbm47 R	TCTGAACATTTGGTGCCACG	194
Rbmx F	AAGGATGCTGCCAGAGACATGA	43
Rbmx R	AGTCCACGTCTACCACTTTCAA	44
RbmxL1 F	GTTACCTTCGAAAGCCCAGC	47
RbmxL1 R	GTCCACGCCTGCCACTTT	48
RbmxL2 F	CGGCTCTTCACCACCTCAAT	141
RbmxL2 R	CTCGATGATGCGGCCATACT	142
Rex1 F	GGCTGCGAGAAGAGCTTTATTCA	19
Rex1 R	AGCATTTCTTCCCGGCCTTT	20
Snail F	CACACGCTGCCTTGTGTCT	BP3
Snail R	GGTCAGCAAAAGCACGGTT	BP4
Sox17 F	AGCCATTTCCTCCGTGGTGT	27
Sox17 R	AACACTGCTTCTGGCCCTCAG	28
Sox7 F	ATGCTGGGAAAGTCATGGAAG	BP9
Sox7 R	CGTGTTCTGGTCACGAGAGA	BP10
T F	TTGAACTTTCCTCCATGTGCTGA	23
T R	TCCCAAGAGCCTGCCACTTT	24

MM Table 7: RTqPCR primers

Gene	Main event (VASTDB ID)	Sequence	ID
1700029G01 Rik	MmuEX0000517	CCCTACACGACGCTCTTCCGATCTCACCCCATCAACTCTCGGGA	1
1700029G01 Rik	MmuEX0000517	TTCAGACGTGTGCTCTTCCGATCTTGGGGTAATTGGCGGTATGT	2
App	MmuEX0005609	TTCAGACGTGTGCTCTTCCGATCTTCCCCGGGTGTCTCCAGGT	199
App	MmuEX0005609	CCCTACACGACGCTCTTCCGATCTGTCCCATTCTTTTACGGCGG	127
Arhgap17	MmuEX0005804	TTCAGACGTGTGCTCTTCCGATCTGTGGCCTTACATCCTCAGTGC	5
Arhgap17	MmuEX0005804	CCCTACACGACGCTCTTCCGATCTGATGTGGCCAGGCACTTCTTT	6
Arhgap39	MmuEX0005870	CCCTACACGACGCTCTTCCGATCTCTGGCTTTCTTCCCACCAAC	75
Arhgap39	MmuEX0005870	TTCAGACGTGTGCTCTTCCGATCTCCTTCTGCAGTTTGTGGTAAC	76
Baiap2	MmuEX0007650	CCCTACACGACGCTCTTCCGATCTGGTCTGGATGACTACGGGGC	7
Baiap2	MmuEX0007650	TTCAGACGTGTGCTCTTCCGATCTAGAGTGTCGATGCAGAGGAG	8
Dennd1b	MmuEX0014234	CCCTACACGACGCTCTTCCGATCTCTCTTAAATCACTCTATAAC-CACCCA	13
Dennd1b	MmuEX0014234	TTCAGACGTGTGCTCTTCCGATCTACTCTCCGGTATCGTTGGGAG	14
Dennd4c	MmuEX0014291	CCCTACACGACGCTCTTCCGATCTGTGGTATTTTCTTGTGGACGAAA	131
Dennd4c	MmuEX0014291	TTCAGACGTGTGCTCTTCCGATCTCGTAGCCCTGGTCAGACTG	132
Dnmt3b	MmuEX0015337	CCCTACACGACGCTCTTCCGATCTCCCACGGTGGCTTCAAGCC	178
Dnmt3b	MmuEX0015337	TTCAGACGTGTGCTCTTCCGATCTGGGACTCAGAAGCAGCAGAG	179
Ehbp1	MmuEX0016447	CCCTACACGACGCTCTTCCGATCTTCGATGCGGAGTGAAGATGT	79
Ehbp1	MmuEX0016447	TTCAGACGTGTGCTCTTCCGATCTGCCTCAAGTCCACCACCTTA	80
Enox1	MmuEX0016887	CCCTACACGACGCTCTTCCGATCTCTGTGACAGCCCGACTAAAA	15
Enox1	MmuEX0016887	TTCAGACGTGTGCTCTTCCGATCTCTCCACCTCGTTCCTGTAGG	134
Fgf5	MmuEX0019139	CCCTACACGACGCTCTTCCGATCTCCGGATGGCAAAGTCAATGG	183
Fgf5	MmuEX0019139	TTCAGACGTGTGCTCTTCCGATCTAGTTCTGTGGATCGCGGAC	184
Gapdh		CCCTACACGACGCTCTTCCGATCTCCAAGTATGATGACATCAAGAA-GGT	185
Gapdh		TTCAGACGTGTGCTCTTCCGATCTTCCTTGGAGGCCATGTAGGC	186
Hook3	MmuEX0023183	CCCTACACGACGCTCTTCCGATCTGCAAGAATGAGCTGGAGACG	87
Hook3	MmuEX0023183	TTCAGACGTGTGCTCTTCCGATCTTCGAGATCTTCAATAATGGCTCT	135
Idh3g	MmuEX0023592	TTCAGACGTGTGCTCTTCCGATCTGTTCTGGCTGCCCATGTGG	17
Idh3g	MmuEX0023592	CCCTACACGACGCTCTTCCGATCTCCCCTGGGATCATAGTCACTGT	18
Insr	MmuEX0024247	CCCTACACGACGCTCTTCCGATCTGCACAACGTGGTTTTTGTCCC	61
Insr	MmuEX0024247	TTCAGACGTGTGCTCTTCCGATCTCTGTCACATTCCCCACCTCT	62
Klc1	MmuEX0025679	CCCTACACGACGCTCTTCCGATCTAGAGCATGGAGAAGCGGAG	97
Klc1	MmuEX0025679	TTCAGACGTGTGCTCTTCCGATCTCGTTTTCCACACAAAGAGGC	98
Klf4 F		CCCTACACGACGCTCTTCCGATCTACCTATACCAAGAGTTCTCATC	191
Klf4 R		TTCAGACGTGTGCTCTTCCGATCTTCTGGCACTGAAAGGGCCGG	192
Ktn1	MmuEX0025961	CCCTACACGACGCTCTTCCGATCTTCCATGAGAAAGATGGACAGATCA	19
Ktn1	MmuEX0025961	TTCAGACGTGTGCTCTTCCGATCTAAGTTCTGCAGCTCCTGGACT	20
Lpin1	MmuEX0026737	CCCTACACGACGCTCTTCCGATCTCCTCGATTTCAACGTACCCCC	21
Lpin1	MmuEX0026737	TTCAGACGTGTGCTCTTCCGATCTCAGCCAGAGCATTTCCAGGTT	22
Lrch2	MmuEX0026786	TTCAGACGTGTGCTCTTCCGATCTTGACAAAGAACAGTTACTTCCAGA	23
Lrch2	MmuEX0026786	CCCTACACGACGCTCTTCCGATCTTGCCATGCAAGAGGTGATAATGG	24
Mapt	MmuEX0027903	CCCTACACGACGCTCTTCCGATCTACTGAGAACCTGAAGCACCA	49
Mapt	MmuEX0027903	TTCAGACGTGTGCTCTTCCGATCTGTCACTTTGCTCAGGTCCACC	50
Mtmr2		CCCTACACGACGCTCTTCCGATCTCTCCAGCTGCGAGAGTCTC	141
Mtmr2		TTCAGACGTGTGCTCTTCCGATCTTGCACCGAGTTCTCTGAATG	82

Gene	Main event (VASTDB ID)	Sequence	ID
Myo1b	MmuEX0030478	CCCTACACGACGCTCTTCCGATCTGCTCGAAAAATCCTGCGGGAG	193
Myo1b	MmuEX0030478	TTCAGACGTGTGCTCTTCCGATCTTCATAGATTTTCTTTCCAGCATTTGC	194
Nbeal1	MmuEX0031005	CCCTACACGACGCTCTTCCGATCTCCCGTTCTTCAGGCCATTTT	27
Nbeal1	MmuEX0031005	TTCAGACGTGTGCTCTTCCGATCTACAGCCAAGCAATCCAACTG	28
Nf1	MmuEX0031550	CCCTACACGACGCTCTTCCGATCTATCAGCTCTTCCTCAGAATTCCC	176
Nf1	MmuEX0031550	TTCAGACGTGTGCTCTTCCGATCTTCTGAGAAACATGGCACTTCCTAC	177
Nf2	MmuEX0031553	CCCTACACGACGCTCTTCCGATCTGGACGTCCTACACAGCGA	142
Nf2	MmuEX0031553	TTCAGACGTGTGCTCTTCCGATCTGGGTCACCTGCTAGAGTTCT	144
Nktr	MmuEX0031809	CCCTACACGACGCTCTTCCGATCTGGTGGAGAATCAATTTATGGTGG	95
Nktr	MmuEX0031809	TTCAGACGTGTGCTCTTCCGATCTCCCTCGATTTGCCATTGACA	96
Npr1	MmuINT0110282	TTCAGACGTGTGCTCTTCCGATCTTTGCCCTAAGAAGTGGGGTC	145
Npr1	MmuINT0110282	CCCTACACGACGCTCTTCCGATCTCAGATCCATGGAAGGTCGGA	146
Nrf1	MmuEX0032327	CCCTACACGACGCTCTTCCGATCTACTCGGACATCCTCAACTCCA	67
Nrf1	MmuEX0032327	TTCAGACGTGTGCTCTTCCGATCTTTCCGTTTCTTCCCTGTTGCC	68
Numb	MmuEX0032648	CCCTACACGACGCTCTTCCGATCTCGTAGAAGTTGATGAGTCAAGAGG	174
Numb	MmuEX0032648	TTCAGACGTGTGCTCTTCCGATCTCGTCCACAACTCTGAGCCCA	175
Picalm	MmuEX0035017	TTCAGACGTGTGCTCTTCCGATCTTCTGTCCACGCCATGTCG	150
Picalm	MmuEX0035017	CCCTACACGACGCTCTTCCGATCTAAGGCCAGCTGAAGAATGTG	151
Ppm1b	MmuEX0036621	CCCTACACGACGCTCTTCCGATCTGCGCCATGTTATTGAAGCTGT	31
Ppm1b	MmuEX0036621	TTCAGACGTGTGCTCTTCCGATCTCTTCCGGGCTCTTGGTAGGTC	32
Rab24	MmuINT0130443	CCCTACACGACGCTCTTCCGATCTGCAAGGAATACGTGGGCAAG	113
Rab24	MmuINT0130443	TTCAGACGTGTGCTCTTCCGATCTCATCACCTTGGCCACGAAAG	114
Rnf138	MmuEX0040027	CCCTACACGACGCTCTTCCGATCTTGTCAAGAGTCAAATTTCACCAG	155
Rnf138	MmuEX0040027	TTCAGACGTGTGCTCTTCCGATCTGACTCTTCCACAGCAGTTTGA	156
Sorbs1	MmuEX0044358	TTCAGACGTGTGCTCTTCCGATCTGAGGCCATTGGTGAAAACCC	157
Sorbs1	MmuEX0044358	CCCTACACGACGCTCTTCCGATCTAATCCAGTGAGGGCTGACTCC	48
Stk11ip	MmuINT0154024	CCCTACACGACGCTCTTCCGATCTTTCCTGCCCTAAAAGAGCCC	111
Stk11ip	MmuINT0154024	TTCAGACGTGTGCTCTTCCGATCTTCCACACTACTGAATCCCGC	112
Supt5h	MmuEX0045761	CCCTACACGACGCTCTTCCGATCTCCAGTGGGAAGATGGAGCTGA	33
Supt5h	MmuEX0045761	TTCAGACGTGTGCTCTTCCGATCTTTCTGCAGACGACGAGCC	34
Tcf12	MmuEX0046651	TTCAGACGTGTGCTCTTCCGATCTGAGGGCAAGCTCCTTCATCTC	35
Tcf12	MmuEX0046651	CCCTACACGACGCTCTTCCGATCTCAACTGCATGGTTTCGTAGC	158
Thrap3	MmuEX0047240	CCCTACACGACGCTCTTCCGATCTCAGAAAGAGGAGAGTGCTGC	159
Thrap3	MmuEX0047240	TTCAGACGTGTGCTCTTCCGATCTACGGAACTCCTGCTCCTTTT	160
Tnrc6b	MmuEX0048388	TTCAGACGTGTGCTCTTCCGATCTAAACCTGCCTACCCCAATGA	161
Tnrc6b	MmuEX0048388	CCCTACACGACGCTCTTCCGATCTCTCCGTCATCCTCCTCTTCC	162
Uspl1	MmuEX0051128	CCCTACACGACGCTCTTCCGATCTATGCGAAGAAGGAGCGTACT	91
Uspl1	MmuEX0051128	TTCAGACGTGTGCTCTTCCGATCTCAATGCCCGTCTGTTGTGAC	92
Yap1	MmuEX0052491	TTCAGACGTGTGCTCTTCCGATCTGCAGCAGTTACAGATGGAGA	197
Yap1	MmuEX0052491	CCCTACACGACGCTCTTCCGATCTATCCCAGGAGAAGACACTGC	198
Zc3h18		TTCAGACGTGTGCTCTTCCGATCTCCCACCTGCCGATTCTTCAT	83
Zc3h18		CCCTACACGACGCTCTTCCGATCTCCGTTGGGTGGAAAAGGTTC	84
Zfp934		CCCTACACGACGCTCTTCCGATCTTGTCATGCTGGAGACCTACA	170
Zfp934		TTCAGACGTGTGCTCTTCCGATCTCAAAGGCTTTATCACGTTGAGT	171

MM Table 8: SPARseq primers

Bibliography

1. Sheng G. Epiblast morphogenesis before gastrulation. *Dev Biol*. 2015;401(1):17-24. doi:10.1016/j.ydbio.2014.10.003
2. Ferner K, Mess A. Evolution and development of fetal membranes and placentation in amniote vertebrates. *Respir Physiol Neurobiol*. 2011;178(1):39-50. doi:10.1016/j.resp.2011.03.029
3. Posfai E, Tam OH, Rossant J. *Mechanisms of Pluripotency In Vivo and In Vitro*. Vol 107. 1st ed. Elsevier Inc.; 2014. doi:10.1016/B978-0-12-416022-4.00001-9
4. Johnston RJ, Desplan C. Stochastic Mechanisms of Cell Fate Specification that Yield Random or Robust Outcomes. *Annu Rev Cell Dev Biol*. 2010;26(1):689-719. doi:10.1146/annurev-cellbio-100109-104113
5. Magnúsdóttir E, Gillich A, Grabole N, Surani MA. Combinatorial control of cell fate and reprogramming in the mammalian germline. *Curr Opin Genet Dev*. 2012;22(5):466-474. doi:10.1016/j.gde.2012.06.002
6. Zernicka-Goetz M, Morris SA, Bruce AW. Making a firm decision: Multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet*. 2009;10(7):467-477. doi:10.1038/nrg2564
7. Condic ML. Totipotency: What it is and what it is not. *Stem Cells Dev*. 2014;23(8):796-812. doi:10.1089/scd.2013.0364
8. Aiken CEM, Swoboda PPL, Skepper JN, Johnson MH. The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. *Reproduction*. 2004;128(5):527-535. doi:10.1530/rep.1.00281
9. Seidel F. Die Entwicklungspotenzen einer isolierten Blastomere des Zweizellenstadiums im Saugetierei. *Naturwissenschaften*. 1952;39(15):355-356. doi:10.1007/BF00589948
10. Nicholas JS, Hall B V. Experiments on developing rats. II. The development of isolated blastomeres and fused eggs. *J Exp Zool*. 1942;90(3):441-459. doi:10.1002/jez.1400900307
11. Tarkowski AK. Experiments on the Development of Isolated Blastomeres of Mouse Eggs. *Nature*. 1959;184(4695):1286-1287. doi:10.1038/1841286a0
12. Antczak M. Oocyte influences on early development: the regulatory proteins leptin and STAT3 are polarized in mouse and human oocytes and differentially distributed within the cells of the preimplantation stage embryo. *Mol Hum Reprod*. 1997;3(12):1067-1086. doi:10.1093/molehr/3.12.1067
13. Piotrowska-Nitsche K, Perea-Gomez A, Haraguchi S, Zernicka-Goetz M. Four-cell stage mouse blastomeres have different developmental properties. *Development*. 2005;132(3):479-490. doi:10.1242/dev.01602
14. Torres-Padilla ME, Parfitt DE, Kouzarides T, Zernicka-Goetz M. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature*. 2007;445(7124):214-218. doi:10.1038/nature05458
15. Goolam M, Scialdone A, Graham SJL, et al. Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell*. 2016;165(1):61-74. doi:10.1016/j.cell.2016.01.047
16. Jedrusik A, Parfitt DE, Guo G, et al. Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes Dev*. 2008;22(19):2692-2706. doi:10.1101/gad.486108
17. Plachta N, Bollenbach T, Pease S, Fraser SE, Pantazis P. Oct4 kinetics predict cell

