

On the molecular basis of mammalian totipotency



Kumulative Dissertation
der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

vorgelegt von
Diego Rodriguez-Terrones

München, den 15.01.2019

Diese Dissertation wurde angefertigt
unter der Leitung von **Prof. Dr. Maria-Elena Torres-Padilla**
am Institut für Epigenetik und Stammzellen
des Helmholtz Zentrum Münchens

Erstgutachter: Prof. Dr. Maria-Elena Torres-Padilla

Zweitgutachter: Prof. Dr. Wolfgang Enard

Tag der Einreichung: 15.01.2019

Tag der mündlichen Prüfung: 23.05.2019

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist. Ich erkläre weiter, dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

Diego Rodriguez-Terrones
München, January 07th, 2019

Table of Contents

LIST OF PUBLICATIONS	7
LIST OF UNPUBLISHED MANUSCRIPTS	7
SUMMARY	9
AIMS	11
INTRODUCTION	13
MAMMALIAN PRE-IMPLANTATION DEVELOPMENT	15
A BRIEF HISTORY OF MAMMALS	15
EMBRYONIC IMPLANTATION IN MAMMALS	16
CELLULAR POTENCY DURING PRE-IMPLANTATION DEVELOPMENT	18
THE TOTIPOTENT EMBRYO	19
TOTIPOTENCY ACROSS MAMMALS	19
THE MOLECULAR BASIS OF MAMMALIAN TOTIPOTENCY	20
METABOLIC FEATURES OF TOTIPOTENT CELLS	21
TOTIPOTENT CELLS UNDERGO AN EXTENSIVE PROCESS OF CHROMATIN REMODELING	21
TRANSCRIPTIONAL FEATURES OF TOTIPOTENT CELLS	23
TRANSPOSABLE ELEMENTS	23
<i>IN VITRO</i> MODEL SYSTEMS FOR THE STUDY OF MAMMALIAN PLURIPOTENCY AND TOTIPOTENCY	24
MOUSE EMBRYONIC STEM CELLS ARE AN <i>IN VITRO</i> MODEL SYSTEM OF PLURIPOTENCY	24
MOUSE 2-CELL-LIKE CELLS ARE AN <i>IN VITRO</i> MODEL SYSTEM OF TOTIPOTENCY	25
RESULTS	27
PART 1 – THE MOLECULAR BASIS OF TOTIPOTENT-LIKE CELLS IN CULTURE	27
EARLY EMBRYONIC-LIKE CELLS ARE INDUCED BY DOWNREGULATING REPLICATION-DEPENDENT CHROMATIN ASSEMBLY	29
A MOLECULAR ROADMAP FOR THE EMERGENCE OF EARLY-EMBRYONIC-LIKE CELLS IN CULTURE	57
PART 2 – THE MOLECULAR BASIS OF MAMMALIAN TOTIPOTENCY <i>IN VIVO</i>	105
A SYSTEMATIC COMPARISON OF MAMMALIAN PRE-IMPLANTATION DEVELOPMENT TRANSCRIPTOMES	107
ABSTRACT	107

INTRODUCTION	107
MATERIALS AND METHODS	109
RESULTS	112
SUV4-20 ACTIVITY IN THE PREIMPLANTATION MOUSE EMBRYO CONTROLS TIMELY REPLICATION	125
DISCUSSION	155
PART 1 – THE MOLECULAR BASIS OF TOTIPOTENT-LIKE CELLS IN CULTURE	157
THE ESC TO ZSCAN4 ⁺ CELL TRANSITION	157
THE ZSCAN4 ⁺ TO 2-CELL-LIKE CELL TRANSITION	159
2-CELL-LIKE CELLS ENABLE THE IDENTIFICATION OF NOVEL MOLECULAR FEATURES OF TOTIPOTENT CELLS	160
PART 2 – THE MOLECULAR BASIS OF MAMMALIAN TOTIPOTENCY <i>IN VIVO</i>	162
CONCLUDING REMARKS	163
REFERENCES	164
ANNEX I – REVIEW ARTICLE	171
ANNEX II – <i>CURRICULUM VITAE</i>	189
ANNEX III – COPYRIGHT STATEMENTS	195
ACKNOWLEDGEMENTS	201

List of publications

Ishiuchi, T.*; Enriquez-Gasca, R., Mizutani, E., Boškovič, A., Ziegler-Birling, C., **Rodriguez-Terrones, D.**, Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671.

Eid, A.*, **Rodriguez-Terrones, D.**, Burton, A., and Torres-Padilla, M.E. (2016). SUV4-20 activity in the preimplantation mouse embryo controls timely replication. *Genes Dev.* 30, 2513–2526.

Rodriguez-Terrones, D.*, Gaume, X.*, Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* 50, 106–119.

Rodriguez-Terrones D. & Torres-Padilla M.E. (2018). Nimble and ready to mingle: transposon outbursts of early development. *Trends in Genetics* 10, 806-820.

List of unpublished manuscripts

Rodriguez-Terrones, D.*, Noll, C., Aksoy, I., Simmet, K., Kurome, M., Zakhartchenko V., Beaujean, N., Wolf, E. and Torres-Padilla, M.E. (Unpublished manuscript). A systematic comparison of mammalian pre-implantation development transcriptomes.

Summary

The transient capacity to autonomously form and organize all of the embryonic and extra-embryonic tissues involved in the development of a complete organism is termed totipotency. In mammals, totipotency is a feature restricted to the earliest cells of the pre-implantation embryo, which harbor this unique capacity during the first 1-3 cell cycles, depending on the species. However, our understanding of the regulatory mechanisms responsible for the establishment, maintenance and termination of such a highly plastic regulatory state remains limited. Mammalian totipotency occurs concomitantly to a set of highly-intermingled biological processes such as global chromatin remodeling, an unusual set of metabolic characteristics and the de-repression of the vast majority of transposable elements, and it is unclear whether these processes act to sustain it. Following a general overview of these processes, in this dissertation I present my contributions to a body of work on an *in vitro* model system for mammalian totipotency, which exhibits certain molecular features of the *in vivo* totipotent state. Afterwards, in the second part of this thesis, I present the transcriptional analyses that I have conducted with the aim of understanding the role of transposable element transcription during pre-implantation development. Overall, this work describes a set of phenomena that arise in totipotent cells *in vivo* and in totipotent-like cells *in vitro* and explores how recapitulating certain molecular features of totipotent cells in pluripotent cells induces a totipotent-like state *in vitro*.

Aims

Part 1. The molecular basis of totipotent-like cells in culture

- Generate a roadmap for the spontaneous emergence of 2-cell-like cells in mouse embryonic stem cell cultures.
- Identify the molecular drivers that underlie the exit of the embryonic stem cell state towards the 2-cell-like state and their mode of action.
- Determine whether 2-cell-like cells recapitulate other features of the totipotent 2-cell embryo beyond gene expression changes.

Part 2. The molecular basis of mammalian totipotency *in vivo*

- Develop an approach to jointly assess gene expression changes and transcription initiation sites from single cell or single embryo samples.
- Systematically generate transcriptional profiles using this method at different timepoints of pre-implantation development across 5 mammalian species.
- Identify a conserved totipotency-associated transcriptional program.
- Determine the degree of evolutionary conservation of transposable element expression during mammalian pre-implantation development.

Introduction

Mammalian pre-implantation development

A brief history of mammals

About 210 million years ago – shortly after the giant supercontinent Pangea began to break up, but prior to the advent of the unresolved cataclysm that marked the end of the Triassic (Figure 1) – a set of little insectivorous carnivores emerged as the first group of **mammals** (Kemp, 2005). Armed with big brains, highly differentiated dentition, active thermal homeostasis and a rather avant-garde jaw, these tiny – yet sturdy – mammals had already procured the near entirety of modern mammalian features. Although the precise timing is yet to be determined (Vaughan et al., 2013), two main lineages would emerge shortly after the appearance of these first mammals: i) the **Monotreme** lineage, which has survived up to our days in the oviparous forms of the platypus and of four species of echidnas; and ii) the **Therian** lineage, that would go on to give rise to all viviparous mammals, including marsupials and eutherians. Notably, all of these pioneering mammals most likely still relied on egg-laying for reproduction, despite at this point already possessing the vast majority of morphological and physiological features that are typically associated with modern mammals.

Over the course of the first two thirds of mammalian history – some 140 million years during which dinosaurs dominated the world’s terrestrial ecosystems – mammals would diversify and give rise to the cladistic diversity with which we are now acquainted. Mesozoic mammals would turn out to be tremendously inventive, proliferating and branching out into numerous lineages despite the fact that none of them seem to have grown past the size of an average cat (Kemp, 2005; Vaughan et al., 2013). In particular, one of the most extraordinary innovations that would arise during this period would be the advent of embryonic **implantation** and its associated viviparity within the

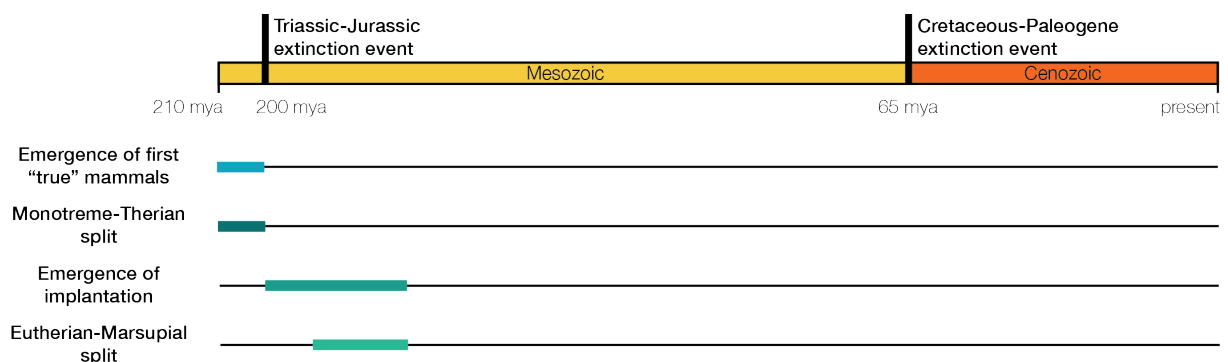


Figure 1. Timeline of mammalian evolution.