lineage patterning in the early mammalian embryo. *Nat Cell Biol*. 2011;13(2):117-123. doi:10.1038/ncb2154
18. Hartshorn C, Eckert JJ, Hartung O, Wangh LJ. Single-cell duplex RT-LATE-PCR reveals Oct4 and Xist RNA gradients in 8-cell embryos. *BMC Biotechnol*. 2007;7:1-14. doi:10.1186/1472-6750-7-87
19. Rossant J, Vijh KM. Ability of outside cells from preimplantation mouse embryos to form inner cell mass derivatives. *Dev Biol*. 1980;76(2):475-482. doi:10.1016/0012-1606(80)90395-4
20. Lawson KA, Dunn NR, Roelen BAJ, et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev*. 1999;13(4):424-436. doi:10.1101/gad.13.4.424
21. Ducibella T, Ukena T, Karnovsky M, Anderson E. Changes in cell surface and cortical cytoplasmic organization during early embryogenesis in the preimplantation mouse embryo. *J Cell Biol*. 1977;74(1):153-167. doi:10.1083/jcb.74.1.153
22. Handyside AH. Distribution of antibody and lectin-binding sites on dissociated blastomeres from mouse morulae: Evidence for polarization at compaction. *J Embryol Exp Morphol*. 1980;VOL.60:99-116.
23. Maro B, Johnson MH, Pickering SJ. Changes in the distribution of membranous organelles during mouse early development. *J Embryol Exp Morphol*. 1985;VOL. 90:287-309.
24. Pratt HPM, Ziomek CA, Reeve WJD, Johnson MH. Compaction of the mouse embryo: An analysis of its components. *J Embryol Exp Morphol*. 1982;Vol. 70:113-132.
25. Stephenson RO, Rossant J, Tam PPL. Intercellular interactions, position, and polarity in establishing blastocyst cell lineages and embryonic axes. *Cold Spring Harb Perspect Biol*. 2012;4(11):1-16. doi:10.1101/cshperspect.a008235
26. Johnson MH, Ziomek CA. Cell interactions influence the fate of mouse blastomeres undergoing the transition from the 16- to the 32-cell stage. *Dev Biol*. 1983;95(1):211-218. doi:10.1016/0012-1606(83)90019-2
27. Nishioka N, Inoue K ichi, Adachi K, et al. The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Dev Cell*. 2009;16(3):398-410. doi:10.1016/j.devcel.2009.02.003
28. Hirate Y, Hirahara S, Inoue KI, et al. Polarity-dependent distribution of angiomotin localizes hippo signaling in preimplantation embryos. *Curr Biol*. 2013;23(13):1181-1194. doi:10.1016/j.cub.2013.05.014
29. Ralston A, Rossant J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev Biol*. 2008;313(2):614-629. doi:10.1016/j.ydbio.2007.10.054
30. Guo G, Huss M, Tong GQ, et al. Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. *Dev Cell*. 2010;18(4):675-685. doi:10.1016/j.devcel.2010.02.012
31. Wu G, Gentile L, Fuchikami T, et al. Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. *Development*. 2010;137(24):4159-4169. doi:10.1242/dev.056630
32. Strumpf D, Mao CA, Yamanaka Y, et al. Cdx2 is required for correct cell fate

specification and differentiation of trophectoderm in the mouse blastocyst. *Development*. 2005;132(9):2093-2102. doi:10.1242/dev.01801

33. Nichols J, Zevnik B, Anastassiadis K, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*. 1998;95(3):379-391. doi:10.1016/S0092-8674(00)81769-9
34. Rossant J, Lis WT. Potential of isolated mouse inner cell masses to form trophectoderm derivatives in vivo. *Dev Biol*. 1979;70(1):255-261. doi:10.1016/0012-1606(79)90022-8
35. Grabarek JB, Zyzyńska K, Saiz N, et al. Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. *Development*. 2012;139(1):129-139. doi:10.1242/dev.067702
36. Ohnishi Y, Huber W, Tsumura A, et al. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat Cell Biol*. 2014;16(1):27-37. doi:10.1038/ncb2881
37. Hermitte S, Chazaud C. Primitive endoderm differentiation: from specification to epithelium formation. *Philos Trans R Soc B Biol Sci*. 2014;369(1657):20130537. doi:10.1098/rstb.2013.0537
38. Boroviak T, Nichols J. The birth of embryonic pluripotency. *Philos Trans R Soc B Biol Sci*. 2014;369(1657):20130541-20130541. doi:10.1098/rstb.2013.0541
39. Bassalert C, Valverde-Estrella L, Chazaud C. Primitive Endoderm Differentiation: From Specification to Epithelialization. *Curr Top Dev Biol*. 2018;128:81-104. doi:10.1016/bs.ctdb.2017.12.001
40. Ryan AQ, Chan CJ, Graner F, Hiiragi T. Lumen Expansion Facilitates Epiblast-Primitive Endoderm Fate Specification during Mouse Blastocyst Formation. *Dev Cell*. 2019;51(6):684-697.e4. doi:10.1016/j.devcel.2019.10.011
41. Bessonnard S, Mot L De, Gonze D, et al. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Dev*. 2014;141(19):3637-3648. doi:10.1242/dev.109678
42. Schro ter C, Rue P, Mackenzie JP, Martinez Arias A. FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. *Development*. 2015;142(24):4205-4216. doi:10.1242/dev.127530
43. Kang M, Garg V, Hadjantonakis AK. Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2. *Dev Cell*. 2017;41(5):496-510.e5. doi:10.1016/j.devcel.2017.05.003
44. Molotkov A, Mazot P, Brewer JR, Cinalli RM, Soriano P. Distinct Requirements for FGFR1 and FGFR2 in Primitive Endoderm Development and Exit from Pluripotency. *Dev Cell*. 2017;41(5):511-526.e4. doi:10.1016/j.devcel.2017.05.004
45. Chazaud C, Yamanaka Y, Pawson T, Rossant J. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Dev Cell*. 2006;10(5):615-624. doi:10.1016/j.devcel.2006.02.020
46. Nichols J, Silva J, Roode M, Smith A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development*. 2009;136(19):3215-3222. doi:10.1242/dev.038893
47. Krawchuk D, Honma-Yamanaka N, Anani S, Yamanaka Y. FGF4 is a limiting factor controlling the proportions of primitive endoderm and epiblast in the ICM of the mouse blastocyst. *Dev Biol*. 2013;384(1):65-71. doi:10.1016/j.ydbio.2013.09.023

48. Yamanaka Y, Lanner F, Rossant J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development*. 2010;137(5):715-724. doi:10.1242/dev.043471
49. Guo F, Li L, Li J, et al. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res*. 2017;27(8):967-988. doi:10.1038/cr.2017.82
50. Plusa B, Piliszek A, Frankenberg S, Artus J, Hadjantonakis AK. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development*. 2008;135(18):3081-3091. doi:10.1242/dev.021519
51. Gerbe F, Cox B, Rossant J, Chazaud C. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Dev Biol*. 2008;313(2):594-602. doi:10.1016/j.ydbio.2007.10.048
52. Frankenberg S, Gerbe F, Bessonnard S, et al. Primitive Endoderm Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling. *Dev Cell*. 2011;21(6):1005-1013. doi:10.1016/j.devcel.2011.10.019
53. Schrode N, Saiz N, Di Talia S, Hadjantonakis AK. GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Dev Cell*. 2014;29(4):454-467. doi:10.1016/j.devcel.2014.04.011
54. Artus J, Piliszek A, Hadjantonakis AK. The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. *Dev Biol*. 2011;350(2):393-404. doi:10.1016/j.ydbio.2010.12.007
55. Bessonnard S, Coqueran S, Vandormael-Pournin S, Dufour A, Artus J, Cohen-Tannoudji M. ICM conversion to epiblast by FGF/ERK inhibition is limited in time and requires transcription and protein degradation. *Sci Rep*. 2017;7(1):1-12. doi:10.1038/s41598-017-12120-0
56. Piras V, Tomita M, Selvarajoo K. Transcriptome-wide variability in single embryonic development cells. *Sci Rep*. 2014;4(Figure 1):1-9. doi:10.1038/srep07137
57. Biase FH, Cao X, Zhong S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome Res*. 2014;24(11):1787-1796. doi:10.1101/gr.177725.114
58. Shi J, Chen Q, Li X, et al. Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell rna-seq. *Dev*. 2015;142(20):3468-3477. doi:10.1242/dev.123950
59. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010;467(7312):167-173. doi:10.1038/nature09326
60. White MD, Zenker J, Bissiere S, Plachta N. Instructions for Assembling the Early Mammalian Embryo. *Dev Cell*. 2018;45(6):667-679. doi:10.1016/j.devcel.2018.05.013
61. Balázsi G, Van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: From microbes to mammals. *Cell*. 2011;144(6):910-925. doi:10.1016/j.cell.2011.01.030
62. Dietrich JE, Hiiragi T. Stochastic patterning in the mouse pre-implantation embryo. *Development*. 2007;134(23):4219-4231. doi:10.1242/dev.003798
63. Hansen CH, Van Oudenaarden A. Allele-specific detection of single mRNA molecules in situ. *Nat Methods*. 2013;10(9):869-871. doi:10.1038/nmeth.2601
64. Abranches E, Guedes AMV, Moravec M, et al. Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Dev*. 2014;141(14):2770-2779.

doi:10.1242/dev.108910
65. White MD, Angiolini JF, Alvarez YD, et al. Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo. *Cell*. 2016;165(1):75-87. doi:10.1016/j.cell.2016.02.032
66. Symmons O, Raj A. What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Mol Cell*. 2016;62(5):788-802. doi:10.1016/j.molcel.2016.05.023
67. Morris SA, Teo RTY, Li H, Robson P, Glover DM, Zernicka-Goetz M. Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc Natl Acad Sci U S A*. 2010;107(14):6364-6369. doi:10.1073/pnas.0915063107
68. Krupa M, Mazur E, Szczepańska K, Filimonow K, Maleszewski M, Suwińska A. Allocation of inner cells to epiblast vs primitive endoderm in the mouse embryo is biased but not determined by the round of asymmetric divisions (8→16- and 16→32-cells). *Dev Biol*. 2014;385(1):136-148. doi:10.1016/j.ydbio.2013.09.008
69. Yamanaka Y. Response: Cell fate in the early mouse embryo - Sorting out the influence of developmental history on lineage choice. *Reprod Biomed Online*. 2011;22(6):525-527. doi:10.1016/j.rbmo.2011.03.011
70. Morris SA. Cell fate in the early mouse embryo: Sorting out the influence of developmental history on lineage choice. *Reprod Biomed Online*. 2011;22(6):521-524. doi:10.1016/j.rbmo.2011.02.009
71. Fleming TP. A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. *Dev Biol*. 1987;119(2):520-531. doi:10.1016/0012-1606(87)90055-8
72. Saiz N, Plusa B. Early cell fate decisions in the mouse embryo. *Reproduction*. 2013;145(3). doi:10.1530/REP-12-0381
73. Ralston A, Cox BJ, Nishioka N, et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development*. 2010;137(3):395-403. doi:10.1242/dev.038828
74. Cho LTY, Wamaitha SE, Tsai IJ, et al. Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Dev*. 2012;139(16):2866-2877. doi:10.1242/dev.078519
75. Schrode N, Xenopoulos P, Piliszek A, Frankenberg S, Plusa B, Hadjantonakis AK. Anatomy of a blastocyst: Cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. *Genesis*. 2013;51(4):219-233. doi:10.1002/dvg.22368
76. Nichols J, Smith A. Naive and Primed Pluripotent States. *Cell Stem Cell*. 2009;4(6):487-492. doi:10.1016/j.stem.2009.05.015
77. Selwood L, Johnson MH. Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: Evolution and origins. *BioEssays*. 2006;28(2):128-145. doi:10.1002/bies.20360
78. Morgani S, Nichols J, Hadjantonakis AK. The many faces of Pluripotency: In vitro adaptations of a continuum of in vivo states. *BMC Dev Biol*. 2017;17(1):10-12. doi:10.1186/s12861-017-0150-4
79. Brook FA, Gardner RL. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A*. 1997;94(11):5709-5712. doi:10.1073/pnas.94.11.5709

80. Kaufman MH, Evans MJ. Establishment in culture of pluripotential cells from mouse embryos. *Nature*. 1981;292(July):154-156.
81. Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A*. 1981;78(12 II):7634-7638. doi:10.1073/pnas.78.12.7634
82. Brons IGM, Smithers LE, Trotter MWB, et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. 2007;448(7150):191-195. doi:10.1038/nature05950
83. Tesar PJ, Chenoweth JG, Brook FA, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 2007;448(7150):196-199. doi:10.1038/nature05972
84. Tanaka S, Kunath T, Hadjantonakis A-K, Nagy A, Rossant J. Promotion of Trophoblast Stem Cell Proliferation by FGF4. *Science (80-)*. 1998;282(5396):2072-2075. doi:10.1126/science.282.5396.2072
85. Kunath T, Arnaud D, Uy GD, et al. Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development*. 2005;132(7):1649-1661. doi:10.1242/dev.01715
86. Carson DD, Bagchi I, Dey SK, et al. Embryo implantation. *Dev Biol*. 2000;223(2):217-237. doi:10.1006/dbio.2000.9767
87. Wang H, Dey SK. Roadmap to embryo implantation: Clues from mouse models. *Nat Rev Genet*. 2006;7(3):185-199. doi:10.1038/nrg1808
88. Ramathal CY, Bagchi IC, Taylor RN, Bagchi MK. Endometrial decidualization: Of mice and men. *Semin Reprod Med*. 2010;28(1):17-26. doi:10.1055/s-0029-1242989
89. Latos PA, Hemberger M. From the stem of the placental tree: Trophoblast stem cells and their progeny. *Dev*. 2016;143(20):3650-3660. doi:10.1242/dev.133462
90. Gardner RL. Origin and differentiation of extraembryonic tissues in the mouse. *Int Rev Exp Pathol*. 1983.
91. Bedzhov I, Zernicka-Goetz M. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell*. 2014;156(5):1032-1044. doi:10.1016/j.cell.2014.01.023
92. Shahbazi MN, Scialdone A, Skorupska N, et al. Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature*. 2017;552(7684):239-243. doi:10.1038/nature24675
93. Copp AJ. The mechanism of mouse egg-cyclinder morphogenesis in vitro. *J Embryol Exp Morphol*. 1981.
94. Rossant J. Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. *Semin Cell Dev Biol*. 2004. doi:10.1016/j.semcdb.2004.04.003
95. Stower MJ, Srinivas S. *The Head's Tale: Anterior-Posterior Axis Formation in the Mouse Embryo*. Vol 128. 1st ed. Elsevier Inc.; 2018. doi:10.1016/bs.ctdb.2017.11.003
96. Gardner RL, Rossant J. Investigation of the fate of 4.5 day post-coitum mouse inner cell mass cells by blastocyst injection. *J Embryol Exp Morphol*. 1979.
97. Viotti M, Foley AC, Hadjantonakis A-K. Gutsy moves in mice: cellular and molecular dynamics of endoderm morphogenesis. *Philos Trans R Soc B Biol Sci*. 2014;369(1657):20130547-20130547. doi:10.1098/rstb.2013.0547
98. Tam P. Endoderm Formation: Not So Black and White Anymore. *Dev Cell*. 2011.