Therian lineage (Figure 2), at some point before their split into the **Marsupial** and **Eutherian** lineages 170-190 million years ago (Williamson et al., 2014). By adhering to the mother's uterus in order to obtain additional nutrients, an implanted therian embryo would be able to develop for an extended period of time within the mother without entirely relying on an egg's yolk. While marsupials would devise rather short-lived placentas to achieve this task and would give birth to comparatively undeveloped young, eutherians would further advance these mechanisms and would undertake the vast majority of their developmental program upon implantation (Roberts et al., 2016). As I will discuss further below, implantation would turn out to be one key novelty with deep implications for the landscape of cellular potency itself.

Finally, the last chapter of mammalian history commenced approximately 65 million years ago, following the mass extinction event that brought the end of the dinosaurs (Kemp, 2005). With the dinosaurs gone, this episode of mammalian evolution would see the emergence of the first large bodied eutherians and marsupials, which would quickly fill the ecological niches left behind by the dinosaurs. Despite initially possessing a virtually identical set of morphological features, the better-implanting eutherian mammals would eventually outcompete their marsupial counterparts and today represent the vast majority of mammalian species. Overall, mammalian evolution can be partially understood as the development of a series of increasingly sophisticated sustenance mechanisms that arose to nourish the developing embryo.

Embryonic implantation in mammals

Implantation conferred early mammals the capacity to develop for an extended period of time within the mother, in the absence of a large provision of yolk in the egg. Already in the last common ancestor of monotremes and therians (Figure 2), a specialized tissue called the **trophoblast** had arisen which would permit the absorption of vital nutrients from the mother (Frankenberg, 2018). For reasons that are not entirely understood, monotremes – and probably their last common ancestor with therians – had dispensed of a substantial fraction of yolk content in favor of absorbing nutrients from the uterine medium through a very primitive **placenta** (Frankenberg, 2018; Selwood and Johnson, 2006). These nutrients are essential for development and will sustain the monotreme embryo *in utero* until it reaches the 19-20 somite stage and is finally laid as an egg.

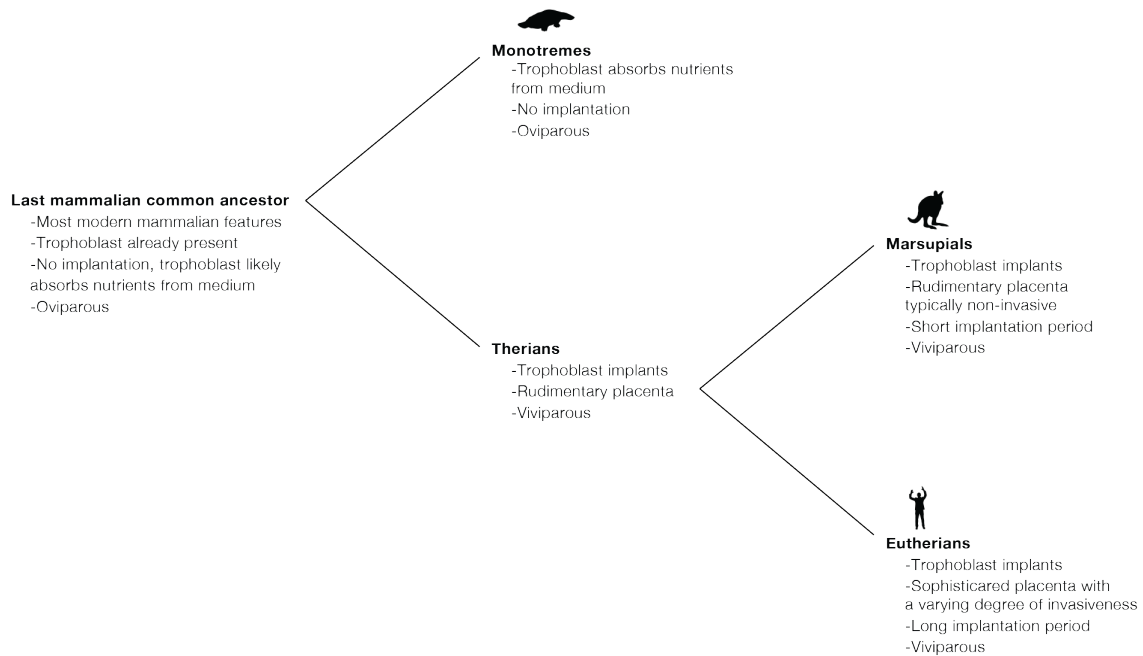


Figure 2. Evolution of mammalian implantation.

While the monotreme trophoblast never lies in direct contact with the uterine endometrium, therian embryos would eventually evolve the capacity to break the outer coating of the egg and enter in direct contact with the uterus in order to exchange nutrients (Frankenberg, 2018). Towards the last third of the marsupial pregnancy, the trophoblast will implant into the uterine endometrium with a varying degree of invasiveness depending on the marsupial species (Hughes, 1974), and a rather rudimentary placenta will be formed. After a very brief period of organogenesis, an extremely altricial neonate will be born and will conclude the greater part of morphogenesis attached to the mother's teat (Smith, 2001). Eutherians, on the other hand, implant considerably earlier and complete a significantly greater fraction of their development inside the womb. In contrast to the two thirds of in-womb development that marsupials complete before implantation, eutherian mice embryos implant already at the 4th day of their ~19 day developmental program (Wang and Dey, 2006).

Despite its crucial character, it remains unclear how exactly did the trophoblast cell lineage that gives rise to the placenta arise over the course of mammalian evolution (Roberts et al., 2016). Although functionally comparable yolk-associated tissues are known to exist in birds and reptiles (Frankenberg, 2018; Selwood and Johnson, 2006), it is unclear whether they are actually homologous to the mammalian trophoblast in terms of cellular and molecular identity. Across all three Monotreme, Marsupial and Eutherian mammalian lineages, the very first cell fate decision in the embryo involves a choice between the trophoblast and the pluriblast fates, and occurs during

the very first 4 - 7 cell cycles, depending on the species (Selwood and Johnson, 2006). Indeed, no obviously homologous cellular lineage is known to exist in any other amniote in terms of timing and function, suggesting that the trophoblast is either a mammalian novelty or radical transposition of the developmental timing of an existing cellular lineage. Thus, the emergence of implantation was only possible because of the appearance of a novel cellular lineage at the beginning of development, and the remodeling of the cell potency landscape to make room for a cellular state with the expanded cellular potency required to produce it.

Cellular potency during pre-implantation development

Pre-implantation development refers to the window of mammalian development that initiates following fertilization of the oocyte by the sperm, and that ends several days later at the time of implantation (Figure 3). Fertilization produces a rather peculiar cell called the zygote, which harbors both of the parental genomes inherited from the gametes in two physically segregated nuclei. One cell division later, both of the parental genomes will be incorporated into the single nuclei of the two daughter cells, although certain epigenetic asymmetries inherent to their parental origin will remain. In the case of the mouse embryo, several cell divisions later during the 8-cell stage, the embryo initiates a process called compaction – through which cells become polarized; upon their next division to the morula stage, outer and inner cells are formed for the first time, and the former become preferentially specified to the trophoblast lineage. Another morphological change, called cavitation, starts to take place at this point and results in the generation of a fluid-filled cavity called the blastocoel. From this point on, the embryo will be termed a blastocyst and will consist of a laminar sphere of trophoblast cells and a cumulus of pluripotent cells – termed the inner cell mass or pluriblast – on the inner side of one of the poles of this sphere.

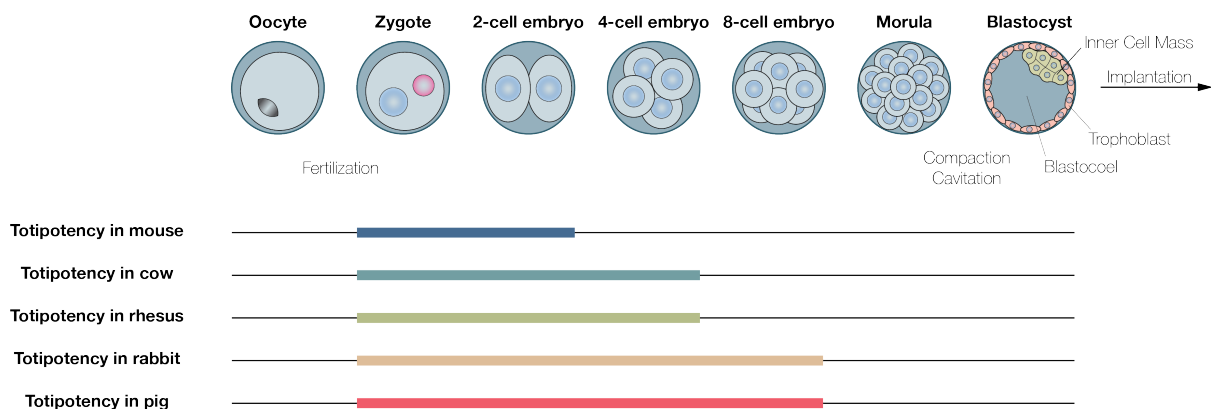


Figure 3. Overview of eutherian pre-implantation development. During this period, the developmental stage is referred to according to the number of cells present in the embryo and individual cells are termed blastomeres.

In terms of cellular potency, the various cells that arise over this window of development can be classified into four main categories: totipotent, plenipotent, pluripotent and trophoblast cells (Condic, 2014). In mouse, the **totipotent** window (Figure 3) includes the zygote and 2-cell stage embryo, since individual blastomeres of each of these two stages can produce an entire organism on their own (Casser et al., 2017; Tarkowski, 1959). In contrast, cells from the 4-cell and 8-cell stages are **plenipotent** because they have yet to be allocated to any of the two lineages of the early blastocyst: the **pluripotent** inner cell mass – which will give rise to the embryo proper and some extra-embryonic tissues – and the **trophoblast** – which will give rise to the placenta (Kelly, 1975; Tabansky et al., 2013). While at the morula stage these two lineages have already started to be specified, these cells are actually still bipotent and will not commit to either fate until the formation of the blastocyst (Rossant and Vijn, 1980). Although in the early blastocyst outer cells have already committed irreversibly to their trophoblast fate, inner cells will not commit to the pluripotent fate for another cell division and can be directed to contribute towards the trophoblast if physically transplanted (Posfai et al., 2017).

The totipotent embryo

How exactly does a single cell give rise to an entire multicellular organism is one of the longest-running and most fascinating questions in biology. This transient capacity to autonomously form and organize all of the embryonic and extra-embryonic tissues involved in the development of a new organism is termed **totipotency**. In mammals, totipotency is a feature restricted to the earliest cells of the pre-implantation embryo. Although its definition entails different interpretations (Condic, 2014) – and is used loosely in some instances (Macfarlan et al., 2012; Morgani et al., 2013) – in this work I will refer to totipotency as the capacity of a single cell to autonomously and successfully form and organize all of the embryonic and extra-embryonic tissues that emerge during the development of a new organism (Ishiuchi and Torres-Padilla, 2013). For example, over the course of **mouse** development, only the zygote and each of its two daughter cells are strictly totipotent since only they can produce an entire organism on their own. In this case, it is said that the last **twinning-stage** over the course of mouse development is the 2-cell stage, since only up to this stage can genetically-identical twin offspring be derived by physically separating each of the individual blastomeres.

Totipotency across mammals

Interestingly, the developmental window during which totipotency manifests is not fixed and varies among mammalian species (Figure 3). In **bovine** and **sheep** embryos, all four of the 4-cell stage

blastomeres are capable of developing into an entire calve or lamb when transferred individually into surrogate mothers (Johnson et al., 1995; Willadsen, 1981). Similarly, in **Rhesus monkeys**, embryos composed of only two 8-cell stage blastomeres are occasionally capable of developing to term, further providing evidence of the surprising fact that in some species, a quarter-embryo – although in this particular case, not a single cell – is sufficient to form an entire organism (Chan, 2000). Even more remarkable, however, are the reports that individual 8-cell stage **pig, sheep** and **rabbit** blastomeres are potentially totipotent and can occasionally give rise to a live piglet, lamb or kitten on their own (Moore et al., 1968; Saito and Niemann, 1991; Willadsen, 1981). Although these experiments have also been attempted in mouse embryos – the best studied model system for early mammalian development – no live mouse pup has been reported to arise from a single 4-cell stage or 8-cell stage blastomere. Notably, single blastomeres of the 4-cell stage do produce a small pseudo blastocyst, but this blastocyst frequently lacks derivatives of the pluripotent ICM and is mostly composed of trophoblast cells (Rossant, 1976). Remarkably, it appears that compaction and cavitation seem to initiate at the same time in mouse embryos derived from a single blastomere and in intact embryos (Tarkowski and Wróblewska, 1967), suggesting the existence of a regulative, time-sensing mechanism that actively limits totipotency. However, considering that it is well established that mouse 4-cell stage blastomeres have the capacity to contribute to both the ICM and the trophoblast (Kelly, 1975; Tabansky et al., 2013), it follows that the apparent loss of totipotency in these 4-cell stage blastomeres might actually result from a premature restriction in cellular potency occurring at a later time, before the specification of sufficient cells of each lineage is achieved.

The molecular basis of mammalian totipotency

Nevertheless, our understanding of the molecular mechanisms that regulate totipotency in mammals is still very limited. In addition to the technical difficulties inherent to working with the pre-implantation embryo, our limited understanding is also a consequence of the complexity of biological processes spanning the totipotent window. In mouse, the totipotent zygote and 2-cell stage embryos are characterized by an unusual metabolism (Kaneko, 2016), the absence of conventional heterochromatin (Eid et al., 2016; Fadloun et al., 2013a), an atypical chromatin organization (Burton and Torres-Padilla, 2014), and a rather unique transcriptional landscape characterized by the expression of usually silent genetic elements such as retrotransposons, tandem repeats and totipotency-specific genes that are only active during this period (Evsikov et al., 2006; Fadloun et al., 2013b; Göke et al., 2015; Peaston et al., 2004; Rodriguez-Terrones and Torres-Padilla, 2018). Below, I will summarize three of the unusual biological processes that occur

at the onset of mammalian development that might underlie – either jointly or individually – the exceptional plasticity of totipotent cells.

Metabolic features of totipotent cells

In terms of **cellular metabolism**, the mouse pre-implantation embryo exhibits a set of rather unusual properties that most likely reflect its disconnection from the mother's blood supply, its exposure to the oviduct and uterine medium – which might possess an evolutionary conserved role dating back to the monotreme means of embryo nutrition (see above) – and the need to produce sufficient amounts of certain critical metabolites to aid in the chromatin remodeling process (Nagaraj et al., 2017). While **glucose** is known to be present in the oviduct (Gardner and Leese, 1990), mouse pre-implantation embryos are unable to use this metabolite for energy production prior to the morula stage (Brinster, 1965). Intriguingly, while glucose is not able to fuel energy metabolism between the zygote and 8-cell stage, its presence is essential for the embryo's developmental progression past the 2-cell stage and a brief exposure to glucose as short as 1 minute is sufficient to prevent these embryos from undergoing developmental arrest (Chatot et al., 1994). Indeed, it appears that the earliest stages of pre-implantation development rely instead on monocarboxylates such as **pyruvate** and **lactate** (Brinster, 1965) to satisfy their energetic demands through oxidative phosphorylation (Houghton et al., 1996), and to generate sufficient amounts of critical metabolites such as acetyl-CoA and α -ketoglutarate (Nagaraj et al., 2017). These metabolites might be necessary to fuel the chromatin remodeling process that occurs after fertilization, and in agreement with this, transposition of the Pyruvate Dehydrogenase enzyme complex from the mitochondria to the nucleus at the 2-cell stage has been shown to be essential for the embryonic genome to be successfully activated (Nagaraj et al., 2017). ***Because it's possible – and even likely – that the unique metabolic features of the early pre-implantation embryo contribute to its exceptional plasticity, during my PhD I have explored whether reproducing these metabolic conditions aids in the reprogramming process to a totipotent-like state which will be described further below. Because of its ongoing nature, however, this work has not been included in this cumulative thesis.***

Totipotent cells undergo an extensive process of chromatin remodeling

Following fertilization, parental genomes are subject to an intense process of **chromatin remodeling** that is thought to underlie the reprogramming of the highly differentiated gametes into totipotent cells. In particular, the paternal genome is enwrapped in protamines rather than histones and is thus the subject of an intense wave of histone deposition (Santenard et al., 2010).

Globally, the earliest cells of the mouse embryo possess a completely atypical chromatin organization, that is evident even in a DAPI staining of genomic DNA. In contrast to the punctual organization of **pericentromeric heterochromatin** into **chromocentres** that is present from the 4-cell stage on, oocytes and 2-cell stage embryos instead organize these same genomic regions into a set of ring-like structure called **nucleolar-like-bodies** (Burton and Torres-Padilla, 2014; Kopečný et al., 1989). In terms of chromatin modifications, mouse pre-implantation embryos exhibit further unique characteristics, such as global DNA demethylation (Eckersley-Maslin et al., 2018) and the apparent lack of conventional constitutive heterochromatin (Fadloun et al., 2013a). Overall, because of the reprogramming process that occurs at the onset of mammalian development, chromatin in totipotent cells is particularly open (Wu et al., 2016) and rather devoid of repressive histone modifications. It remains an open question how exactly this unique chromatin context underlies the totipotent capacity of these cells. *In this dissertation, Eid et al. describe the developmental arrest of mouse pre-implantation embryos upon induction of the H4K20me3 constitutive heterochromatin mark. By analyzing gene expression data in manipulated and control embryos, I contributed to the determination of the molecular mechanisms underlying this developmental arrest. Overall, the downregulation of constitutive heterochromatin at the beginning of mammalian development appears to be essential for normal developmental progression and, perhaps, totipotency.*

	Totipotency window (range of twinning stages)	ZGA timing (major wave)	Earliest transcription detected (start of minor wave)
Mouse	Zygote – 2-cell stage	2-cell stage (Woodland and Graham, 1969)	Zygote (Mintz, 1964)
Cow	Zygote – 4-cell stage	8-cell – 16-cell stages (Camous et al., 1986)	Zygote (Memili and First, 1999)
Rhesus	Zygote – 4-cell stage	4-cell – 8-cell stages (Schramm and Bavister, 1999)	2-cell stage (Schramm and Bavister, 1999)
Rabbit	Zygote – 8-cell stage	8-cell – 16-cell stages (Christians et al., 1994)	Zygote (Christians et al., 1994)
Pig	Zygote – 8-cell stage	4-cell stage (Arrell et al., 1991)	4-cell stage (Tománek et al., 1989)

Table 1. Timing of the totipotency window and of zygotic genome activation across several eutherian species. In Rhesus monkeys and Pigs, it's possible that technical factors might underlie the failure to detect transcription at the Zygote stage.

Transcriptional features of totipotent cells

Shortly after fertilization, the mammalian genome will be transcriptionally activated for the first time in an event called **Zygotic Genome Activation** (ZGA) (Table 1). In mouse, this transcriptional event can be subdivided because of its timing into two parts: the **minor wave** of zygotic genome activation that occurs at the zygote stage (Mintz, 1964), and the larger **major wave** of zygotic genome activation that occurs later in the 2-cell stage (Woodland and Graham, 1969). While numerous developmentally-critical genes that are specific to the genome activation event are transcribed at this time and might underlie the totipotent capacity of the early embryo, it is most tempting to speculate on the contribution of the other major substrate of these transcriptional waves: transposable elements. Across both humans and mice, specific families of transposable elements are activated during genome activation (Fadloun et al., 2013b; Göke et al., 2015; Hendrickson et al., 2017; Peaston et al., 2004), in many cases enabling the transcription of chimeric genes that can only be transcribed downstream of the promoters of these repetitive elements. *During my PhD, I have characterized the transcriptional landscapes of pre-implantation development across five different mammalian species in an effort to assess the evolutionary conservation of the totipotent transcriptional program and of transposable element expression, among other questions. Although at the time of submission of this dissertation these analyses are still ongoing, the methodological advances that enabled the generation of these datasets as well as some preliminary findings are included further ahead.*

Transposable elements

A supplementary introductory section on transposable elements and their regulatory roles during mammalian pre-implantation development has been included in **Appendix I**. Copyright to include this publication in this dissertation has been retained and can be consulted in **Appendix III**.