doi:10.1016/j.devcel.2011.10.003
99. Kwon GS, Viotti M, Hadjantonakis AK. The Endoderm of the Mouse Embryo Arises by Dynamic Widespread Intercalation of Embryonic and Extraembryonic Lineages. *Dev Cell*. 2008. doi:10.1016/j.devcel.2008.07.017
100. Beck S, Le Good JA, Guzman M, et al. Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nat Cell Biol*. 2002. doi:10.1038/ncb890
101. Kimura C, Yoshinaga K, Tian E, Suzuki M, Aizawa S, Matsuo I. Visceral endoderm mediates forebrain development by suppressing posteriorizing signals. *Dev Biol*. 2000. doi:10.1006/dbio.2000.9835
102. Rivera-Pérez JA, Mager J, Magnuson T. Dynamic morphogenetic events characterize the mouse visceral endoderm. *Dev Biol*. 2003. doi:10.1016/S0012-1606(03)00302-6
103. Srinivas S, Rodriguez T, Clements M, Smith JC, Beddington RSP. Active cell migration drives the unilateral movements of the anterior visceral endoderm. *Development*. 2004. doi:10.1242/dev.01005
104. Thomas P, Beddington R. Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr Biol*. 1996. doi:10.1016/S0960-9822(96)00753-1
105. Yamamoto M, Saijoh Y, Perea-Gomez A, et al. Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature*. 2004. doi:10.1038/nature02418
106. Beddington RSP, Robertson EJ. Axis development and early asymmetry in mammals. *Cell*. 1999. doi:10.1016/S0092-8674(00)80560-7
107. Tam PPL, Loebel DAF. Gene function in mouse embryogenesis: Get set for gastrulation. *Nat Rev Genet*. 2007;8(5):368-381. doi:10.1038/nrg2084
108. Ohinata Y, Payer B, O'Carroll D, et al. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature*. 2005. doi:10.1038/nature03813
109. Leitch HG, Tang WWC, Surani MA. Primordial Germ-Cell Development and Epigenetic Reprogramming in Mammals. In: *Current Topics in Developmental Biology*. ; 2013. doi:10.1016/B978-0-12-416027-9.00005-X
110. Peng G, Suo S, Cui G, et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature*. 2019;572(7770):528-532. doi:10.1038/s41586-019-1469-8
111. Mohammed H, Hernando-Herraez I, Savino A, et al. Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Rep*. 2017;20(5):1215-1228. doi:10.1016/j.celrep.2017.07.009
112. Chew J-L, Loh Y-H, Zhang W, et al. Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Mol Cell Biol*. 2005. doi:10.1128/mcb.25.14.6031-6046.2005
113. Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem*. 2005;280(7):5307-5317. doi:10.1074/jbc.M410015200
114. Zhang J, Tam WL, Tong GQ, et al. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol*. 2006. doi:10.1038/ncb1481
115. Wu Q, Chen X, Zhang J, et al. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem*. 2006. doi:10.1074/jbc.C600122200

116. Sun LT, Yamaguchi S, Hirano K, Ichisaka T, Kuroda T, Tada T. Nanog co-regulated by Nodal/Smad2 and Oct4 is required for pluripotency in developing mouse epiblast. *Dev Biol*. 2014. doi:10.1016/j.ydbio.2014.06.002
117. Di-Gregorio A, Sancho M, Stuckey DW, et al. BMP signalling inhibits premature neural differentiation in the mouse embryo. *Development*. 2007. doi:10.1242/dev.005967
118. Smith A. Formative pluripotency: the executive phase in a developmental continuum. *Development*. 2017;144(3):365-373. doi:10.1242/dev.142679
119. Cheng S, Pei Y, He L, et al. Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development. *Cell Rep*. 2019;26(10):2593-2607.e3. doi:10.1016/j.celrep.2019.02.031
120. Kojima Y, Kaufman-Francis K, Studdert JB, et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell*. 2014;14(1):107-120. doi:10.1016/j.stem.2013.09.014
121. Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*. 2011;146(4):519-532. doi:10.1016/j.cell.2011.06.052
122. Stevens LC, Little CC. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc Natl Acad Sci*. 1954. doi:10.1073/pnas.40.11.1080
123. Martin GR. Teratocarcinomas and mammalian embryogenesis. *Science (80-)*. 1980. doi:10.1126/science.6250214
124. Bradley A, Evans M, Kaufman MH, Robertson E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*. 1984. doi:10.1038/309255a0
125. Capecchi MR. Gene targeting in mice: Functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet*. 2005. doi:10.1038/nrg1619
126. Gardner RL, Lyon MF, Evans EP, Burtenshaw MD. Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. *J Embryol Exp Morphol*. 1985.
127. Yang Y, Liu B, Xu J, et al. Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell*. 2017;169(2):243-257.e25. doi:10.1016/j.cell.2017.02.005
128. Yang J, Ryan DJ, Wang W, et al. Establishment of mouse expanded potential stem cells. *Nature*. 2017;550(7676). doi:10.1038/nature24052
129. Matsui Y, Zsebo K, Hogan BLM. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell*. 1992;70(5):841-847. doi:10.1016/0092-8674(92)90317-6
130. Kanatsu-Shinohara M, Inoue K, Lee J, et al. Generation of pluripotent stem cells from neonatal mouse testis. *Cell*. 2004. doi:10.1016/j.cell.2004.11.011
131. Ko K, Tapia N, Wu G, et al. Induction of Pluripotency in Adult Unipotent Germline Stem Cells. *Cell Stem Cell*. 2009. doi:10.1016/j.stem.2009.05.025
132. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. 2006;126(4):663-676. doi:10.1016/j.cell.2006.07.024
133. Wakayama T, Tabar V, Rodriguez I, Perry ACF, Studer L, Mombaerts P. Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer.

Science (80-). 2001;292(5517):740-743. doi:10.1126/science.1059399
134. Najm FJ, Chenoweth JG, Anderson PD, et al. Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell*. 2011. doi:10.1016/j.stem.2011.01.016
135. Yilmaz A, Benvenisty N. Defining Human Pluripotency. *Cell Stem Cell*. 2019;25(1):9-22. doi:10.1016/j.stem.2019.06.010
136. Thomson JA, Kalishman J, Golos TG, et al. Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A*. 1995. doi:10.1073/pnas.92.17.7844
137. Thomson JA. Embryonic stem cell lines derived from human blastocysts. *Science (80-)*. 1998. doi:10.1126/science.282.5391.1145
138. Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science (80-)*. 2007. doi:10.1126/science.1151526
139. Liu H, Zhu F, Yong J, et al. Generation of Induced Pluripotent Stem Cells from Adult Rhesus Monkey Fibroblasts. *Cell Stem Cell*. 2008. doi:10.1016/j.stem.2008.10.014
140. Lovell-Badge R. Many ways to pluripotency. *Nat Biotechnol*. 2007;25(10):1114-1116. doi:10.1038/nbt1007-1114
141. Hanna J, Markoulaki S, Mitalipova M, et al. Metastable Pluripotent States in NOD-Mouse-Derived ESCs. *Cell Stem Cell*. 2009;4(6):513-524. doi:10.1016/j.stem.2009.04.015
142. Hanna J, Cheng AW, Saha K, et al. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A*. 2010. doi:10.1073/pnas.1004584107
143. Gafni O, Weinberger L, Mansour AA, et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature*. 2013. doi:10.1038/nature12745
144. Chan YS, Göke J, Ng JH, et al. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell*. 2013. doi:10.1016/j.stem.2013.11.015
145. Theunissen TW, Powell BE, Wang H, et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell*. 2014. doi:10.1016/j.stem.2014.07.002
146. Takashima Y, Guo G, Loos R, et al. Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell*. 2014;158(6):1254-1269. doi:10.1016/j.cell.2014.08.029
147. Ware CB, Nelson AM, Mecham B, et al. Derivation of naïve human embryonic stem cells. *Proc Natl Acad Sci U S A*. 2014. doi:10.1073/pnas.1319738111
148. Theunissen TW, Friedli M, He Y, et al. Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell*. 2016;19(4):502-515. doi:10.1016/j.stem.2016.06.011
149. Liu X, Nefzger CM, Rossello FJ, et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat Methods*. 2017. doi:10.1038/nmeth.4436
150. Guo G, Von Meyenn F, Santos F, et al. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports*. 2016;6(4):437-446. doi:10.1016/j.stemcr.2016.02.005
151. Qin H, Hejna M, Liu Y, et al. YAP Induces Human Naive Pluripotency. *Cell Rep*. 2016;14(10):2301-2312. doi:10.1016/j.celrep.2016.02.036

152. Guo G, von Meyenn F, Rostovskaya M, et al. Epigenetic resetting of human pluripotency. *Development*. 2017;144(15):2748-2763. doi:10.1242/dev.146811
153. Chen RJ, Zhang G, Garfield SH, et al. Variations in Glycogen Synthesis in Human Pluripotent Stem Cells with Altered Pluripotent States. *PLoS One*. 2015;10(11):e0142554. doi:10.1371/journal.pone.0142554
154. Duggal G, Warrier S, Ghimire S, et al. Alternative Routes to Induce Naïve Pluripotency in Human Embryonic Stem Cells. *Stem Cells*. 2015. doi:10.1002/stem.2071
155. Boroviak T, Loos R, Bertone P, Smith A, Nichols J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol*. 2014;16(6):513-525. doi:10.1038/ncb2965
156. Tesar PJ. Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proc Natl Acad Sci U S A*. 2005. doi:10.1073/pnas.0503231102
157. Delhaise F, Bralion V, Schuurbiers N, Dessy F. Establishment of an embryonic stem cell line from 8-cell stage mouse embryos. *Eur J Morphol*. 1996. doi:10.1076/ejom.34.4.237.13046
158. Bao S, Tang F, Li X, et al. Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature*. 2009;461(7268):1292-1295. doi:10.1038/nature08534
159. Kim H, Wu J, Ye S, et al. Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat Commun*. 2013;4. doi:10.1038/ncomms3403
160. Greber B, Wu G, Bernemann C, et al. Conserved and Divergent Roles of FGF Signaling in Mouse Epiblast Stem Cells and Human Embryonic Stem Cells. *Cell Stem Cell*. 2010;6(3):215-226. doi:10.1016/j.stem.2010.01.003
161. Bernemann C, Greber B, Ko K, et al. Distinct Developmental Ground States of Epiblast Stem Cell Lines Determine Different Pluripotency Features. *Stem Cells*. 2011;29(10):1496-1503. doi:10.1002/stem.709
162. Zhou H, Li W, Zhu S, et al. Conversion of mouse epiblast stem cells to an earlier pluripotency state by small molecules. *J Biol Chem*. 2010;285(39):29676-29680. doi:10.1074/jbc.C110.150599
163. Guo G, Yang J, Nichols J, et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development*. 2009;136(7):1063-1069. doi:10.1242/dev.030957
164. Okashita N, Suwa Y, Nishimura O, et al. PRDM14 Drives OCT3/4 Recruitment via Active Demethylation in the Transition from Primed to Naive Pluripotency. *Stem Cell Reports*. 2016;7(6):1072-1086. doi:10.1016/j.stemcr.2016.10.007
165. Silva J, Nichols J, Theunissen TW, et al. Nanog Is the Gateway to the Pluripotent Ground State. *Cell*. 2009. doi:10.1016/j.cell.2009.07.039
166. Yang J, Van Oosten AL, Theunissen TW, Guo G, Silva JCR, Smith A. Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell*. 2010. doi:10.1016/j.stem.2010.06.022
167. Guo G, Smith A. A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development*. 2010. doi:10.1242/dev.052753

168. Williams CAC, Fernandez-Alonso R, Wang J, Toth R, Gray NS, Findlay GM. Erk5 Is a Key Regulator of Naive-Primed Transition and Embryonic Stem Cell Identity. *Cell Rep*. 2016;16(7):1820-1828. doi:10.1016/j.celrep.2016.07.033
169. Stuart HT, Stirparo GG, Lohoff T, et al. Distinct Molecular Trajectories Converge to Induce Naive Pluripotency. *Cell Stem Cell*. 2019;25(3):388-406.e8. doi:10.1016/j.stem.2019.07.009
170. Yu S, Zhou C, Cao S, et al. BMP4 resets mouse epiblast stem cells to naive pluripotency through ZBTB7A/B-mediated chromatin remodelling. *Nat Cell Biol*. May 2020. doi:10.1038/s41556-020-0516-x
171. Acampora D, Di Giovannantonio LG, Garofalo A, et al. Functional Antagonism between OTX2 and NANOG Specifies a Spectrum of Heterogeneous Identities in Embryonic Stem Cells. *Stem Cell Reports*. 2017;9(5):1642-1659. doi:10.1016/j.stemcr.2017.09.019
172. Ying QL, Nichols J, Chambers I, Smith A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*. 2003. doi:10.1016/S0092-8674(03)00847-X
173. Smith AG, Heath JK, Donaldson DD, et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*. 1988. doi:10.1038/336688a0
174. Williams RL, Hilton DJ, Pease S, et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*. 1988. doi:10.1038/336684a0
175. Smith AG, Hooper ML. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol*. 1987. doi:10.1016/0012-1606(87)90132-1
176. Hirai H, Karian P, Kikyo N. Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. *Biochem J*. 2011. doi:10.1042/BJ20102152
177. Morikawa M, Koinuma D, Mizutani A, et al. BMP Sustains Embryonic Stem Cell Self-Renewal through Distinct Functions of Different Krüppel-like Factors. *Stem Cell Reports*. 2016. doi:10.1016/j.stemcr.2015.12.004
178. Qi X, Li TG, Hao J, et al. BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A*. 2004. doi:10.1073/pnas.0401367101
179. Li Z, Fei T, Zhang J, et al. BMP4 signaling acts via dual-specificity phosphatase 9 to control ERK activity in mouse embryonic stem cells. *Cell Stem Cell*. 2012. doi:10.1016/j.stem.2011.12.016
180. Boeuf H, Hauss C, De Graeve F, Baran N, Kedinger C. Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *J Cell Biol*. 1997. doi:10.1083/jcb.138.6.1207
181. Martello G, Bertone P, Smith A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J*. 2013;32(19):2561-2574. doi:10.1038/emboj.2013.177
182. Matsuda T, Nakamura T, Nakao K, et al. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J*. 1999. doi:10.1093/emboj/18.15.4261
183. Niwa H, Ogawa K, Shimosato D, Adachi K. A parallel circuit of LIF signalling

pathways maintains pluripotency of mouse ES cells. *Nature*. 2009. doi:10.1038/nature08113
184. Niwa H, Burdon T, Chambers I, Smith A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev*. 1998. doi:10.1101/gad.12.13.2048
185. Schlesinger S, Meshorer E. Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. *Dev Cell*. 2019;48(2):135-150. doi:10.1016/j.devcel.2019.01.003
186. Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One*. 2010. doi:10.1371/journal.pone.0010531
187. Efroni S, Duttagupta R, Cheng J, et al. Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*. 2008;2(5):437-447. doi:10.1016/j.stem.2008.03.021
188. Singer ZS, Yong J, Tischler J, et al. Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells. *Mol Cell*. 2014;55(2):319-331. doi:10.1016/j.molcel.2014.06.029
189. Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008. doi:10.1038/nature07107
190. Azuara V, Perry P, Sauer S, et al. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*. 2006. doi:10.1038/ncb1403
191. Christophorou MA, Castelo-Branco G, Halley-Stott RP, et al. Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature*. 2014. doi:10.1038/nature12942
192. Xiao S, Lu J, Sridhar B, et al. SMARCAD1 Contributes to the Regulation of Naive Pluripotency by Interacting with Histone Citrullination. *Cell Rep*. 2017. doi:10.1016/j.celrep.2017.02.070
193. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic Dynamics of Imprinted X Inactivation during Early Mouse Development. *Science (80-)*. 2004. doi:10.1126/science.1092727
194. Mak W, Nesterova TB, De Napoles M, et al. Reactivation of the Paternal X Chromosome in Early Mouse Embryos. *Science (80-)*. 2004. doi:10.1126/science.1092674
195. Sheardown SA, Duthie SM, Johnston CM, et al. Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell*. 1997. doi:10.1016/S0092-8674(01)80012-X
196. Sousa EJ, Stuart HT, Bates LE, et al. Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. *Cell Stem Cell*. 2018;22(6):919-928.e6. doi:10.1016/j.stem.2018.05.001
197. Panning B, Dausman J, Jaenisch R. X chromosome inactivation is mediated by Xist RNA stabilization. *Cell*. 1997. doi:10.1016/S0092-8674(00)80355-4
198. Aoto T, Saitoh N, Ichimura T, Niwa H, Nakao M. Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. *Dev Biol*. 2006. doi:10.1016/j.ydbio.2006.04.450
199. Yeom YII, Fuhrmann G, Ovitt CE, et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development*. 1996.
200. Yeo JC, Ng HH. The transcriptional regulation of pluripotency. *Cell Res*. 2013;23(1):20-32. doi:10.1038/cr.2012.172