The reference is:

Rodríguez-Terrones D. & Torres-Padilla M.E. (2018). Nimble and ready to mingle: transposon outbursts of early development. Trends in Genetics 10, 806-820.

***In vitro* model systems for the study of mammalian pluripotency and totipotency**

Mouse embryonic stem cells are an *in vitro* model system of pluripotency

Mouse **embryonic stem cells** (ESCs) are an *in vitro* model system for the study of the pluripotent cells from the mouse embryo. Although these cells are derived from the inner cell mass of the mouse pre-implantation blastocyst, *in vitro* they actually correspond to a heterogeneous mixture of at least two different developmental stages (Wray et al., 2010). Depending on culture conditions, these cells can transition back and forth between a state reminiscent of the pre-implantation epiblast – termed **naïve pluripotency** and defined by high expression of the transcription factor Rex1 – and another one reminiscent of the post-implantation epiblast – termed **primed pluripotency**, defined by low Rex1 expression and more prone to differentiate. As is evident from these heterogeneities, mouse ESCs recapitulate developmental features of their *in vivo* counterparts to a great extent and are even capable of resuming their developmental roles when transplanted back into the embryo.

In terms of their molecular identity, ESCs are characterized by the activity of a conserved gene regulatory network involving the transcription factors OCT4 (Nichols et al., 1998), SOX2 (Masui et al., 2007) and NANOG (Chambers et al., 2003), among several others (Kim et al., 2008). Indeed, this regulatory network not only modulates the pluripotent state of the cell but is even capable of inducing pluripotency in differentiated cell types (Takahashi and Yamanaka, 2006). With regards to epigenetic marks, naïve pluripotent stem cells are characterized by low levels of DNA methylation that become upregulated upon entry to the primed state, a situation that is also observed to occur in the mouse embryo (Leitch et al., 2013). Intriguingly, mouse ESCs also recapitulate several of the metabolic features observed in the early embryo, such as a mixed reliance on glycolysis and oxidative phosphorylation in the naïve state that shifts to an almost absolute use of glycolysis in the primed state (Zhou et al., 2012). Overall, in addition to reproducing the developmental characteristics present in the pluripotent cells of the mouse embryo, mouse embryonic stem cells also recapitulate many of the molecular features observed in their *in vivo* counterparts.

Mouse 2-cell-like cells are an *in vitro* model system of totipotency

In addition to the aforementioned heterogeneities, a small subpopulation of cells resembling the mouse 2-cell stage embryo spontaneously arises in mouse embryonic stem cell cultures and constitutes approximately 0.5% of the culture at any given time. This totipotent-like **2-cell-like** population of mouse ESCs was originally identified on the basis of its expression of the 2-cell-stage specific **MERV-L** family of retrotransposons (Macfarlan et al., 2012) and later found to exhibit a series of chromatin features in common with the 2-cell stage embryo such as higher histone mobility (Bošković et al., 2014). The ERV-L family of retrotransposons stands out for having been co-opted as promoters and other regulatory sequences for a wide range of genes activated at zygotic genome activation in both mouse and human, and its activation is suggestive of a reversion to a regulatory state reminiscent of the 2-cell stage embryo. In addition, given that these cells arise in pluripotent cultures, it is very surprising to note that these cells have been reported to downregulate protein levels of OCT4 (Macfarlan et al., 2012), calling into question whether they are even pluripotent.

Further adding to the list of mouse ESC heterogeneities, an altogether different subpopulation expressing the 2-cell stage specific gene *Zscan4* has been described to occur in approximately 5% of the cells in a mouse ESC culture (Falco et al., 2007; Zalzman et al., 2010). By periodically entering into this *Zscan4*⁺ state every ~10 passages, mouse ESCs elongate their telomeres through not completely resolved molecular mechanisms. ***What exactly is the relationship between the 2-cell-like subpopulation and *Zscan4*⁺ population is a question that I aim to address in the first part of this thesis. Beyond the actual cellular potency of these subpopulations, understanding how pluripotent cells rewind back to a molecular state reminiscent of the totipotent embryo could greatly aid us in understanding the molecular dynamics underlying totipotency in vivo. To this end, we also describe our search for molecular determinants of this transition and explore their mechanisms of action.***

Results

Part 1 – The molecular basis of totipotent-like cells in culture

Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly

Statement of contribution

I hereby state that my contribution to the publication:

Ishiuchi, T.*, Enriquez-Gasca, R., Mizutani, E., Boškoviä, A., Ziegler-Birling, C., **Rodriguez-Terrones, D.**, Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat. Struct. Mol. Biol. 22, 662–671.

consisted in the computational analysis of transcriptomic data that contributed to the determination of an enrichment of 2-cell-stage specific genes in the 2-cell-like cell transcriptome.



Diego Rodriguez-Terrones

München, December 12th, 2018

Confirmation of contribution

I hereby confirm that the statement of contribution reproduced above is both truthful and accurate.



Prof. Dr. Maria-Elena Torres-Padilla

München, December 12th, 2018

Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly

Takashi Ishiuchi¹, Rocio Enriquez-Gasca², Eiji Mizutani³, Ana Bošković¹, Celine Ziegler-Birling¹, Diego Rodriguez-Terrones¹, Teruhiko Wakayama³, Juan M Vaquerizas² & Maria-Elena Torres-Padilla¹

Cellular plasticity is essential for early embryonic cells. Unlike pluripotent cells, which form embryonic tissues, totipotent cells can generate a complete organism including embryonic and extraembryonic tissues. Cells resembling 2-cell-stage embryos (2C-like cells) arise at very low frequency in embryonic stem (ES) cell cultures. Although induced reprogramming to pluripotency is well established, totipotent cells remain poorly characterized, and whether reprogramming to totipotency is possible is unknown. We show that mouse 2C-like cells can be induced *in vitro* through downregulation of the chromatin-assembly activity of CAF-1. Endogenous retroviruses and genes specific to 2-cell embryos are the highest-upregulated genes upon CAF-1 knockdown. Emerging 2C-like cells exhibit molecular characteristics of 2-cell embryos and higher reprogrammability than ES cells upon nuclear transfer. Our results suggest that early embryonic-like cells can be induced by modulating chromatin assembly and that atypical histone deposition may trigger the emergence of totipotent cells.

The amazing capacity of the early embryo to generate all cells of an organism is transient. This feature, referred to as totipotency, implies that a full organism can be derived from a single cell and contrasts with pluripotency, which refers to the ability of a cell to contribute to all three germ layers of the embryo but not to the extraembryonic lineages^{1–4}. Thus, totipotent cells have greater cellular plasticity. In mice, only the zygote and 2-cell-stage blastomeres are fully totipotent because they can generate an organism on their own without the need of carrier cells³. Totipotent-like 2C-like cells have been shown to arise spontaneously in ES-cell cultures, but only in the extremely low proportion of around 0.5% (ref. 1). Similarly to 2-cell-stage embryos, 2C-like cells also reactivate transcription of endogenous retroviruses (ERVs), in particular MERVL. Although most somatic cells can be readily induced to pluripotency^{2,5}, it is unknown whether totipotent cells can be induced *in vitro*. Moreover, the molecular and epigenetic features of totipotent cells are poorly characterized. Here, we set out to identify molecular players associated with transitions between pluripotent and totipotent states. We show that depletion of either the p150 or the p60 subunits of chromatin assembly factor-1 (CAF-1) in ES cells increases the population of 2C-like-cells. CAF-1 depletion leads to increased accessibility at MERVL and to upregulation of neighboring genes, thus generating a transcriptional profile similar to that of 2-cell-stage embryos. Induced 2C-like cells after CAF-1 loss share the molecular characteristics of endogenous 2C-like cells and exhibit molecular features of 2-cell-stage embryos.

RESULTS

To obtain insights into the molecular mechanisms underlying the transition toward the 2-cell-like state, we sought to identify common

features between early embryos and 2C-like cells that differ from those of pluripotent cells. We generated stable ES-cell lines containing an EGFP reporter under control of the MERVL long terminal repeat (2C::EGFP) (Fig. 1a). Examination of several clones revealed that 2C-like cells, identified by EGFP expression and the absence of OCT4 protein¹, lack chromocenters (Fig. 1b), similarly to zygotes and 2-cell-stage embryos⁶, thus suggesting that, compared to pluripotent and somatic cells, they display major global differences in nuclear organization. Instead of the well-defined 4',6-diamidino-2-phenylindole (DAPI)-stained foci indicative of pericentromeric clustering into chromocenters in ES cells, the DAPI-rich regions in 2C-like cells appeared spread and were strongly enriched in acetylated histones, results suggesting accelerated histone turnover (Fig. 1b and Supplementary Fig. 1a). Absence of chromocenters was accompanied by a robust transcriptional activation of the major satellite (MajSat) repeats of the pericentromeric chromatin (Fig. 1c), a feature of zygotes and 2-cell-stage embryos^{7–9}. Thus, both early embryos and 2C-like cells lack chromocenters and reactivate transcription of major satellites and MERVL (Fig. 1d).

Depletion of CAF-1 induces 2C-like cells

To identify molecular pathways that regulate features of 2-cell-stage embryos and 2C-like cells, we searched for candidate proteins known to regulate chromocenter integrity and/or histone turnover. CAF-1 is a trimeric complex responsible for deposition of histones H3 and H4 during DNA synthesis^{10,11}, and depletion of the p150 subunit of CAF-1 results in chromocenter loss in pluripotent ES cells and in pericentric heterochromatin instability in *Drosophila*^{12,13}. We thus performed RNA interference (RNAi) for the two major CAF-1 subunits, p150

¹Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS UMR7104 and INSERM U964, Illkirch, France. ²Max Planck Institute for Molecular Biomedicine, Münster, Germany. ³Faculty of Life and Environmental Sciences, University of Yamanashi, Yamanashi, Japan. Correspondence should be addressed to M.-E.T.-P. (metp@igbmc.fr).