201. Svoboda P, Flemr M. The role of miRNAs and endogenous siRNAs in maternal-to-zygotic reprogramming and the establishment of pluripotency. *EMBO Rep*. 2010. doi:10.1038/embor.2010.102
202. Marson A, Levine SS, Cole MF, et al. Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. *Cell*. 2008. doi:10.1016/j.cell.2008.07.020
203. Houbaviy HB, Murray MF, Sharp PA. Embryonic stem cell-specific microRNAs. *Dev Cell*. 2003. doi:10.1016/S1534-5807(03)00227-2
204. Divisato G, Passaro F, Russo T, Parisi S. The key role of micrornas in self-renewal and differentiation of embryonic stem cells. *Int J Mol Sci*. 2020;21(17):1-21. doi:10.3390/ijms21176285
205. Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*. 2010. doi:10.1038/nature08725
206. Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol*. 2016. doi:10.1038/nrm.2016.8
207. Judson RL, Babiarz JE, Venere M, Blelloch R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol*. 2009. doi:10.1038/nbt.1535
208. Anokye-Danso F, Trivedi CM, Juhr D, et al. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*. 2011. doi:10.1016/j.stem.2011.03.001
209. Chambers I, Silva J, Colby D, et al. Nanog safeguards pluripotency and mediates germline development. *Nature*. 2007. doi:10.1038/nature06403
210. Kalmar T, Lim C, Hayward P, et al. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol*. 2009;7(7):33-36. doi:10.1371/journal.pbio.1000149
211. Karwacki-Neisius V, Göke J, Osorno R, et al. Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell*. 2013. doi:10.1016/j.stem.2013.04.023
212. Macarthur BD, Sevilla A, Lenz M, et al. Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nat Cell Biol*. 2012. doi:10.1038/ncb2603
213. Reynolds N, Latos P, Hynes-Allen A, et al. NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell*. 2012. doi:10.1016/j.stem.2012.02.020
214. Hayashi K, Lopes SMC de S, Tang F, Surani MA. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell*. 2008;3(4):391-401. doi:10.1016/j.stem.2008.07.027
215. Martinez Arias A, Brickman JM. Gene expression heterogeneities in embryonic stem cell populations: Origin and function. *Curr Opin Cell Biol*. 2011. doi:10.1016/j.ceb.2011.09.007
216. Graf T, Stadtfeld M. Heterogeneity of Embryonic and Adult Stem Cells. *Cell Stem Cell*. 2008. doi:10.1016/j.stem.2008.10.007
217. Cherry A, Daley GQ. Another horse in the meta-stable state of pluripotency. *Cell Stem Cell*. 2010. doi:10.1016/j.stem.2010.11.020
218. Loh KM, Lim B. A precarious balance: Pluripotency factors as lineage specifiers. *Cell Stem Cell*. 2011. doi:10.1016/j.stem.2011.03.013

219. Cahan P, Daley GQ. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat Rev Mol Cell Biol*. 2013. doi:10.1038/nrm3584
220. Kumar RM, Cahan P, Shalek AK, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*. 2014;516(729):56-61. doi:10.1038/nature13920
221. Huang S. Non-genetic heterogeneity of cells in development: More than just noise. *Development*. 2009. doi:10.1242/dev.035139
222. Kolodziejczyk AA, Kim JK, Tsang JCH, et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell*. 2015;17(4):471-485. doi:10.1016/j.stem.2015.09.011
223. Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*. 2008;135(5):909-918. doi:10.1242/dev.017400
224. Canham MA, Sharov AA, Ko MSH, Brickman JM. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol*. 2010. doi:10.1371/journal.pbio.1000379
225. Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. 2012. doi:10.1038/nature11244
226. Falco G, Lee SL, Stanghellini I, Bassey UC, Hamatani T, Ko MSH. Zscan4: A novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev Biol*. 2007. doi:10.1016/j.ydbio.2007.05.003
227. Davenport TG, Jerome-Majewska LA, Papaioannou VE. Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. *Development*. 2003. doi:10.1242/dev.00431
228. Luo J, Sladek R, Bader JA, Matthyssen A, Rossant J, Giguère V. Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-β. *Nature*. 1997. doi:10.1038/42022
229. Dahl JA, Reiner AH, Klungland A, Wakayama T, Collas P. Histone H3 lysine 27 methylation asymmetry on developmentally-regulated promoters distinguish the first two lineages in mouse preimplantation embryos. *PLoS One*. 2010;5(2). doi:10.1371/journal.pone.0009150
230. Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. *Genes Dev*. 2013. doi:10.1101/gad.219626.113
231. Harikumar A, Meshorer E. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO Rep*. 2015. doi:10.15252/embr.201541011
232. Tang F, Barbacioru C, Bao S, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. *Cell Stem Cell*. 2010;6(5):468-478. doi:10.1016/j.stem.2010.03.015
233. Davey RE, Onishi K, Mahdavi A, Zandstra PW. LIF-mediated control of embryonic stem cell self-renewal emerges due to an autoregulatory loop. *FASEB J*. 2007. doi:10.1096/fj.06-7852com
234. Kunath T, Saba-El-Leil MK, Almousailleakh M, Wray J, Meloche S, Smith A. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*. 2007. doi:10.1242/dev.02880
235. Ying QL, Wray J, Nichols J, et al. The ground state of embryonic stem cell self-renewal.

Nature. 2008;453(7194):519-523. doi:10.1038/nature06968
236. Buehr M, Smith A, Camus A, Merry C, Kumar PM. Genesis of embryonic stem cells. *Philos Trans R Soc B Biol Sci*. 2003. doi:10.1098/rstb.2003.1327
237. Wray J, Kalkan T, Gomez-Lopez S, et al. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol*. 2011;13(7):838-845. doi:10.1038/ncb2267
238. Buehr M, Meek S, Blair K, et al. Capture of Authentic Embryonic Stem Cells from Rat Blastocysts. *Cell*. 2008. doi:10.1016/j.cell.2008.12.007
239. Marks H, Kalkan T, Menafra R, et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. 2012;149(3):590-604. doi:10.1016/j.cell.2012.03.026
240. Wray J, Kalkan T, Smith AG. The ground state of pluripotency. *Biochem Soc Trans*. 2010;38(4):1027-1032. doi:10.1042/BST0381027
241. Morgani SM, Canham MA, Nichols J, et al. Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. *Cell Rep*. 2013;3(6):1945-1957. doi:10.1016/j.celrep.2013.04.034
242. Yagi M, Kishigami S, Tanaka A, et al. Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature*. 2017;548(7666):224-227. doi:10.1038/nature23286
243. Choi J, Huebner AJ, Clement K, et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature*. 2017. doi:10.1038/nature23274
244. Shimizu T, Ueda J, Ho JC, et al. Dual inhibition of Src and GSK3 maintains mouse embryonic stem cells, whose differentiation is mechanically regulated by Src signaling. *Stem Cells*. 2012. doi:10.1002/stem.1119
245. Hassani SN, Totonchi M, Sharifi-Zarchi A, et al. Inhibition of TGFβ Signaling Promotes Ground State Pluripotency. *Stem Cell Rev Reports*. 2014. doi:10.1007/s12015-013-9473-0
246. Li R, Zhong C, Yu Y, et al. Generation of Blastocyst-like Structures from Mouse Embryonic and Adult Cell Cultures. *Cell*. 2019;179(3):687-702.e18. doi:10.1016/j.cell.2019.09.029
247. Martin Gonzalez J, Morgani SM, Bone RA, et al. Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. *Stem Cell Reports*. 2016;7(2):177-191. doi:10.1016/j.stemcr.2016.07.009
248. Boroviak T, Loos R, Lombard P, et al. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev Cell*. 2015;35(3):366-382. doi:10.1016/j.devcel.2015.10.011
249. Papatsenko D, Darr H, Kulakovskiy I V., et al. Single-Cell Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal. *Stem Cell Reports*. 2015;5(2):207-220. doi:10.1016/j.stemcr.2015.07.004
250. Ying Q-L, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol*. 2003;21(2):183-186. doi:10.1038/nbt780
251. Hamazaki T, Oka M, Yamanaka S, Terada N. Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. *J Cell Sci*. 2004.

doi:10.1242/jcs.01489
252. Price FD, Yin H, Jones A, Van Ijcken W, Grosveld F, Rudnicki MA. Canonical Wnt signaling induces a primitive endoderm metastable state in mouse embryonic stem cells. *Stem Cells*. 2013. doi:10.1002/stem.1321
253. Doetschman TC, Eistetter H, Katz M. The in vitro development of blastocyst-derived embryonic stem cell lines: Formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol*. 1985.
254. ten Berge D, Koole W, Fuerer C, Fish M, Eroglu E, Nusse R. Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies. *Cell Stem Cell*. 2008. doi:10.1016/j.stem.2008.09.013
255. Lancaster MA, Knoblich JA. Organogenesisin a dish: Modeling development and disease using organoid technologies. *Science (80-)*. 2014. doi:10.1126/science.1247125
256. Clevers H. Modeling Development and Disease with Organoids. *Cell*. 2016. doi:10.1016/j.cell.2016.05.082
257. Van Den Brink SC, Baillie-Johnson P, Balayo T, et al. Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Dev*. 2014. doi:10.1242/dev.113001
258. Shahbazi MN, Zernicka-Goetz M. Deconstructing and reconstructing the mouse and human early embryo. *Nat Cell Biol*. 2018;20(8):878-887. doi:10.1038/s41556-018-0144-x
259. Fuchs C, Scheinast M, Pasteiner W, et al. Self-organization phenomena in embryonic stem cell-derived embryoid bodies: Axis formation and breaking of symmetry during cardiomyogenesis. *Cells Tissues Organs*. 2012. doi:10.1159/000328712
260. Turner DA, Girgin M, Alonso-Crisostomo L, et al. Anteroposterior polarity and elongation in the absence of extraembryonic tissues and of spatially localised signalling in gastruloids: Mammalian embryonic organoids. *Dev*. 2017. doi:10.1242/dev.150391
261. Beddington RSP, Robertson EJ. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development*. 1989.
262. Huang J, Deng K, Wu H, et al. Efficient Production of Mice from Embryonic Stem Cells Injected into Four- or Eight-Cell Embryos by Piezo Micromanipulation. *Stem Cells*. 2008. doi:10.1634/stemcells.2008-0164
263. Poueymirou WT, Auerbach W, Frendewey D, et al. F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat Biotechnol*. 2007. doi:10.1038/nbt1263
264. Huang Y, Osorno R, Tsakiridis A, Wilson V. In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. *Cell Rep*. 2012;2(6):1571-1578. doi:10.1016/j.celrep.2012.10.022
265. Eggan K, Akutsu H, Loring J, et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci U S A*. 2001. doi:10.1073/pnas.101118898
266. Nagy A, Gocza E, Merentes Diaz E, et al. Embryonic stem cells alone are able to support fetal development in the mouse. *Development*. 1990.
267. Eakin GS, Hadjantonakis AK, Papaioannou VE, Behringer RR. Developmental potential and behavior of tetraploid cells in the mouse embryo. *Dev Biol*. 2005.

doi:10.1016/j.ydbio.2005.09.028
268. Osorno R, Tsakiridis A, Wong F, et al. The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development*. 2012;139(13):2288-2298. doi:10.1242/dev.078071
269. Turco MY, Furia L, Dietze A, et al. Cellular heterogeneity during embryonic stem cell differentiation to epiblast stem cells is revealed by the ShcD/RaLP adaptor protein. *Stem Cells*. 2012. doi:10.1002/stem.1217
270. Camus A, Perea-Gomez A, Moreau A, Collignon J. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev Biol*. 2006. doi:10.1016/j.ydbio.2006.03.047
271. Mesnard D, Guzman-Ayala M, Constam DB. Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development*. 2006. doi:10.1242/dev.02413
272. Vallier L, Mendjan S, Brown S, et al. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development*. 2009. doi:10.1242/dev.033951
273. Feldman B, Poueymirou W, Papaioannou VE, DeChiara TM, Goldfarb M. Requirement of FGF-4 for postimplantatior mouse development. *Science (80-)*. 1995. doi:10.1126/science.7809630
274. Kurek D, Neagu A, Tastemel M, et al. Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports*. 2015. doi:10.1016/j.stemcr.2014.11.007
275. Hunt GC, Singh P, Schwarzbauer JE. Endogenous production of fibronectin is required for self-renewal of cultured mouse embryonic stem cells. *Exp Cell Res*. 2012. doi:10.1016/j.yexcr.2012.06.009
276. Factor DC, Corradin O, Zentner GE, et al. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. *Cell Stem Cell*. 2014;14(6):854-863. doi:10.1016/j.stem.2014.05.005
277. Han DW, Tapia N, Joo JY, et al. Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell*. 2010;143(4):617-627. doi:10.1016/j.cell.2010.10.015
278. Jouneau A, Ciaudo C, Sismeiro O, et al. Naive and primed murine pluripotent stem cells have distinct miRNA expression profiles. *RNA*. 2012;18(2):253-264. doi:10.1261/rna.028878.111
279. Gillich A, Bao S, Grabole N, et al. Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell*. 2012. doi:10.1016/j.stem.2012.01.020
280. Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. *Cell*. 2011;145(6):835-850. doi:10.1016/j.cell.2011.05.019
281. Song J, Saha S, Gokulrangan G, Tesar PJ, Ewing RM. DNA and chromatin modification networks distinguish stem cell pluripotent ground states. *Mol Cell Proteomics*. 2012. doi:10.1074/mcp.M111.011114
282. Zhang H, Gayen S, Xiong J, et al. MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency. *Cell Stem Cell*. 2016;18(4):481-494. doi:10.1016/j.stem.2016.02.004
283. Zhou W, Choi M, Margineantu D, et al. HIF1α induced switch from bivalent to

exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J*. 2012;31(9):2103-2116. doi:10.1038/emboj.2012.71
284. Sperber H, Mathieu J, Wang Y, et al. The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat Cell Biol*. 2015;17(12):1523-1535. doi:10.1038/ncb3264
285. Altshuler A, Verbuk M, Bhattacharya S, et al. RAS Regulates the Transition from Naive to Primed Pluripotent Stem Cells. *Stem Cell Reports*. 2018;10(3):1088-1101. doi:10.1016/j.stemcr.2018.01.004
286. Tsogtbaatar E, Landin C, Minter-Dykhouse K, Folmes CDL. Energy Metabolism Regulates Stem Cell Pluripotency. *Front Cell Dev Biol*. 2020;8(February):1-16. doi:10.3389/fcell.2020.00087
287. Sun B, Ito M, Mendjan S, et al. Status of genomic imprinting in epigenetically distinct pluripotent stem cells. *Stem Cells*. 2012. doi:10.1002/stem.793
288. Sugimoto M, Kondo M, Koga Y, et al. A simple and robust method for establishing homogeneous mouse epiblast stem cell lines by Wnt inhibition. *Stem Cell Reports*. 2015. doi:10.1016/j.stemcr.2015.02.014
289. Tsakiridis A, Huang Y, Blin G, et al. Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Dev*. 2014. doi:10.1242/dev.101014
290. Ohtsuka S, Nishikawa-Torikai S, Niwa H. E-Cadherin Promotes Incorporation of Mouse Epiblast Stem Cells into Normal Development. *PLoS One*. 2012. doi:10.1371/journal.pone.0045220
291. Butz S, Larue L. Expression of catenins during mouse embryonic development and in adult tissues. *Cell Commun Adhes*. 1995. doi:10.3109/15419069509081018
292. Sharaireh AM, Fitzpatrick LM, Ward CM, McKay TR, Unwin RD. Epithelial cadherin regulates transition between the naïve and primed pluripotent states in mouse embryonic stem cells. *Stem Cells*. 2020;38(10):1292-1306. doi:10.1002/stem.3249
293. Soncin F, Mohamet L, Eckardt D, et al. Abrogation of E-cadherin-mediated cell-cell contact in mouse embryonic stem cells results in reversible LIF-independent self-renewal. *Stem Cells*. 2009. doi:10.1002/stem.134
294. Hayashi K, Surani MA. Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development*. 2009. doi:10.1242/dev.037747
295. Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, Saitou M. A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. *Cell*. 2009. doi:10.1016/j.cell.2009.03.014
296. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*. 1977;74(8):3171-3175. doi:10.1073/pnas.74.8.3171
297. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell*. 1977;12(1):1-8. doi:10.1016/0092-8674(77)90180-5
298. Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J Exp Med*. 1944;79(2):137-158. doi:10.1084/jem.79.2.137
299. CRICK F. Central Dogma of Molecular Biology. *Nature*. 1970;227(5258):561-563. doi:10.1038/227561a0

300. Gilbert W. Why genes in pieces? *Nature*. 1978;271(5645):501-501. doi:10.1038/271501a0
301. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol*. 2014;6(6). doi:10.1101/cshperspect.a016071
302. Rogozin IB, Carmel L, Csuros M, Koonin E V. Origin and evolution of spliceosomal introns. *Biol Direct*. 2012. doi:10.1186/1745-6150-7-11
303. Hong X, Scofield DG, Lynch M. Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol*. 2006. doi:10.1093/molbev/msl111
304. Purugganan M, Wessler S. The splicing of transposable elements and its role in intron evolution. *Genetica*. 1992. doi:10.1007/BF00133728
305. CAVALIER-SMITH T. Selfish DNA and the origin of introns. *Nature*. 1985;315(6017):283-284. doi:10.1038/315283b0
306. Wang W, Yu H, Long M. Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. *Nat Genet*. 2004. doi:10.1038/ng1338
307. Li W, Tucker AE, Sung W, Kelley Thomas W, Lynch M. Extensive, recent lntron gains in Daphnia populations. *Science (80-)*. 2009. doi:10.1126/science.1179302
308. Yenerall P, Zhou L. Identifying the mechanisms of intron gain: progress and trends. *Biol Direct*. 2012. doi:10.1186/1745-6150-7-29
309. Huff JT, Zilberman D, Roy SW. Mechanism for DNA transposons to generate introns on genomic scales. *Nature*. 2016;538(7626):533-536. doi:10.1038/nature20110
310. Saha A, Kim Y, Gewirtz ADH, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res*. 2017. doi:10.1101/gr.216721.116
311. Tapial J, Ha KCH, Sterne-Weiler T, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res*. 2017;27(10):1759-1768. doi:10.1101/gr.220962.117
312. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008;40(12):1413-1415. doi:10.1038/ng.259
313. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010. doi:10.1038/nature08909
314. Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res*. 2007. doi:10.1093/nar/gkl924
315. Barbosa-Morais NL, Irimia M, Pan Q, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science (80-)*. 2012;338(6114):1587-1593. doi:10.1126/science.1230612
316. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008. doi:10.1038/nature07509
317. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. *Nature*. 2014. doi:10.1038/nature13302
318. Buljan M, Chalancon G, Eustermann S, et al. Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. *Mol Cell*. 2012. doi:10.1016/j.molcel.2012.05.039

319. Ellis JD, Barrios-Rodiles M, Çolak R, et al. Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. *Mol Cell*. 2012. doi:10.1016/j.molcel.2012.05.037
320. Yang X, Coulombe-Huntington J, Kang S, et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell*. 2016;164(4):805-817. doi:10.1016/j.cell.2016.01.029
321. Porter RS, Jaamour F, Iwase S. Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. *Mol Cell Neurosci*. 2018. doi:10.1016/j.mcn.2017.10.006
322. Kelemen O, Convertini P, Zhang Z, et al. Function of alternative splicing. *Gene*. 2013. doi:10.1016/j.gene.2012.07.083
323. Irimia M, Blencowe BJ. Alternative splicing: Decoding an expansive regulatory layer. *Curr Opin Cell Biol*. 2012. doi:10.1016/j.ceb.2012.03.005
324. Irimia M, Roy SW. Spliceosomal introns as tools for genomic and evolutionary analysis. *Nucleic Acids Res*. 2008. doi:10.1093/nar/gkn012
325. Ruskin B, Krainer AR, Maniatis T, Green MR. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*. 1984. doi:10.1016/0092-8674(84)90553-1
326. Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science (80-)*. 1984. doi:10.1126/science.6206566
327. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science (80-)*. 2008. doi:10.1126/science.1153803
328. Lambowitz AM, Zimmerly S. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*. 2011. doi:10.1101/cshperspect.a003616
329. Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*. 2009. doi:10.1016/j.cell.2009.02.009
330. Parker R, Siliciano PG, Guthrie C. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*. 1987. doi:10.1016/0092-8674(87)90564-2
331. Valcárcel J, Gaur RK, Singh R, Green MR. Interaction of U2AF65 RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA. *Science (80-)*. 1996. doi:10.1126/science.273.5282.1706
332. Grabowski PJ, Seiler SR, Sharp PA. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*. 1985. doi:10.1016/S0092-8674(85)80130-6
333. Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol*. 2011. doi:10.1101/cshperspect.a003707
334. Daguenet E, Dujardin G, Valcárcel J. The pathogenicity of splicing defects: mechanistic insights into pre- mRNA processing inform novel therapeutic approaches . *EMBO Rep*. 2015. doi:10.15252/embr.201541116
335. Konarska MM, Vilardell J, Query CC. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell*. 2006. doi:10.1016/j.molcel.2006.01.017
336. Yan C, Wan R, Shi Y. Molecular mechanisms of pre-mRNA splicing through structural

biology of the spliceosome. *Cold Spring Harb Perspect Biol*. 2019. doi:10.1101/cshperspect.a032409
337. Berget SM. Exon Recognition in Vertebrate Splicing. *J Biol Chem*. 1995;270(6):2411-2414. doi:10.1074/jbc.270.6.2411
338. Robberson BL, Cote GJ, Berget SM. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol*. 1990. doi:10.1128/mcb.10.1.84
339. Lang KM, Spritz RA. RNA splice site selection: Evidence for a 5′→3′ scanning model. *Science (80-)*. 1983. doi:10.1126/science.6304877
340. Grau-bové X, Grau-bové X. Genome Biology Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. 2012:1-21.
341. Amit M, Donyo M, Hollander D, et al. Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Rep*. 2012;1(5):543-556. doi:10.1016/j.celrep.2012.03.013
342. Chen M, Manley JL. Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*. 2009. doi:10.1038/nrm2777
343. Huelga SC, Vu AQ, Arnold JD, et al. Integrative Genome-wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins. *Cell Rep*. 2012. doi:10.1016/j.celrep.2012.02.001
344. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. In: *Journal of Computational Biology*. ; 2004. doi:10.1089/1066527041410418
345. Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008. doi:10.1261/rna.876308
346. Van Nostrand EL, Freese P, Pratt GA, et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature*. 2020;583(7818):711-719. doi:10.1038/s41586-020-2077-3
347. Cereda M, Pozzoli U, Rot G, et al. RNAmotifs: Prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol*. 2014. doi:10.1186/gb-2014-15-1-r20
348. Chou MY, Underwood JG, Nikolic J, Luu MHT, Black DL. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol Cell*. 2000. doi:10.1016/S1097-2765(00)80260-9
349. Fu XD, Ares M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet*. 2014. doi:10.1038/nrg3778
350. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (80-)*. 2003. doi:10.1126/science.1090095
351. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: Organizers of cellular biochemistry. *Nat Rev Mol Cell Biol*. 2017. doi:10.1038/nrm.2017.7
352. Kato M, McKnight SL. A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. *Annu Rev Biochem*. 2018. doi:10.1146/annurev-biochem-061516-044700
353. Li P, Banjade S, Cheng HC, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*. 2012. doi:10.1038/nature10879
354. Sterner DA, Carlo T, Berget SM. Architectural limits on split genes. *Proc Natl Acad Sci U S A*. 1996. doi:10.1073/pnas.93.26.15081

355. Gelfman S, Burstein D, Penn O, et al. Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res*. 2012. doi:10.1101/gr.119834.110
356. Pai AA, Henriques T, McCue K, Burkholder A, Adelman K, Burge CB. The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *Elife*. 2017;6:1-26. doi:10.7554/eLife.32537
357. Hiller M, Zhang Z, Backofen R, Stamm S. Pre-mRNA secondary structures influence exon recognition. *PLoS Genet*. 2007. doi:10.1371/journal.pgen.0030204
358. Estes PA, Cooke NE, Liebhaber SA. A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. *J Biol Chem*. 1992.
359. Blanchette M, Chabot B. A highly stable duplex structure sequesters the 5′ splice site region of hnRNP A1 alternative exon 7B. *RNA*. 1997.
360. Warf MB, Berglund JA. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci*. 2010. doi:10.1016/j.tibs.2009.10.004
361. Nasim FUH, Hutchison S, Cordeau M, Chabot B. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5′ splice site selection in support of a common looping out and repression mechanism. *RNA*. 2002. doi:10.1017/S1355838202024056
362. Lovci MT, Ghanem D, Marr H, et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol*. 2013. doi:10.1038/nsmb.2699
363. Baraniak AP, Lasda EL, Wagner EJ, Garcia-Blanco MA. A Stem Structure in Fibroblast Growth Factor Receptor 2 Transcripts Mediates Cell-Type-Specific Splicing by Approximating Intronic Control Elements. *Mol Cell Biol*. 2003. doi:10.1128/mcb.23.24.9327-9337.2003
364. Buratti E, Muro AF, Giombi M, Gherbassi D, Iaconcig A, Baralle FE. RNA Folding Affects the Recruitment of SR Proteins by Mouse and Human Polypurinic Enhancer Elements in the Fibronectin EDA Exon. *Mol Cell Biol*. 2004. doi:10.1128/mcb.24.3.1387-1400.2004
365. Huang H, Zhang J, Harvey SE, Hu X, Cheng C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev*. 2017. doi:10.1101/gad.305862.117
366. Han SP, Tang YH, Smith R. Functional diversity of the hnRNPs: Past, present and perspectives. *Biochem J*. 2010. doi:10.1042/BJ20100396
367. Zhou Z, Fu XD. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*. 2013. doi:10.1007/s00412-013-0407-z
368. Singh R, Valcárcel J. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol*. 2005. doi:10.1038/nsmb961
369. Krecic AM, Swanson MS. hnRNP complexes: Composition, structure, and function. *Curr Opin Cell Biol*. 1999. doi:10.1016/S0955-0674(99)80051-9
370. Long JC, Caceres JF. The SR protein family of splicing factors: Master regulators of gene expression. *Biochem J*. 2009. doi:10.1042/BJ20081501
371. Fu XD. The superfamily of arginine/serine-rich splicing factors. *RNA*. 1995.
372. Pandit S, Zhou Y, Shiue L, et al. Genome-wide Analysis Reveals SR Protein

Cooperation and Competition in Regulated Splicing. *Mol Cell*. 2013. doi:10.1016/j.molcel.2013.03.001

373. Dvinge H. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett*. 2018;592(17):2987-3006. doi:10.1002/1873-3468.13119
374. Sharma S, Falick AM, Black DL. Polypyrimidine tract binding protein blocks the 5′ splice site-dependent assembly of U2AF and the prespliceosomal e complex. *Mol Cell*. 2005. doi:10.1016/j.molcel.2005.07.014
375. Oberstrass FC. Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science (80-)*. 2005;309(5743):2054-2057. doi:10.1126/science.1114066
376. Xue Y, Zhou Y, Wu T, et al. Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping. *Mol Cell*. 2009. doi:10.1016/j.molcel.2009.12.003
377. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol*. 2008. doi:10.1038/nsmb.1375
378. Torres-Méndez A, Bonnal S, Marquez Y, et al. A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nat Ecol Evol*. 2019;3(4):691-701. doi:10.1038/s41559-019-0813-6
379. Bhate A, Parker DJ, Bebee TW, et al. ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. *Nat Commun*. 2015;6(1):8768. doi:10.1038/ncomms9768
380. Irimia M, Weatheritt RJ, Ellis JD, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*. 2014. doi:10.1016/j.cell.2014.11.035
381. Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet*. 2011;12(10):715-729. doi:10.1038/nrg3052
382. Licatalosi DD, Darnell RB. RNA processing and its regulation: Global insights into biological networks. *Nat Rev Genet*. 2010. doi:10.1038/nrg2673
383. Mallory MJ, Allon SJ, Qiu J, et al. Induced transcription and stability of CELF2 mRNA drives widespread alternative splicing during T-cell signaling. *Proc Natl Acad Sci U S A*. 2015. doi:10.1073/pnas.1423695112
384. Gabut M, Samavarchi-Tehrani P, Wang X, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*. 2011;147(1):132-146. doi:10.1016/j.cell.2011.08.023
385. Han H, Irimia M, Ross PJ, et al. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature*. 2013;498(7453):241-245. doi:10.1038/nature12270
386. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*. 2017;18(7):437-451. doi:10.1038/nrm.2017.27
387. Wong JJL, Ritchie W, Ebner OA, et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*. 2013;154(3):583-595. doi:10.1016/j.cell.2013.06.052
388. Zhang X, Chen MH, Wu X, et al. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell*. 2016. doi:10.1016/j.cell.2016.07.025
389. Cieply B, Park JW, Nakauka-Ddamba A, et al. Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by

Numerous RNA-Binding Proteins. *Cell Rep*. 2016;15(2):247-255. doi:10.1016/j.celrep.2016.03.025
390. Bebee TW, Park JW, Sheridan KI, et al. The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife*. 2015;4(September2015):1-27. doi:10.7554/eLife.08954
391. Burguera D, Marquez Y, Racioppi C, et al. Evolutionary recruitment of flexible Esrp-dependent splicing programs into diverse embryonic morphogenetic processes. *Nat Commun*. 2017;8(1). doi:10.1038/s41467-017-01961-y
392. Boutz PL, Bhutkar A, Sharp PA. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev*. 2015;29(1):63-80. doi:10.1101/gad.247361.114
393. Gotic I, Omidi S, Fleury-Olela F, Molina N, Naef F, Schibler U. Temperature regulates splicing efficiency of the cold-inducible RNA-binding protein gene Cirbp. *Genes Dev*. 2016. doi:10.1101/gad.287094.116
394. Low KH, Lim C, Ko HW, Edery I. Natural Variation in the Splice Site Strength of a Clock Gene and Species-Specific Thermal Adaptation. *Neuron*. 2008. doi:10.1016/j.neuron.2008.10.048
395. Preußner M, Goldammer G, Neumann A, et al. Body Temperature Cycles Control Rhythmic Alternative Splicing in Mammals. *Mol Cell*. 2017;67(3):433-446.e4. doi:10.1016/j.molcel.2017.06.006
396. Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. Dynamic integration of splicing within gene regulatory pathways. *Cell*. 2013;152(6):1252-1269. doi:10.1016/j.cell.2013.02.034
397. Ule J, Blencowe BJ. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell*. 2019;76(2):329-345. doi:10.1016/j.molcel.2019.09.017
398. Atlasi Y, Mowla SJ, Ziaee SAM, Gokhale PJ, Andrews PW. OCT4 Spliced Variants Are Differentially Expressed in Human Pluripotent and Nonpluripotent Cells. *Stem Cells*. 2008. doi:10.1634/stemcells.2008-0530
399. Yamazaki T, Liu L, Lazarev D, et al. TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev*. 2018:1-14. doi:10.1101/gad.316984.118
400. Das S, Jena S, Levasseur DN. Alternative splicing produces nanog protein variants with different capacities for self-renewal and pluripotency in embryonic stem cells. *J Biol Chem*. 2011. doi:10.1074/jbc.M111.290189
401. Gopalakrishnan S, Van Emburgh BO, Shan J, et al. A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding. *Mol Cancer Res*. 2009. doi:10.1158/1541-7786.MCR-09-0018
402. Ohta S, Nishida E, Yamanaka S, Yamamoto T. Global Splicing Pattern Reversion during Somatic Cell Reprogramming. *Cell Rep*. 2013;5(2):357-366. doi:10.1016/j.celrep.2013.09.016
403. Tanaka Y, Hysolli E, Su J, et al. Transcriptome Signature and Regulation in Human Somatic Cell Reprogramming. *Stem Cell Reports*. 2015;4(6):1125-1139. doi:10.1016/j.stemcr.2015.04.009
404. Kanitz A, Syed AP, Kaji K, Zavolan M. Conserved regulation of RNA processing in

somatic cell reprogramming. *BMC Genomics*. 2019;20(1):1-19. doi:10.1186/s12864-019-5438-2