Received 9 June; accepted 9 July; published online 3 August 2015; doi:10.1038/nsmb.3066

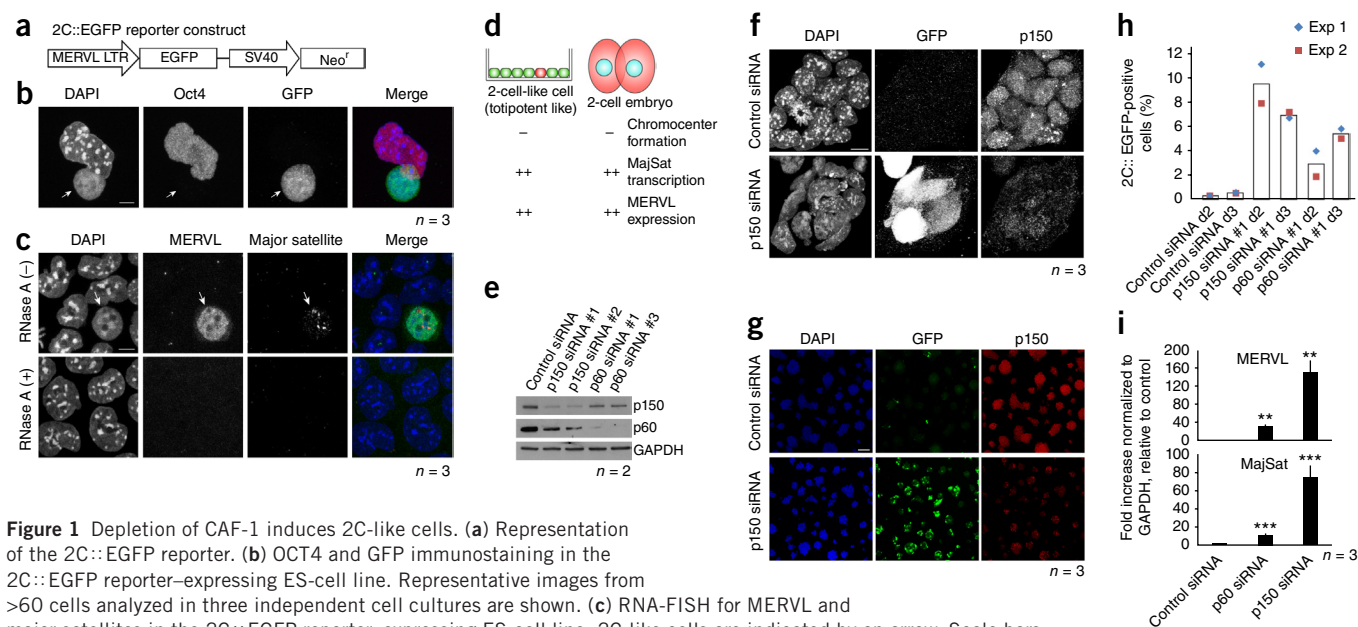


Figure 1 Depletion of CAF-1 induces 2C-like cells. (a) Representation of the 2C::EGFP reporter. (b) OCT4 and GFP immunostaining in the 2C::EGFP reporter-expressing ES-cell line. Representative images from >60 cells analyzed in three independent cell cultures are shown. (c) RNA-FISH for MERVL and major satellites in the 2C::EGFP reporter-expressing ES-cell line. 2C-like cells are indicated by an arrow. Scale bars, 5 μ m (b,c). Representative images from three independent cell cultures are shown. (d) Common features between 2C-like cells and 2-cell-stage embryos. (e) KD with different siRNAs for p150 and p60 and analysis of whole cell extracts with the indicated antibodies. Results are representative of two independent experiments. GAPDH is a loading control. (f) Loss of chromocenters and reactivation of the 2C::EGFP reporter in cells lacking p150 protein, after p150 KD. Representative images from three independent cell cultures are shown, with >50 cells analyzed for each experiment. Scale bar, 10 μ m. (g) Immunostaining of the 2C::EGFP line with antibodies to GFP and p150 after p150 KD. Scale bar, 100 μ m. *n*, number of independent experiments with independent cell cultures. (h) Quantification of GFP-positive cells by FACS 2 d and 3 d after transfection of control, p150 or p60 siRNA. Bars show the mean, and raw values from two independent cell cultures are also shown. (i) RT-qPCR for MERVL and MajSat in 2C::EGFP reporter-expressing ES-cell line after KD of p60 and p150. Shown are the mean \pm s.d. of two technical replicates for three cell cultures in experiments performed on different days. ** $P < 0.01$; *** $P < 0.001$ (two-tailed Student's *t* test, compared to control siRNA).

and p60, in ES cells containing the 2C::EGFP reporter. We confirmed depletion of CAF-1 subunits by western blotting and immunofluorescence (Fig. 1e,f). As previously reported¹², RNAi for p150 abolished detectable chromocenters (Fig. 1f) and led to a dramatic upregulation of EGFP-positive cells (Fig. 1g,h). Reverse-transcription quantitative PCR (RT-qPCR) and RNA fluorescence *in situ* hybridization (RNA-FISH) revealed that activation of the 2C::EGFP reporter reflected transcriptional reactivation of endogenous MERVL (Fig. 1i and Supplementary Fig. 1b). Moreover, p150 depletion resulted in strong upregulation of major-satellite transcription (Fig. 1i). We obtained similar results upon p60 knockdown (KD) (Supplementary Fig. 2). Remarkably, the overall number of cells expressing the 2C::EGFP reporter increased by 10- to 30-fold in the absence of p60 or p150 (Fig. 1g,h), thus suggesting that depletion of CAF-1 can dramatically induce the emergence of 2C-like cells.

Members of the ERVL family of retrotransposons act as alternative promoters of mouse genes and generate chimeric transcripts in 2-cell-stage embryos¹⁴ and in 2C-like cells¹. We therefore asked whether CAF-1 depletion leads to upregulation of 2-cell-specific genes, including chimeric transcripts. p150 depletion resulted in strong upregulation of the 2-cell-specific transcripts *Eif1a*-like and *Zscan4* as well as *Gm6763*, which generates a chimeric transcript linked to the MERVL long terminal repeat (LTR) (Fig. 2a and Supplementary Fig. 2). Although major satellites and MERVL transcripts were strongly upregulated, other active retrotransposons such as IAP and LINE1 were largely unaffected, thus suggesting that p150 depletion does not affect global transcription of all repetitive elements (Fig. 2a). Similarly to spontaneous 2C-like cells¹, EGFP-positive cells induced by p150 depletion were also devoid of OCT4 protein (Fig. 2b). Thus, we conclude that CAF-1 controls the

stability of stem-cell states and that the cells induced upon CAF-1 depletion are similar to 2C-like cells that emerge spontaneously.

Chromocenter disruption does not induce 2C-like cells

We next investigated the mechanism through which depletion of CAF-1 induces cells with totipotent-like features. Because the lack of chromocenters and reactivation of major-satellite transcription are commonly found in 2-cell embryos, we first asked whether conferring these features to ES cells would be sufficient to induce 2C-like cells, by using a transcription activator-like effector (TALE) that specifically binds major satellites¹⁵, fused to the VP64 transactivator (Fig. 2c). Targeting VP64 to major satellites in ES cells led to decondensation of chromocenters, as visualized with DAPI, and to accumulation of H4 K16 acetylation therein (Fig. 2d). Expression of TALE-MajSat-VP64 robustly increased major-satellite transcription (Fig. 2e). However, TALE-MajSat-VP64 expression did not result in increased MERVL transcription or in loss of OCT4 protein (Fig. 2e,f), and the number of ES cells expressing the 2C::EGFP reporter remained unchanged (Supplementary Fig. 2). Thus, although chromocenters and major-satellite transcription appear to be characteristic features of totipotent cells, loss of chromocenters and reactivation of major satellites are not sufficient to induce 2C-like cells.

Second, we asked whether the passage through replication was necessary for induction of the 2C::EGFP reporter upon p150 depletion. After confirming that thymidine treatment efficiently arrested ES cells at G1-S phase (Supplementary Fig. 3), we performed RNAi for p150, blocked cells at G1-S and analyzed ES cells by immunostaining and RT-qPCR (Fig. 3a). Thymidine treatment suppressed the phenotype caused by p150 depletion, thus leading to a strong reduction in the number of cells expressing the 2C::EGFP reporter as compared to the