405. Toh CXD, Chan JW, Chong ZS, et al. RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming. *Cell Rep*. 2016;15(12):2597-2607. doi:10.1016/j.celrep.2016.05.049
406. Lu X, Göke J, Sachs F, et al. SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat Cell Biol*. 2013;15(10):1141-1152. doi:10.1038/ncb2839
407. Hirsch CL, Akdemir ZC, Wang L, et al. Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. *Genes Dev*. 2015;29(8):803-816. doi:10.1101/gad.255109.114
408. Lu Y, Loh YH, Li H, et al. Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell Stem Cell*. 2014. doi:10.1016/j.stem.2014.04.002
409. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*. 2009. doi:10.1038/nsmb.1545
410. Zhang T, Lin Y, Liu J, et al. Rbm24 Regulates Alternative Splicing Switch in Embryonic Stem Cell Cardiac Lineage Differentiation. *Stem Cells*. 2016. doi:10.1002/stem.2366
411. Corsini NS, Peer AM, Moeseneder P, et al. Coordinated Control of mRNA and rRNA Processing Controls Embryonic Stem Cell Pluripotency and Differentiation. *Cell Stem Cell*. 2018;22(4):543-558.e12. doi:10.1016/j.stem.2018.03.002
412. Thompson S, Clarke AR, Pow AM, Hooper ML, Melton DW. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell*. 1989;56(2):313-321. doi:10.1016/0092-8674(89)90905-7
413. Nesterova TB, Popova BC, Cobb BS, et al. Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics Chromatin*. 2008;1(1):2. doi:10.1186/1756-8935-1-2
414. Collier AJ, Panula SP, Schell JP, et al. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. *Cell Stem Cell*. 2017;20(6):874-890.e7. doi:10.1016/j.stem.2017.02.014
415. Taei A, Rasooli P, Braun T, Hassani S, Baharvand H. Signal regulators of human naïve pluripotency. *Exp Cell Res*. 2020:111924. doi:10.1016/j.yexcr.2020.111924
416. Irie N, Weinberger L, Tang WWC, et al. SOX17 is a critical specifier of human primordial germ cell fate. *Cell*. 2015. doi:10.1016/j.cell.2014.12.013
417. Liu Y, Bourgeois CF, Pang S, et al. The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon. *PLoS Genet*. 2009;5(11). doi:10.1371/journal.pgen.1000707
418. Heinrich B, Zhang Z, Raitskin O, et al. Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regionsin pre-mRNA. *J Biol Chem*. 2009;284(21):14303-14315. doi:10.1074/jbc.M901026200
419. Hofmann Y. hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum Mol Genet*. 2002;11(17):2037-2049. doi:10.1093/hmg/11.17.2037
420. Wang Y, Wang J, Gao L, Stamm S, Andreadis A. An SRp75/hnRNPG complex interacting with hnRNPE2 regulates the 5' splice site of tau exon 10, whose

misregulation causes frontotemporal dementia. *Gene*. 2011;485(2):130-138. doi:10.1016/j.gene.2011.06.020
421. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res*. 2017;45(10):6051-6063. doi:10.1093/nar/gkx141
422. Zhou KI, Shi H, Lyu R, et al. Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. *Mol Cell*. 2019;76(1):70-81.e9. doi:10.1016/j.molcel.2019.07.005
423. Ehrmann I, Crichton JH, Gazzara MR, et al. An ancient germ cell-specific RNA-binding protein protects the germline from cryptic splice site poisoning. *Elife*. 2019;8. doi:10.7554/eLife.39304
424. Moursy A, Allain FHT, Cléry A. Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic Acids Res*. 2014;42(10):6659-6672. doi:10.1093/nar/gku244
425. Yang Y, Park JW, Bebee TW, et al. Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. *Mol Cell Biol*. 2016;36(11):1704-1719. doi:10.1128/MCB.00019-16
426. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005. doi:10.1073/pnas.0506580102
427. Cano A, Pérez-Moreno MA, Rodrigo I, et al. The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol*. 2000. doi:10.1038/35000025
428. Orr-Urtreger A, Bedford MT, Do MS, Eisenbach L, Lonai P. Developmental expression of the α receptor for platelet-derived growth factor, which is deleted in the embryonic lethal Patch mutation. *Development*. 1992.
429. Sakurai H, Era T, Jakt LM, et al. In Vitro Modeling of Paraxial and Lateral Mesoderm Differentiation Reveals Early Reversibility. *Stem Cells*. 2006. doi:10.1634/stemcells.2005-0256
430. Yamaguchi TP, Dumont DJ, Conlon RA, Breitman ML, Rossant J. Flk-1, an flt-related receptor tyrosine kinase is an early marker for endothelial cell precursors. *Development*. 1993;118(2):489-498.
431. Yamashita J, Itoh H, Hirashima M, et al. Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature*. 2000. doi:10.1038/35040568
432. Warzecha CC, Jiang P, Amirikian K, et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J*. 2010;29(19):3286-3300. doi:10.1038/emboj.2010.195
433. Warzecha CC, Shen S, Xing Y, Carstens RP. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol*. 2009;6(5):546-562. doi:10.4161/rna.6.5.9606
434. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. *Mol Cell*. 2009;33(5):591-601. doi:10.1016/j.molcel.2009.01.025
435. Fernando RI, Litzinger M, Trono P, Hamilton DH, Schlom J, Palena C. The T-box

transcription factor Brachyury promotes epithelial-mesenchymal transition in human tumor cells. *J Clin Invest*. 2010. doi:10.1172/JCI38379

436. Roselli M, Fernando RI, Guadagni F, et al. Brachyury, a driver of the epithelial-mesenchymal transition, is overexpressed in human lung tumors: An opportunity for novel interventions against lung cancer. *Clin Cancer Res*. 2012. doi:10.1158/1078-0432.CCR-11-3211
437. Tosic J, Kim GJ, Pavlovic M, et al. Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nat Cell Biol*. 2019. doi:10.1038/s41556-019-0423-1
438. Arnold SJ, Hofmann UK, Bikoff EK, Robertson EJ. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development*. 2008. doi:10.1242/dev.014357
439. Han H, Braunschweig U, Gonatopoulos-Pournatzis T, et al. Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol Cell*. 2017;65(3):539-553.e7. doi:10.1016/j.molcel.2017.01.011
440. Zhao L, Tian D, Xia M, Macklin WB, Feng Y. Rescuing qkv dysmyelination by a single isoform of the selective RNA-binding protein QKI. *J Neurosci*. 2006;26(44):11278-11286. doi:10.1523/JNEUROSCI.2677-06.2006
441. Zhao L, Mandler MD, Yi H, Feng Y. Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proc Natl Acad Sci*. 2010;107(44):19061-19066. doi:10.1073/pnas.1007487107
442. Hardy RJ. Molecular defects in the dysmyelinating mutant quaking. *J Neurosci Res*. 1998;51(4):417-422. doi:10.1002/(SICI)1097-4547(19980215)51:4<417::AID-JNR1>3.0.CO;2-F
443. Darbelli L, Choquet K, Richard S, Kleinman CL. Transcriptome profiling of mouse brains with qkI-deficient oligodendrocytes reveals major alternative splicing defects including self-splicing. *Sci Rep*. 2017;7(1):1-13. doi:10.1038/s41598-017-06211-1
444. Fujita N, Sato S, Kurihara T, Inuzuka T, Takahashi Y, Miyatake T. Developmentally regulated alternative splicing of brain myelin-associated glycoprotein mRNA is lacking in the quaking mouse. *FEBS Lett*. 1988;232(2):323-327. doi:10.1016/0014-5793(88)80762-2
445. De Bruin RG, Shiue L, Prins J, et al. Quaking promotes monocyte differentiation into pro-atherogenic macrophages by controlling pre-mRNA splicing and gene expression. *Nat Commun*. 2016;7(2). doi:10.1038/ncomms10846
446. Noveroske JK, Lai L, Gaussin V, et al. Quaking is essential for blood vessel development. *Genesis*. 2002;32(3):218-230. doi:10.1002/gene.10060
447. Hall MP, Nagel RJ, Fagg WS, et al. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *Rna*. 2013;19(5):627-638. doi:10.1261/rna.038422.113
448. Van Der Veer EP, De Bruin RG, Kraaijeveld AO, et al. Quaking, an RNA-binding protein, is a critical regulator of vascular smooth muscle cell phenotype. *Circ Res*. 2013;113(9):1065-1075. doi:10.1161/CIRCRESAHA.113.301302
449. Conn SJ, Pillman KA, Toubia J, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell*. 2015;160(6):1125-1134. doi:10.1016/j.cell.2015.02.014
450. Hayakawa-Yano Y, Suyama S, Nogami M, et al. An RNA-binding protein, Qki5,

regulates embryonic neural stem cells through pre-mRNA processing in cell adhesion signaling. *Genes Dev*. 2017;31(18):1910-1925. doi:10.1101/gad.300822.117
451. Zong FY, Fu X, Wei WJ, et al. The RNA-Binding Protein QKI Suppresses Cancer-Associated Aberrant Splicing. *PLoS Genet*. 2014;10(4). doi:10.1371/journal.pgen.1004289
452. Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY. Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res*. 2015;43(10):5130-5144. doi:10.1093/nar/gkv210
453. Wu JI, Reed RB, Grabowski PJ, Artzt K. Function of quaking in myelination: Regulation of alternative splicing. *Proc Natl Acad Sci*. 2002;99(7):4233-4238. doi:10.1073/pnas.072090399
454. Wang Y, Vogel G, Yu Z, Richard S. The QKI-5 and QKI-6 RNA Binding Proteins Regulate the Expression of MicroRNA 7 in Glial Cells. *Mol Cell Biol*. 2013;33(6):1233-1243. doi:10.1128/MCB.01604-12
455. Chen AJ, Paik JH, Zhang H, et al. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. *Genes Dev*. 2012. doi:10.1101/gad.189001.112
456. YU F, JIN L, YANG G, JI L, WANG F, LU Z. Post-transcriptional repression of FOXO1 by QKI results in low levels of FOXO1 expression in breast cancer cells. *Oncol Rep*. 2014;31(3):1459-1465. doi:10.3892/or.2013.2957
457. Lakiza O, Frater L, Yoo Y, et al. STAR proteins quaking-6 and GLD-1 regulate translation of the homologues GLI1 and tra-1 through a conserved RNA 3′UTR-based mechanism. *Dev Biol*. 2005;287(1):98-110. doi:10.1016/j.ydbio.2005.08.038
458. Hafner M, Landthaler M, Burger L, et al. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*. 2010;141(1):129-141. doi:10.1016/j.cell.2010.03.009
459. Zhang Y, Lu Z, Ku L, Chen Y, Wang H, Feng Y. Tyrosine phosphorylation of QKI mediates developmental signals to regulate mRNA metabolism. *EMBO J*. 2003;22(8):1801-1810. doi:10.1093/emboj/cdg171
460. Paronetto MP, Achsel T, Massiello A, Chalfant CE, Sette C. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J Cell Biol*. 2007. doi:10.1083/jcb.200701005
461. Stoss O, Novoyatleva T, Gencheva M, Olbrich M, Benderska N, Stamm S. P59fyn-mediated phosphorylation regulates the activity of the tissue-specific splicing factor rSLM-1. *Mol Cell Neurosci*. 2004. doi:10.1016/j.mcn.2004.04.011
462. Matter N, Herrlich P, König H. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*. 2002;420(6916):691-695. doi:10.1038/nature01153
463. Buecker C, Srinivasan R, Wu Z, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*. 2014;14(6):838-853. doi:10.1016/j.stem.2014.04.003
464. Acampora D, Di Giovannantonio LG, Simeone A. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development*. 2013;140(1):43-55. doi:10.1242/dev.085290
465. Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. *Science*

(80-). 1999;286(5441):964-967. doi:10.1126/science.286.5441.964
466. Lingenfelter PA, Delbridge ML, Thomas S, et al. Expression and conservation of processed copies of the RBMX gene. *Mamm Genome*. 2001;12(7):538-545. doi:10.1007/s00335001-0003-z
467. Venables JP. RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. *Hum Mol Genet*. 2000;9(5):685-694. doi:10.1093/hmg/9.5.685
468. Dreyfuss G. hnRNP Proteins and the Biogenesis of mRNA. *Annu Rev Biochem*. 1993. doi:10.1146/annurev.biochem.62.1.289
469. Elliott DJ, Oghene K, Makarov G, et al. Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. *J Cell Sci*. 1998;111(9):1255-1265.
470. Bellott DW, Hughes JF, Skaletsky H, et al. Mammalian y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*. 2014. doi:10.1038/nature13206
471. Casola C, Betrán E. The genomic impact of gene retrocopies: What have we learned from comparative genomics, population genomics, and transcriptomic analyses? *Genome Biol Evol*. 2017;9(6):1351-1373. doi:10.1093/gbe/evx081
472. de la Grange P, Gratadou L, Delord M, Dutertre M, Auboeuf D. Splicing factor and exon profiling across human tissues. *Nucleic Acids Res*. 2010. doi:10.1093/nar/gkq008
473. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biol*. 2004. doi:10.1186/gb-2004-5-10-r74
474. Grosso AR, Gomes AQ, Barbosa-Morais NL, et al. Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res*. 2008. doi:10.1093/nar/gkn463
475. Clark TA, Schweitzer AC, Chen TX, et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol*. 2007. doi:10.1186/gb-2007-8-4-r64
476. Soumillon M, Necsulea A, Weier M, et al. Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. *Cell Rep*. 2013. doi:10.1016/j.celrep.2013.05.031
477. Licatalosi DD. Roles of RNA-binding proteins and post-transcriptional regulation in driving male germ cell development in the mouse. In: *Advances in Experimental Medicine and Biology*. ; 2016. doi:10.1007/978-3-319-29073-7_6
478. Xia B, Yan Y, Baron M, et al. Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell*. 2020;180(2):248-262.e21. doi:10.1016/j.cell.2019.12.015
479. Elliott DJ, Dalgliesh C, Hysenaj G, Ehrmann I. RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology. *Int J Biochem Cell Biol*. 2019;108(September 2018):1-6. doi:10.1016/j.biocel.2018.12.014
480. Elliott DJ. An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene. *Hum Mol Genet*. 2000;9(14):2117-2124. doi:10.1093/hmg/9.14.2117
481. Maymon BB, Paz G, Elliott DJ, et al. Localization of the germ cell-specific protein, hnRNP G-T, in testicular biopsies of azoospermic men. *Acta Histochem*. 2002;104(3):255-261. doi:10.1078/0065-1281-00657
482. Wang J, Gao QS, Wang Y, Lafyatis R, Stamm S, Andreadis A. Tau exon 10, whose

missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. *J Neurochem*. 2004;88(5):1078-1090. doi:10.1046/j.1471-4159.2003.02232.x
483. Wang PJ. X chromosomes, retrogenes and their role in male reproduction. *Trends Endocrinol Metab*. 2004;15(2):79-83. doi:10.1016/j.tem.2004.01.007
484. Blomen VA, Májek P, Jae LT, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science (80-)*. 2015. doi:10.1126/science.aac7557
485. Hart T, Chandrashekhar M, Aregger M, et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. 2015;163(6):1515-1526. doi:10.1016/j.cell.2015.11.015
486. Wang T, Birsoy K, Hughes NW, et al. Identification and characterization of essential genes in the human genome. *Science (80-)*. 2015. doi:10.1126/science.aac7041
487. Dickinson ME, Flenniken AM, Ji X, et al. High-throughput discovery of novel developmental phenotypes. *Nature*. 2016;537(7621):508-514. doi:10.1038/nature19356
488. Adamson B, Smogorzewska A, Sigoillot FD, King RW, Elledge SJ. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat Cell Biol*. 2012;14(3):318-328. doi:10.1038/ncb2426
489. Zhao S, Korzan WJ, Chen CC, Fernald RD. Heterogeneous nuclear ribonucleoprotein A/B and G inhibits the transcription of gonadotropin-releasing-hormone 1. *Mol Cell Neurosci*. 2008;37(1):69-84. doi:10.1016/j.mcn.2007.08.015
490. Takemoto T, Nishio Y, Sekine O, et al. RBMX is a novel hepatic transcriptional regulator of SREBP-1c gene response to high-fructose diet. *FEBS Lett*. 2007;581(2):218-222. doi:10.1016/j.febslet.2006.12.014
491. Becker JS, McCarthy RL, Sidoli S, et al. Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol Cell*. 2017;68(6):1023-1037.e15. doi:10.1016/j.molcel.2017.11.030
492. Mazeyrat S, Saut N, Mattei M-G, Mitchell MJ. RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. *Nat Genet*. 1999;22(3):224-226. doi:10.1038/10282
493. Kanhoush R, Beenders B, Perrin C, Moreau J, Bellini M, Penrad-Mobayed M. Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts. *Nucleus*. 2010;1(1):109-122. doi:10.4161/nucl.1.1.10857
494. Shashi V, Xie P, Schoch K, et al. The RBMX gene as a candidate for the Shashi X-linked intellectual disability syndrome. *Clin Genet*. 2015. doi:10.1111/cge.12511
495. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem Sci*. 1998. doi:10.1016/S0968-0004(98)01208-0
496. Jangi M, Sharp PA. Building robust transcriptomes with master splicing factors. *Cell*. 2014;159(3):487-498. doi:10.1016/j.cell.2014.09.054
497. Yang P, Humphrey SJ, Cinghu S, et al. Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. *Cell Syst*. 2019;8(5):427-445.e10. doi:10.1016/j.cels.2019.03.012
498. Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by deleterious

mutations but not gene knockdowns. *Nature*. 2015;524(7564):230-233. doi:10.1038/nature14580