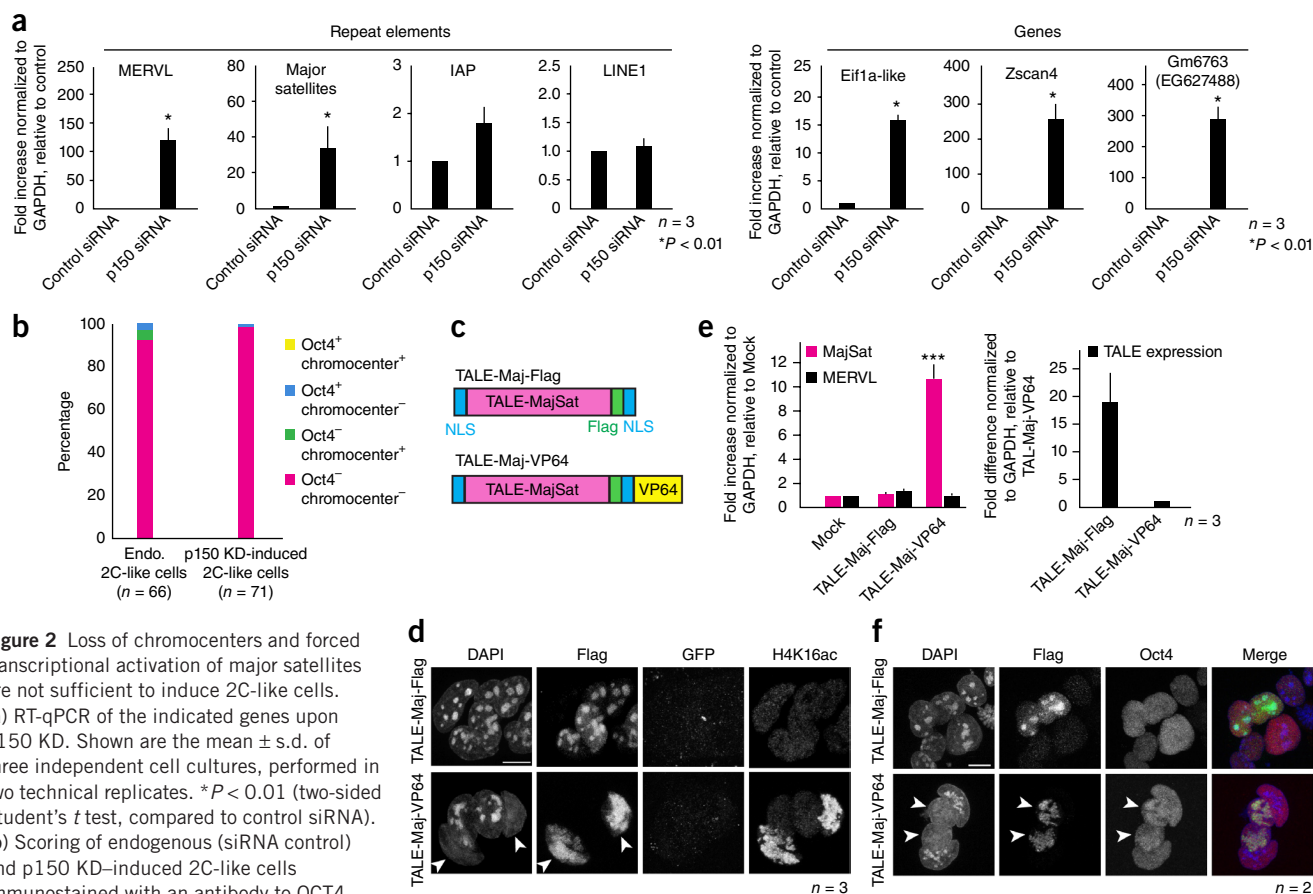


Figure 2 Loss of chromocenters and forced transcriptional activation of major satellites are not sufficient to induce 2C-like cells. (a) RT-qPCR of the indicated genes upon p150 KD. Shown are the mean \pm s.d. of three independent cell cultures, performed in two technical replicates. $*P < 0.01$ (two-sided Student's *t* test, compared to control siRNA). (b) Scoring of endogenous (siRNA control) and p150 KD-induced 2C-like cells immunostained with an antibody to OCT4, according to the presence or absence of OCT4 and chromocenters. *n*, number of cells analyzed for each combination. (c) Schematic representation of the TALE constructs used in d. NLS, nuclear localization signal. (d) Immunostaining of 2C::EGFP-reporter ES cells for Flag, GFP and acetylated H4 K16 after transfection of two different TALE constructs as indicated. Cells expressing TALE-Maj-VP64 are indicated by arrowheads. Scale bar, 10 μ m. Shown are representative images of >20 cells analyzed in each of three independent cultures. (e) RT-qPCR analysis for major satellites (MajSat) and MERVL after transfection of two different TALE constructs (left) and for expression levels for each of the constructs (right). Data are mean \pm s.d. of three independent cell cultures. $***P < 0.001$ (two-sided Student's *t* test, compared to mock transfection (mock)). (f) Flag and OCT4 immunostaining of cells transfected as in d. Cells expressing TALE-Maj-VP64 are indicated by arrowheads. Scale bar, 10 μ m. Representative images of >20 cells analyzed in each of two independent cultures are shown.

control (Fig. 3b). G1-S arrest also rescued the loss of chromocenters, which were instead readily detectable by DAPI staining in spite of the absence of p150 (Fig. 3c). Finally, induction of MERVL and major-satellite transcription was also diminished in p150-KD ES cells after treatment with thymidine, whereas the abundance of LINE1 and IAP transcripts was unaffected (Fig. 3d). Thus, inhibition of cell-cycle progression through S phase suppresses induction of 2C-like cells after p150 depletion, a result suggesting that induction occurs concomitantly with or after replication.

Loss of chromatin-assembly activity induces 2C-like cells

We performed a complementation assay to determine how p150 is involved in suppressing the emergence of 2C-like cells. To discern whether the phenotype induced by p150 depletion reflects a deficiency in DNA synthesis and/or in chromatin assembly, we focused on four domains within p150: two noncanonical PIP domains that interact with PCNA, the HP1-interaction motif and the ED domain (Fig. 3e). The N-terminal PIP1 is dispensable for chromatin assembly and does not contribute significantly to p150 targeting to replication sites, whereas ablation of PIP2 impairs chromatin assembly without inhibiting DNA synthesis¹⁶. Likewise, the p150 HP1-interaction domain is required for CAF-1 localization to heterochromatin outside of S phase

but is dispensable for *de novo* chromatin assembly and for CAF-1 targeting to replication foci¹⁷. In contrast, the ED domain, composed of clusters of glutamate and aspartate residues, is essential for assembly activity¹⁸. We generated ES cells expressing either full-length p150 or each of four Flag-tagged p150 mutants lacking these domains, and we confirmed that all five constructs were homogeneously and efficiently expressed, although the p150 Δ PIP1 and Δ ED deletions displayed slightly lower protein levels (Fig. 3f and Supplementary Fig. 3). We then depleted endogenous p150 with a short interfering RNA (siRNA) (p150 siRNA #1) that does not target the transgenes used for complementation (Fig. 3e). Expression of exogenous full-length p150 and of the Δ PIP1 and Δ HPI1 mutants suppressed the increase of 2C-like cells induced by depletion of endogenous p150 (Supplementary Fig. 3). RT-qPCR revealed that expression of the Δ PIP1 and Δ HPI1 mutants, compared to the full-length p150, led to similar transcript levels for MERVL, Zscan4, Gm6763 and major satellites (Fig. 3g). Likewise, chromocenters were readily visible after expression of exogenous full-length p150 or the Δ PIP1 and Δ HPI1 mutants (Fig. 3h). Strikingly, however, the Δ ED and the Δ PIP2 constructs, which each lack domains required for the chromatin-assembly activity of CAF1, were unable to rescue p150 deficiency, and cells instead exhibited significant expression of 2-cell-specific genes and repetitive elements analyzed

and failed to maintain chromocenter integrity (Fig. 3g,h). Moreover, whereas expression of the full length, Δ PIP1 or Δ HP1 constructs suppressed the emergence of 2C-like cells, the Δ PIP2 or the Δ ED mutants failed to do so (Supplementary Fig. 3). Because RNAi for p60 in human cells leads to loss of chromatin assembly during DNA synthesis¹⁹, these data are in agreement with the observation that depletion of p60 alone also increases the number of cells expressing the 2C::EGFP reporter. Importantly, a second siRNA (p150 siRNA #2) targeting both endogenous p150 and the exogenously expressed p150 mutants recapitulated the emergence of 2C-like cells

(Supplementary Fig. 3). Thus, CAF-1 can regulate the stem-cell state through its chromatin-assembly activity, but p150's functional interaction with HP1 and with PCNA through the PIP1 is dispensable. Moreover, the data point to a role for the chromatin-assembly activity of CAF-1 in orchestrating the emergence of 2C-like cells. Importantly, ablation of another histone-deposition pathway through downregulation of the HIRA chaperone did not affect chromocenter integrity or 2C::EGFP expression (Supplementary Fig. 4), thus suggesting that the induction of totipotent-like features upon CAF-1 KD is not a result of altered histone-deposition pathways in general.

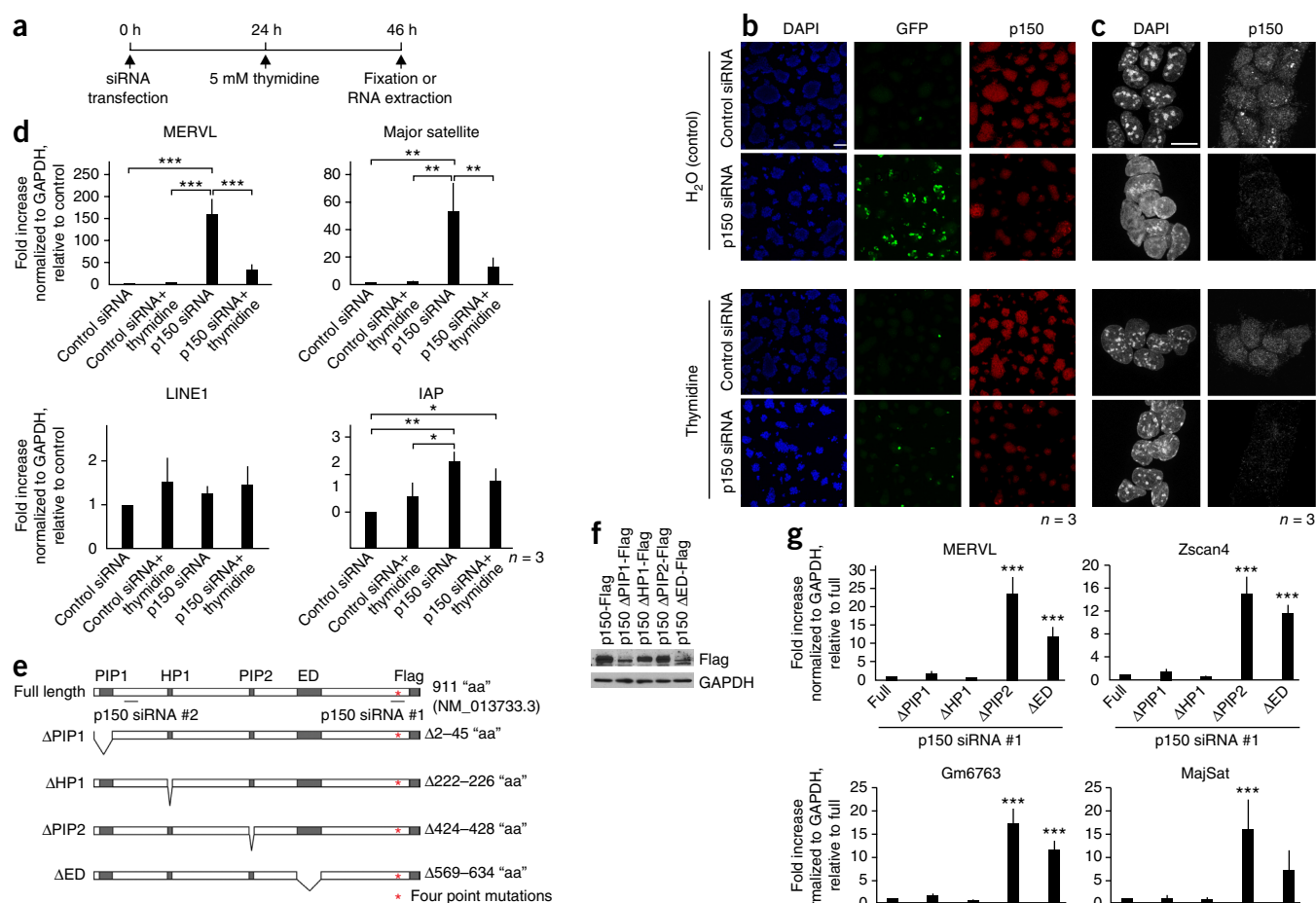


Figure 3 Induction of 2C-like cells is triggered by the loss of the chromatin assembly activity of CAF-1. **(a)** Schematic representation of the experimental procedure used for **b–d**. **(b,c)** Immunostaining for GFP and p150 of 2C::EGFP reporter ES cells transfected with p150 siRNA and treated with either vehicle (water) or thymidine. The population of GFP⁺ cells is indicated in **b**, and the chromocenters are visualized by DAPI in **c**. Scale bars, 10 μ m (**b**) and 100 μ m (**c**). Representative images from three independent cell cultures are shown. **(d)** RT-qPCR analysis for MERVL, MajSat, LINE1 and IAP in the four different conditions used in **b** and **c**. Data are mean \pm s.d. of three cell cultures in experiments performed on different days in technical duplicates. * P < 0.05; ** P < 0.01; *** P < 0.001 (Tukey's honest significant difference (HSD) test). **(e)** Schematic representation of full-length p150 with the functional domains and the mutants used in the complementation assay indicated. Red asterisks indicate the position of four point mutations at the p150 siRNA #1 target sequence, which render the exogenous transgene resistant to p150 siRNA #1 but not to p150 siRNA #2. **(f)** Expression level of p150 full-length or mutants analyzed by western blotting. GAPDH, loading control. **(g,h)** Analysis by RT-qPCR for MERVL, Zscan4, Gm6763 and MajSat (**g**) or by immunostaining for Flag (**h**). Samples are cells stably expressing p150 full-length or mutants and transfected with p150 siRNA #1 to deplete endogenous p150. Scale bar, 10 μ m. Representative data of three independent cell cultures are shown. *** P < 0.001 (Tukey's HSD test). In **h**, at least 50 cells were analyzed per replicate for each condition.

Transcriptome upon CAF-1 depletion resembles 2-cell stage

Next, we asked whether CAF-1 depletion results in a global transcriptional profile similar to that of 2-cell-stage embryos, paying particular attention to retrotransposon activation. We performed RNA sequencing (RNA-seq) in ES cells after KD of either of the two CAF-1 subunits, p150 and p60. Analysis of differentially expressed transcripts with application of a false discovery rate (FDR)-adjusted P -value cutoff revealed 2,517 and 1,676 upregulated genes with more than two-fold change in gene expression after treatment with p150 or p60 siRNA, respectively, and only 96 and 31 downregulated genes (Fig. 4a and Supplementary Tables 1 and 2). The greater number of upregulated than downregulated genes suggests a general openness of the chromatin upon CAF-1 depletion, in agreement with a hypothesis of general repressor function for CAF-1 (ref. 13). Comparison of differentially expressed genes in p150- and p60-depleted ES cells revealed a high similarity between the two data sets (Supplementary Fig. 5), results in agreement with the observations that p150 and p60 are interdependent in mediating chromatin assembly. Among the upregulated genes upon CAF-1 KD, we identified Eif1a-like genes (which include *Gm5662*, *Gm2022*, *Gm4027*, *BB287469*, *Gm2016*, *Gm21319*, *Gm8300* and *Gm10264*), *Zscan4* genes (*Zscan4b*–*Zscan4f*), *Zfp352* and *Tdpoz* genes (*Tdpoz1*–*Tdpoz5*) expressed in 2-cell embryos²⁰. Pluripotency-associated genes such as *Nanog* and *Pou5f1* (*Oct4*) were unaffected (Fig. 4a and Supplementary Tables 1 and 2), in line with the observation that endogenous 2C-like cells maintain pluripotency-associated gene transcription but have downregulated protein levels¹. We also performed a comprehensive analysis of the expression of repetitive elements by aligning reads to the genome, applying as annotation the coordinates of repeat elements obtained with RepeatMasker (v4.0.3; <http://www.repeatmasker.org/>). Remarkably, only a subset of retrotransposons were highly upregulated upon treatment with p150 or p60 siRNA compared to control siRNA (Fig. 4b and Supplementary Tables 3 and 4). In particular, *gsat_mm* (major satellite repeat) and the MERVL families of retrotransposons *mt2_mm*, *mervl-int* and *mt2b1*—which are specifically activated in 2-cell embryos—were the highest-upregulated elements after either p60 or p150 KD (Fig. 4b and Supplementary Tables 3 and 4). Consistently with the RT-qPCR results, the upregulation of repetitive elements upon CAF-1 depletion was not global but displayed specificity: LTR retrotransposons of the MERVL family were very highly upregulated after KD of either p150 or p60, but IAPez, LINE1 and other types of active retrotransposons were unaffected (Fig. 4b).

To establish the extent of similarity between induced 2C-like cells and 2-cell-stage embryos, we compared the list of differentially expressed genes with those extracted from a comparison between oocyte and 2-cell embryos and therefore deemed to be specifically activated in 2-cell-stage embryos¹. A statistically significant subset of genes were commonly upregulated (Supplementary Fig. 5; $P = 5.93 \times 10^{-14}$ and 1.56×10^{-14} for p150 and p60 RNAi, respectively, by two-sided Fisher's exact test), thus suggesting that 2-cell embryos and CAF-1-KD cells share overlapping gene-expression patterns. To further assess the identity of 2C-like cells induced upon CAF-1 KD, we performed a systematic comparison of the expression profiles of endogenous 2C-like cells, p60 KD-induced 2C-like cells and p150 KD-induced 2C-like cells against all stages of preimplantation development. We first generated RNA-seq data derived from the GFP-positive sorted population of cells for each condition (Supplementary Tables 5 and 6), which further revealed that induced 2C-like cells upon CAF-1 KD have highly similar transcriptional profiles to endogenous 2C-like cells ($P < 2.2 \times 10^{-16}$ and $P < 2.2 \times 10^{-16}$, by FDR-adjusted two-sided Fisher's exact test, for KD of p150 and p60, respectively) (Online Methods). We then extracted the genes specifically expressed

in zygote, 2-cell, 4-cell, 8-cell and 16-cell stages, and early blastocyst from ref. 21 (Online Methods) and compared them with the sets of genes upregulated in endogenous 2C-like, p60 KD-induced 2C-like and p150 KD-induced 2C-like cells. This analysis revealed that the gene subset specific to 2-cell-stage embryos is overrepresented in the set of upregulated genes in 2C-like cells induced after KD of p60 or p150, in comparison with any of the other stages analyzed and similarly to endogenous 2C-like cells ($P = 4.44 \times 10^{-15}$ and 2.09×10^{-13} for KD of p150 and p60, respectively, and $P = 2.90 \times 10^{-16}$ for endogenous 2C-like, by Bonferroni-adjusted two-sided Fisher's exact test) (Supplementary Fig. 6). This implies that a substantial proportion of the transcriptional network specific to 2-cell-stage embryos is activated after CAF-1 depletion in ES cells. Chimeric transcripts such as *Esp24*, *Gm8994*, *Gm7102* and *Gm12114*, which derive from MERVLs, are overrepresented within the commonly upregulated genes in 2-cell-stage embryos¹ and in p150- or p60-depleted cells (Supplementary Tables 7 and 8). Moreover, we observed readthrough transcription from endogenous retrovirus into genes located in proximity to MERVL-related elements after KD of p150 and p60 (data not shown). These observations prompted us to ask whether CAF-1 regulates expression of genes associated with or located close to retrotransposons, thereby generating a transcriptional profile similar to that of early embryos¹⁴. We determined the percentage of upregulated genes out of the total number of genes located within a fixed distance of a repeat (Supplementary Fig. 5). Strikingly, this analysis revealed that a large proportion of genes proximal to the MERVL LTR (*mt2_mm*, *mt2b* and *mt2c*) were upregulated upon CAF-1 KD (Fig. 4c and Supplementary Fig. 7). Interestingly, we did not find significant proximal distance association between upregulated genes and repeat elements unrelated to the MERVL family, such as LINE1 (Fig. 4c). These results suggest that the overlap of upregulated genes in both 2-cell embryo and CAF-1-depleted ES cells can be largely ascribed to the activation of genes proximal to retrotransposon elements normally activated in 2-cell-stage embryos, in particular the MERVL family of retrotransposons. Indeed, a global distance analysis indicated that genes that were commonly upregulated in CAF-1 depleted cells and expressed in 2-cell-stage embryos are significantly closer to MERVL family of retrotransposons (*mt2_mm*, *mt2b1* or *mt2c_mm*) compared to nonoverlapping genes and are also significantly closer when compared to the rest of the genes in the genome (Fig. 4d; $P < 2 \times 10^{-16}$ for both p150 and p60, by two-sided Wilcoxon rank-sum test). Altogether, these results imply that CAF-1 might regulate a subset of 2-cell stage-specific genes that are dependent on the presence of retrotransposon elements.