499. El-Brolosy MA, Kontarakis Z, Rossi A, et al. Genetic compensation triggered by mutant mRNA degradation. *Nature*. 2019;568(7751):193-197. doi:10.1038/s41586-019-1064-z
500. Lunde BM, Moore C, Varani G. RNA-binding proteins: Modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007. doi:10.1038/nrm2178
501. Dreumont N, Bourgeois CF, Lejeune F, et al. Human RBMY regulates germline-specific splicing events by modulating the function of the serine/arginine-rich proteins 9G8 and Tra2-β. *J Cell Sci*. 2010;123(1):40-50. doi:10.1242/jcs.055889
502. Stoilov P, Dauod R, Nayler O, Stamm S. Human tra2-beta1 autoregulates its protein concentration by infuencing alternative splicing of its pre-mRNA. *Hum Mol Genet*. 2004;13(5):509-524. doi:10.1093/hmg/ddh051
503. Soulard M, Valle V Della, Siomi MC, et al. hnRNP G: Sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res*. 1993;21(18):4210-4217. doi:10.1093/nar/21.18.4210
504. Corley SM, Gready JE. Identification of the RGG Box Motif in Shadoo: RNA-Binding and Signaling Roles? *Bioinform Biol Insights*. 2008;2:BBI.S1075. doi:10.4137/bbi.s1075
505. Thandapani P, O'Connor TR, Bailey TL, Richard S. Defining the RGG/RG Motif. *Mol Cell*. 2013;50(5):613-623. doi:10.1016/j.molcel.2013.05.021
506. Fornerod M. RS and RGG repeats as primitive proteins at the transition between the RNA and RNP worlds. *Nucleus*. 2012;3(1):4-5. doi:10.4161/nucl.18631
507. Ozdilek BA, Thompson VF, Ahmed NS, White CI, Batey RT, Schwartz JC. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res*. 2017;45(13):7984-7996. doi:10.1093/nar/gkx460
508. Lemaire S, Fontrodona N, Aubé F, et al. Characterizing the interplay between gene nucleotide composition bias and splicing. *Genome Biol*. 2019;20(1):1-21. doi:10.1186/s13059-019-1869-y
509. Betts MJ, Russell RB. Amino-Acid Properties and Consequences of Substitutions. In: *Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data: Second Edition*. ; 2007. doi:10.1002/9780470059180.ch13
510. Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*. 2005. doi:10.1038/nature04209
511. Hegele A, Kamburov A, Grossmann A, et al. Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Mol Cell*. 2012. doi:10.1016/j.molcel.2011.12.034
512. Yu H, Tardivo L, Tam S, et al. Next-generation sequencing to generate interactome datasets. *Nat Methods*. 2011. doi:10.1038/nmeth.1597
513. Rolland T, Taşan M, Charloteaux B, et al. A proteome-scale map of the human interactome network. *Cell*. 2014. doi:10.1016/j.cell.2014.10.050
514. Baker BS. Sex in flies: The splice of life. *Nature*. 1989;340(6234):521-524. doi:10.1038/340521a0
515. Barbieri I, Tzelepis K, Pandolfini L, et al. Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature*. 2017;552(7683):126-131. doi:10.1038/nature24678

516. Ke S, Pandya-Jones A, Saito Y, et al. m 6 A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev*. 2017;31(10):990-1006. doi:10.1101/gad.301036.117
517. Louloupi A, Ntini E, Conrad T, Ørom UAV. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell Rep*. 2018;23(12):3429-3437. doi:10.1016/j.celrep.2018.05.077
518. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012;485(7397):201-206. doi:10.1038/nature11112
519. Xiao W, Adhikari S, Dahal U, et al. Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell*. 2016;61(4):507-519. doi:10.1016/j.molcel.2016.01.012
520. Alarcón CR, Goodarzi H, Lee H, Liu X, Tavazoie S, Tavazoie SF. HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell*. 2015;162(6):1299-1308. doi:10.1016/j.cell.2015.08.011
521. Geula S, Moshitch-Moshkovitz S, Dominissini D, et al. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science (80-)*. 2015;347(6225):1002-1006. doi:10.1126/science.1261417
522. Batista PJ, Molinie B, Wang J, et al. M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*. 2014;15(6):707-719. doi:10.1016/j.stem.2014.09.019
523. Waisman A, Vazquez Echegaray C, Solari C, et al. Inhibition of Cell Division and DNA Replication Impair Mouse-Naïve Pluripotency Exit. *J Mol Biol*. 2017;429(18):2802-2815. doi:10.1016/j.jmb.2017.06.020
524. Soldi M, Bonaldi T. The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. *Mol Cell Proteomics*. 2013. doi:10.1074/mcp.M112.024307
525. Strom AR, Emelyanov A V., Mir M, Fyodorov D V., Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. *Nature*. 2017. doi:10.1038/nature22989
526. Larson AG, Elnatan D, Keenen MM, et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*. 2017. doi:10.1038/nature22822
527. Wu J, Huang B, Chen H, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. 2016. doi:10.1038/nature18606
528. Tosolini M, Brochard V, Adenot P, et al. Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. *Sci Rep*. 2018;8(1):1-14. doi:10.1038/s41598-018-23822-4
529. Hassan-Zadeh V, Rugg-Gunn P, Bazett-Jones DP. DNA methylation is dispensable for changes in global chromatin architecture but required for chromocentre formation in early stem cell differentiation. *Chromosoma*. 2017;126(5):605-614. doi:10.1007/s00412-017-0625-x
530. Fagoonee S, Bearzi C, Di Cunto F, et al. The RNA Binding Protein ESRP1 Fine-Tunes the Expression of Pluripotency-Related Factors in Mouse Embryonic Stem Cells. *PLoS One*. 2013;8(8):1-13. doi:10.1371/journal.pone.0072300
531. Su AI, Cooke MP, Ching KA, et al. Large-scale analysis of the human and mouse

transcriptomes. *Proc Natl Acad Sci U S A*. 2002. doi:10.1073/pnas.012025199
532. Göttgens E-L, Span PN, Zegers MM. Roles and Regulation of Epithelial Splicing Regulatory Proteins 1 and 2 in Epithelial–Mesenchymal Transition. In: *International Review of Cell and Molecular Biology*. Vol 327. Elsevier Inc.; 2016:163-194. doi:10.1016/bs.ircmb.2016.06.003
533. Brown RL, Reinke LM, Damerow MS, et al. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest*. 2011. doi:10.1172/JCI44540
534. Leontieva O V., Ionov Y. RNA-binding motif protein 35A is a novel tumor suppressor for colorectal cancer. *Cell Cycle*. 2009;8(3):490-497. doi:10.4161/cc.8.3.7679
535. Díaz-Díaz C, Fernandez de Manuel L, Jimenez-Carretero D, Montoya MC, Clavería C, Torres M. Pluripotency Surveillance by Myc-Driven Competitive Elimination of Differentiating Cells. *Dev Cell*. 2017;42(6):585-599.e4. doi:10.1016/j.devcel.2017.08.011
536. Guan R, El-Rass S, Spillane D, et al. rbm47, a novel RNA binding protein, regulates zebrafish head development. *Dev Dyn*. 2013;242(12):1395-1404. doi:10.1002/dvdy.24039
537. Fossat N, Radziewic T, Jones V, Tourle K, Tam PPL. Conditional restoration and inactivation of Rbm47 reveal its tissue-context requirement for viability and growth. *Genesis*. 2016;54(3):115-122. doi:10.1002/dvg.22920
538. Fossat N, Tourle K, Radziewic T, et al. C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep*. 2014;15(8):903-910. doi:10.15252/embr.201438450
539. Vanharanta S, Marney CB, Shu W, et al. Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer Cancer Biology and Genetics Program , Memorial Sloan-Kettering Cancer MRC Cancer Unit , MRC / Hutchison Research Centre , University of Camb. *Elife*. 2014:1-24. doi:10.7554/eLife.02734
540. Treiber T, Treiber N, Plessmann U, et al. A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Mol Cell*. 2017;66(2):270-284.e13. doi:10.1016/j.molcel.2017.03.014
541. Gu KL, Zhang Q, Yan Y, et al. Pluripotency-associated miR-290/302 family of microRNAs promote the dismantling of naive pluripotency. *Cell Res*. 2016. doi:10.1038/cr.2016.2
542. Moradi S, Sharifi-Zarchi A, Ahmadi A, et al. Small RNA Sequencing Reveals Dlk1-Dio3 Locus-Embedded MicroRNAs as Major Drivers of Ground-State Pluripotency. *Stem Cell Reports*. 2017;9(6):2081-2096. doi:10.1016/j.stemcr.2017.10.009
543. Wang Y, Hussein AM, Somasundaram L, et al. MicroRNAs regulating human and mouse naïve pluripotency. *Int J Mol Sci*. 2019;20(23):1-19. doi:10.3390/ijms20235864
544. Du P, Pirouz M, Choi J, et al. An Intermediate Pluripotent State Controlled by MicroRNAs Is Required for the Naive-to-Primed Stem Cell Transition. *Cell Stem Cell*. 2018;22(6):851-864.e5. doi:10.1016/j.stem.2018.04.021
545. Stadler B, Ivanovska I, Mehta K, et al. Characterization of microRNAs Involved in Embryonic Stem Cell States. *Stem Cells Dev*. 2010;19(7):935-950. doi:10.1089/scd.2009.0426

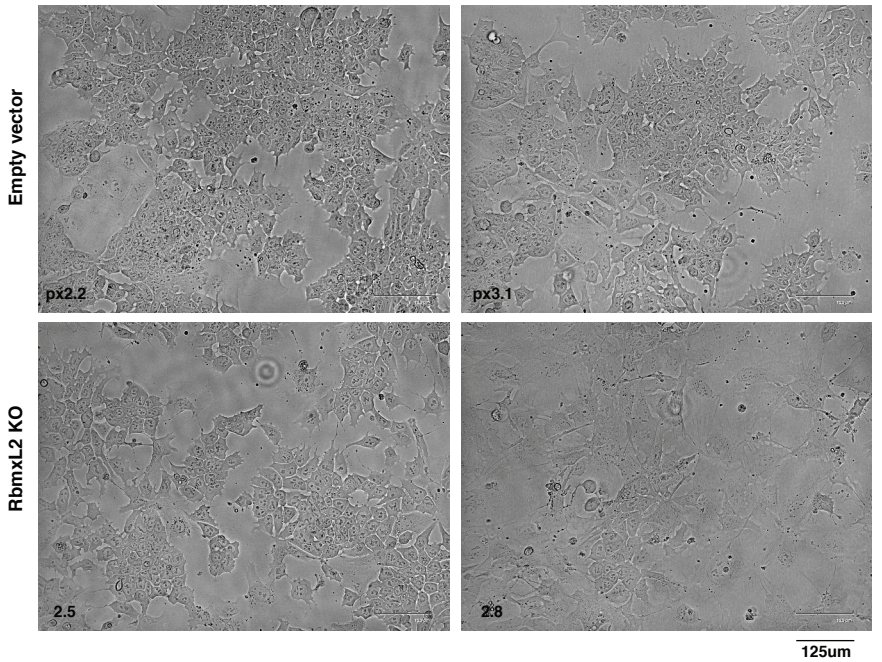
546. Yeganeh M, Seyedjafari E, Kamrani FA, Ghaemi N. RNA-binding protein Rbm47 binds to Nanog in mouse embryonic stem cells. *Mol Biol Rep*. 2013;40(7):4391-4396. doi:10.1007/s11033-013-2528-0
547. Navarro P, Festuccia N, Colby D, et al. OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. *EMBO J*. 2012;31(24):4547-4562. doi:10.1038/emboj.2012.321
548. Barral A, Rollan I, Sanchez-Iranzo H, et al. Nanog regulates Pou3f1 expression at the exit from pluripotency during gastrulation. *Biol Open*. 2019;8(11):1-10. doi:10.1242/bio.046367
549. Shapiro IM, Cheng AW, Flytzanis NC, et al. An EMT–Driven Alternative Splicing Program Occurs in Human Breast Cancer and Modulates Cellular Phenotype. Spinner NB, ed. *PLoS Genet*. 2011;7(8):e1002218. doi:10.1371/journal.pgen.1002218
550. Ishii H, Saitoh M, Sakamoto K, et al. Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *J Biol Chem*. 2014;289(40):27386-27399. doi:10.1074/jbc.M114.589432
551. Ficz G, Branco MR, Seisenberger S, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*. 2011. doi:10.1038/nature10008
552. Ficz G, Hore TA, Santos F, et al. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*. 2013. doi:10.1016/j.stem.2013.06.004
553. Pillman KA, Phillips CA, Roslan S, et al. miR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein Quaking. *EMBO J*. 2018;37(13):e99016. doi:10.15252/embj.201899016
554. Shingu T, Ho AL, Yuan L, et al. Qki deficiency maintains stemness of glioma stem cells in suboptimal environment by downregulating endolysosomal degradation. *Nat Genet*. 2017;49(1):75-86. doi:10.1038/ng.3711
555. Mellman I. ENDOCYTOSIS AND MOLECULAR SORTING. *Annu Rev Cell Dev Biol*. 1996;12(1):575-625. doi:10.1146/annurev.cellbio.12.1.575
556. Keppetipola NM, Yeom KH, Hernandez AL, Bui T, Sharma S, Black DL. Multiple determinants of splicing repression activity in the polypyrimidine tract binding proteins, PTBP1 and PTBP2. *RNA*. 2016. doi:10.1261/rna.057505.116
557. Amir-Ahmady B, Boutz PL, Markovtsov V, Phillips ML, Black DL. Exon repression by polypyrimidine tract binding protein. *RNA*. 2005. doi:10.1261/rna.2250405
558. Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou M-Y, Black DL. Cooperative Assembly of an hnRNP Complex Induced by a Tissue-Specific Homolog of Polypyrimidine Tract Binding Protein. *Mol Cell Biol*. 2000. doi:10.1128/mcb.20.20.7463-7479.2000
559. Boutz PL, Stoilov P, Li Q, et al. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev*. 2007. doi:10.1101/gad.1558107
560. Spellman R, Llorian M, Smith CWJ. Crossregulation and Functional Redundancy between the Splicing Regulator PTB and Its Paralogs nPTB and ROD1. *Mol Cell*. 2007. doi:10.1016/j.molcel.2007.06.016
561. Llorian M, Schwartz S, Clark TA, et al. Position-dependent alternative splicing

activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol*. 2010. doi:10.1038/nsmb.1881

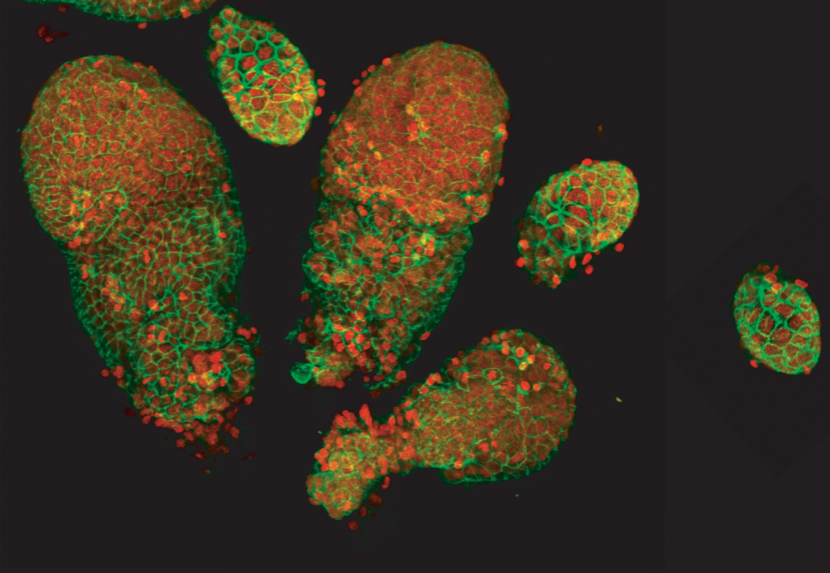
562. Linares AJ, Lin CH, Damianov A, Adams KL, Novitch BG, Black DL. The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. *Elife*. 2015. doi:10.7554/eLife.09268
563. Chan RC, Black DL. The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol Cell Biol*. 1997. doi:10.1128/mcb.17.8.4667
564. Modafferi EF, Black DL. Combinatorial control of a neuron-specific exon. *RNA*. 1999. doi:10.1017/S1355838299990155
565. Vuong JK, Lin CH, Zhang M, Chen L, Black DL, Zheng S. PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell Rep*. 2016. doi:10.1016/j.celrep.2016.11.034
566. Côté J, Boisvert FM, Boulanger MC, Bedford MT, Richard S. Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. *Mol Biol Cell*. 2003. doi:10.1091/mbc.E02-08-0484
567. Babic I, Jakymiw A, Fujita DJ. The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity. *Oncogene*. 2004. doi:10.1038/sj.onc.1207484
568. Justice MJ, Bode VC. Three ENU-induced alleles of the murine quaking locus are recessive embryonic lethal mutations. *Genet Res*. 1988. doi:10.1017/S0016672300024101
569. Cornacchia D, Zhang C, Zimmer B, et al. Lipid Deprivation Induces a Stable , Naive-to-Primed Intermediate State of Pluripotency in Human PSCs Article Lipid Deprivation Induces a Stable , Naive-to-Primed Intermediate State of Pluripotency in Human PSCs. *Stem Cell*. 2019:1-17. doi:10.1016/j.stem.2019.05.001
570. Kalkan T, Bornelöv S, Mulas C, et al. Complementary Activity of ETV5, RBPJ, and TCF3 Drives Formative Transition from Naive Pluripotency. *Cell Stem Cell*. 2019:785-801. doi:10.1016/j.stem.2019.03.017
571. Neagu A, van Genderen E, Escudero I, et al. In vitro capture and characterization of embryonic rosette-stage pluripotency between naïve and primed states. *Nat Cell Biol*. doi:10.1038/s41556-020-0508-x
572. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013. doi:10.1038/nprot.2013.143
573. Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep*. 2014;4:4-9. doi:10.1038/srep05400
574. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009. doi:10.1186/gb-2009-10-3-r25
575. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009. doi:10.1093/bioinformatics/btp616
576. Braunschweig U, Barbosa-Morais NL, Pan Q, et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*. 2014;24(11):1774-1786.

doi:10.1101/gr.177790.114
577. Labbé RM, Irimia M, Currie KW, et al. A Comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells*. 2012;30(8):1734-1745. doi:10.1002/stem.1144
578. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009. doi:10.1186/1471-2105-10-48
579. Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011. doi:10.1371/journal.pone.0021800
580. Nowotschin S, Setty M, Kuo YY, et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature*. 2019;569(7756):361-367. doi:10.1038/s41586-019-1127-1
581. Arkell RM, Tam PPL. Initiating head development in mouse embryos: Integrating signalling and transcriptional activity. *Open Biol*. 2012;2(MARCH). doi:10.1098/rsob.120030
582. McDole K, Guignard L, Amat F, et al. In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell*. 2018;175(3):859-876.e33. doi:10.1016/j.cell.2018.09.031
583. Park E, Pan Z, Zhang Z, Lin L, Xing Y. The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am J Hum Genet*. 2018;102(1):11-26. doi:10.1016/j.ajhg.2017.11.002
584. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17(1):19-32. doi:10.1038/nrg.2015.3

Annex

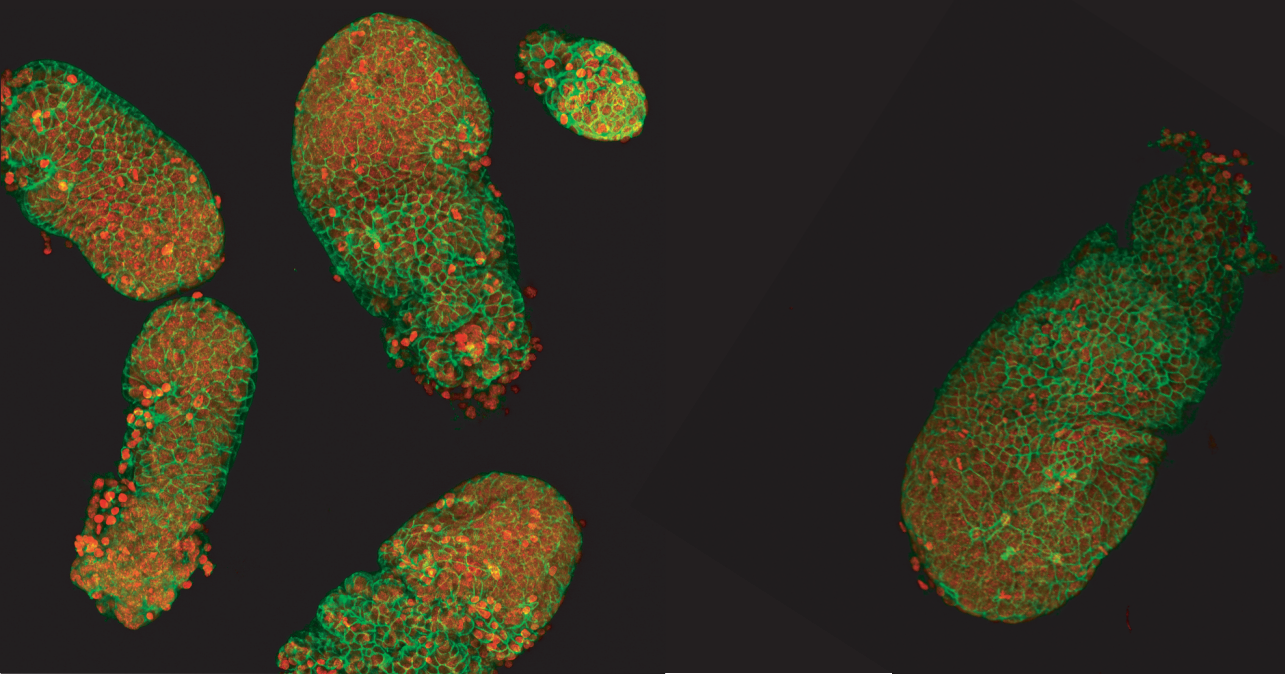


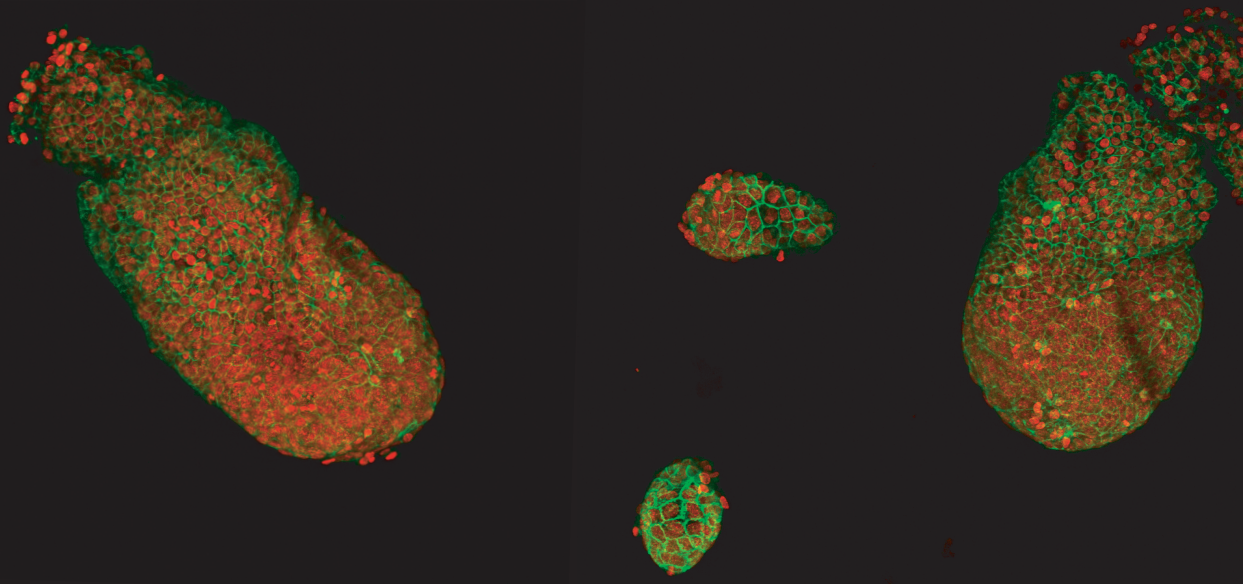
Supplementary Figure 1: Distinct morphology of RbmxL2 KO cell line in SL condition. Bright-field images of cell lines generated using CRISPR/Cas9 gene editing system for RbmxL2 gene depletion shows evident morphology difference in one of the KO clones. Control cell lines correspond to cells transfected with the empty CRISPR/Cas9 vector, clones were selected and expanded in parallel to the KO cell lines.



Manu Bárbara Laura Yamile
Marta Demi Chris Javi Patryk
Lucía Nicco Thom Andre Jon
Cede Cristina Beth Ludo Reza
Mireya Mireia Carla Lara Marco
Alex Marina
MamaTaper
Ali Nino
Marc Juan
Gaia Esther
Vero Lisa Diego Adriano
Fede Lukas Marina A
Anna David Sherell Julia
Claudia Paul Cate Fernando
PhDreps Sophie
Juan Fátima AbuelaVictoria
AbuelaMiguel Miguel Elena
Mima Perico Franklin Olga
Maleli Manolo Elia Juanito
Manolito Carlitos Miguelito Arturo
Jose Maite JuanCarlos Raquel Ichi
María Lizy Ana Lucía Gabi Ainara
Angel Flori Jenny Enza Peter
Javi Sara Mamá Papá Antonio.

Llega el final de este camino, aún resuena en mí una mezcla de emoción, energía, locura y cansancio, resaca de la escritura una la tesis supongo, que hace que a pesar de haber pensado en que llegase en este momento durante ocho largos años, me sea súper difícil expresar en palabras lo agradecida que estoy a tantas personas que han hecho que este momento sea posible. Pero he de hacerlo, es importante porque todos vosotros tenéis un pedacito de este tocho y de todos los hallazgos que contiene. Lo primero, sobretodo agradecer a Manu por haberme dado una oportunidad desde el principio, por creer en mí y en mis capacidades incluso cuando yo dudaba de ellas. También por darme la esperanza y la felicidad de pensar que había otra forma de hacer ciencia. A Bárbara porque desde el principio me conoció, me enseñó, supo ver que la transición naïve-to-primed era una pregunta fascinante estando al pie del cañón en todas las grandes preguntas y dilemas de la tesis, apoyándome para que pudiese sacar un sentido y que no me saliera demasiado del camino. Gracias también por la libertad que los dos me habéis dado, quizás no lo veía cuando pasaba, pero creo que ha sido clave para que pudiera desarrollar un espíritu científico crítico e independiente. También he de agradecer





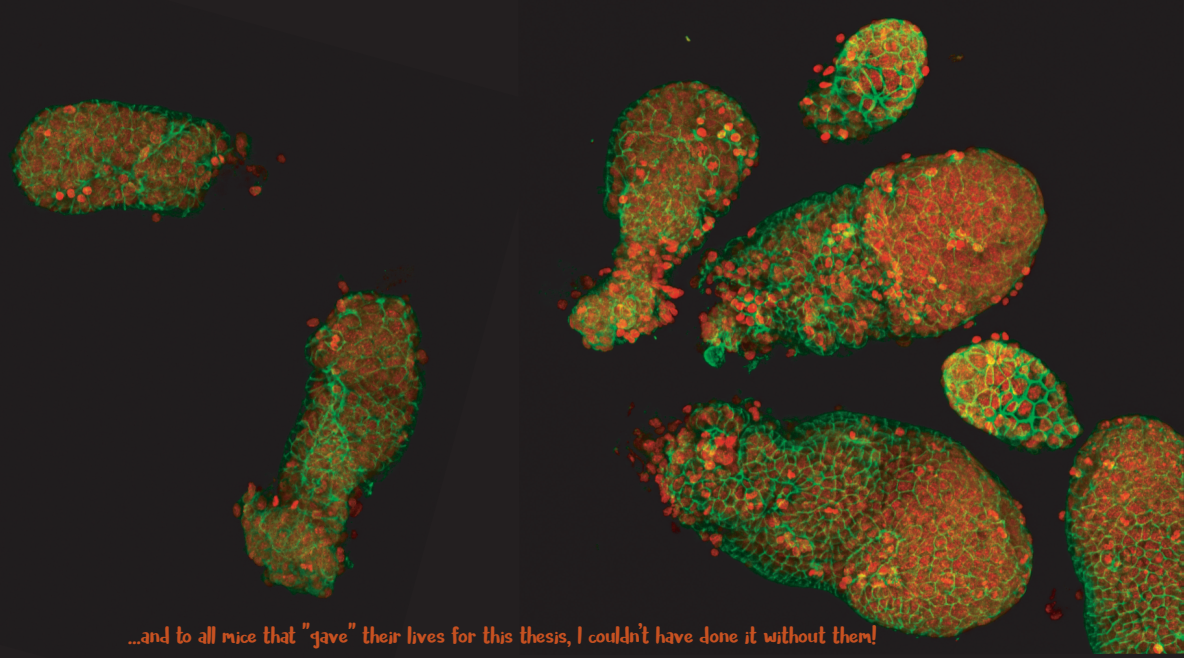
la labor de Marta y Lucía que tanto me ayudaron con las sopotocientasmil líneas que genere, ha sido genial trabajar codo con codo y desde luego, esto no hubiera sido posible sin vosotras.

I also want to thank the rest of the lab for the cool discussions, the fun nights, the late nights, for teaching me, for learning together and because you all make a great team where working in science is always exciting.

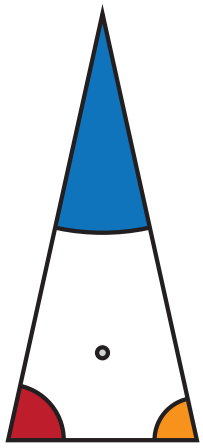
A Laura, la meva companya de viatge des del principi, gràcies per escoltar-me i donar-me suport sempre. A Jami, que pasase lo que pasase siempre estaba ahí para todo, desde una buena discusión científica, un análisis bioinformático, un gran viaje o una necesitada chela. Thanks to all the people in the CRG that have helped and were there for science and fun. To Peter & Enza for their contagious love for science and because my stay with you in Cambridge changed me and this thesis.

I also want to thank Boehringer Ingelheim Fonds for believing in me and in my future, giving me the confidence and the means to do so, a privilege that is granted to so few. A los taperos ¡Qué familia más magnífica! os quiero a todos y cada uno de vosotros, no os imagináis el efecto terapeutico que tiene el cruzar la terraza hasta la cuarta todos los días y el haber compartido tantas comidas y viajes con vosotros. La tesis es el taper, y el taper es la tesis, sin uno no hubiera conseguido el otro y desde luego, a pesar de los momentos más duros de la tesis, vosotros sois los que habeis hecho que todo esto haya merecido la pena.

Gracias. A mi familia, porque me habeis apoyado tantísimo todos estos años. Papá y mamá gracias, por estar siempre al otro lado del telefono, por creer en mí absolutamente siempre y por entender el gran sacrificio que suponía la enorme aventura del doctorado. Os quiero muchísimo. Y a Antonio ¡porque esta tesis también es tuya y porque quiero muchas más aventuras juntos!



...and to all mice that "gave" their lives for this thesis, I couldn't have done it without them!



The naïve-to-primed AS triangle. Each corner represents pluripotent cells grown in one condition (2iL, SL and AF), each line represents a comparison with another condition with the length representing the total number of differential AS exons identified. Finally, each angle connects two lines and represents the overlapping exon skipping events in the two connected comparisons.