To address how CAF-1 depletion mediates changes in retrotransposon expression in these induced 2C-like cells, we asked whether CAF-1 modulates chromatin accessibility as it does in HeLa cells¹⁹ by using micrococcal nuclease (MNase) digestion. We isolated nuclei from control siRNA-treated ES cells and from fluorescence-activated cell sorting (FACS)-sorted GFP-positive 2C-like cells after p60 siRNA transfection and performed a time course of MNase digestion. Analysis of the digested DNA revealed a higher global accessibility to MNase in EGFP-positive 2C-like cells induced after p60 depletion (Supplementary Fig. 7). To specifically address whether MERVL elements are in a more open chromatin configuration in these cells, we performed Southern blot hybridization after MNase digestion, using a MERVL LTR probe, which revealed that MERVL LTRs are indeed more accessible in EGFP-positive 2C-like cells induced after p60 depletion compared to control ES cells (Fig. 4e). These results suggest that CAF-1 knockdown promotes expression of MERVL through changes in chromatin accessibility.

Chromatin features similar to those of 2-cell embryos upon CAF-1 loss

We recently reported that high chromatin mobility is a hallmark of totipotent cells and that 2-cell-stage nuclei display unusually high core-histone mobility²². To address whether CAF-1 KD-induced 2C-like cells share similar chromatin features with 2-cell-stage embryos, we asked whether CAF-1 depletion results in changes in histone mobility comparable to those observed in both 2-cell-stage

embryos and endogenous 2C-like cells. We performed fluorescence recovery after photobleaching (FRAP) analyses of H3.1-GFP in ES cells containing a MERVL LTR-driven tdTomato reporter (2C::tdTomato)²² after p60 and p150 KD. Remarkably, tdTomato-positive cells induced upon p60 and p150 KD displayed a very high H3.1-GFP mobility (Fig. 5a), and the H3.1-GFP mobile fractions were highly similar to those in 2-cell-stage embryos and in spontaneously emerging 2C-like cells²² (Fig. 5b). We conclude that global core-histone

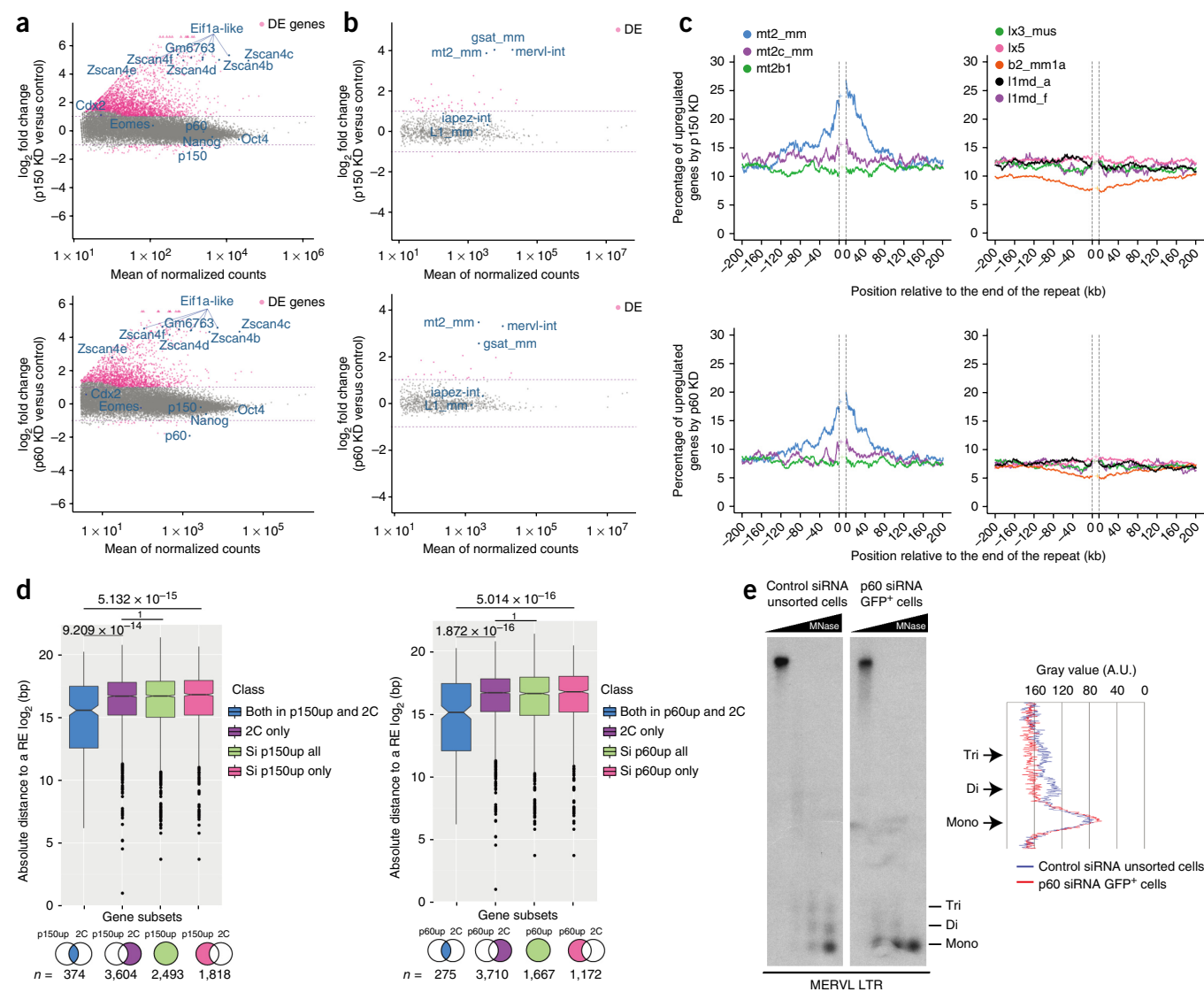


Figure 4 CAF-1 depletion induces expression of genes located in proximity to repetitive elements. **(a)** MA plots displaying differentially expressed genes in p150- and p60-depleted ES cells. $n = 2$ cell cultures in experiments performed on different days (Spearman's $\rho \geq 0.97$ for each pair of replicates; **Supplementary Note**). Arrows represent data points outside of the plotting area. DE, differentially expressed. **(b)** MA plots showing differentially expressed repeats in p150- and p60-depleted ES cells. $n = 2$ cell cultures in experiments performed on different days (Spearman's $\rho \geq 0.99$ for each pair of replicates; **Supplementary Note**). **(c)** Proportion of differentially expressed genes in p150- or p60-KD samples located in the vicinity of specific repeat types. Colored lines represent the percentage of upregulated genes among all overlapping genes up to a distance of 200 kb surrounding each repeat type. At left: blue, mt2_mm; green, mt2b1; purple, mt2c_mm; pink, lx3_mus; orange, lx5; black, b2_mm1a; purple, l1md_a; purple, l1md_f. **(d)** Distribution of distances (\log_2) between differentially expressed genes upon p150 or p60 KD and/or genes upregulated in 2-cell embryos to the 2-cell-specific repeats. The absolute distance between the transcription start site of a gene and the end of the closest 2-cell-specific MERVL family of repeats (either mt2_mm, mt2b1 or mt2c_mm) was measured for each gene. Box plots indicate the median \log_2 base-pair distance distribution, and notches indicate confidence interval. The box covers the interquartile range. Whiskers extend to 1.5 times the interquartile range, and the center lines indicate medians. Dots indicate outliers. P values were calculated by two-sided Wilcoxon rank-sum tests with Bonferroni correction for multiple testing. RE, repetitive element. **(e)** MNase analysis with a MERVL probe after p60 KD. Images are representative of two Southern blots from two independent experiments. The positions of the DNA corresponding to mono-, di- and trinucleosomes are indicated. Quantification of the signal intensity over the region indicated is shown on the right and was performed with ImageJ/PlotProfiler. A.U., arbitrary units.

mobility in CAF-1 depletion–induced 2C-like cells is similar to that in 2-cell-stage embryos and that induced 2C-like cells share molecular features of totipotent cells.

Induced 2C-like cells have higher reprogrammability on SCNT

The above data indicate that induced 2C-like cells after CAF-1 depletion display a transcriptional program and the chromatin features of 2-cell-stage embryos. Cells can be tested for their bipotential capacity to give rise to inner cell mass and trophectoderm derivatives through physical aggregation with morula-stage embryos or through injection into blastocysts. However, we reasoned that because 2C-like cells correspond to a different, much earlier, developmental timing than that of the recipient embryos used to test bipotentiality, a more appropriate approach should be used to test functional relevance. Therefore, to address the impact on embryonic development of the chromatin and transcriptional features of 2C-like cells induced upon CAF-1 KD, we performed somatic-cell nuclear transfer (SCNT) into enucleated mouse oocytes. We reasoned that if nuclei of 2C-like cells regain totipotent features of 2-cell-stage embryos, their success rate upon SCNT should be greatly improved compared to ES cells because early, totipotent embryos are substantially better donors for SCNT²³. Indeed, the capacity to reprogram the donor nucleus also depends on the potency state of the donor nucleus itself²³. For donors, we used FACS-sorted GFP-negative cells, GFP-positive endogenous 2C-like cells and GFP-positive 2C-like cells induced after p60 KD (Fig. 5c). We also controlled for the effects of FACS on SCNT success rate (Supplementary Fig. 7). The number of embryos that cleaved to the 2-cell stage was greatly increased when 2C-like (GFP-positive) cells were used as donors compared to ES cells (GFP⁻ cells) (68% compared to 28%, $n = 210$ and 200, respectively). Importantly, p60-KD GFP-positive cells showed a similar 2-cell-stage cleavage rate to endogenous 2C-like cells (61%, $n = 197$) (Fig. 5d and Supplementary

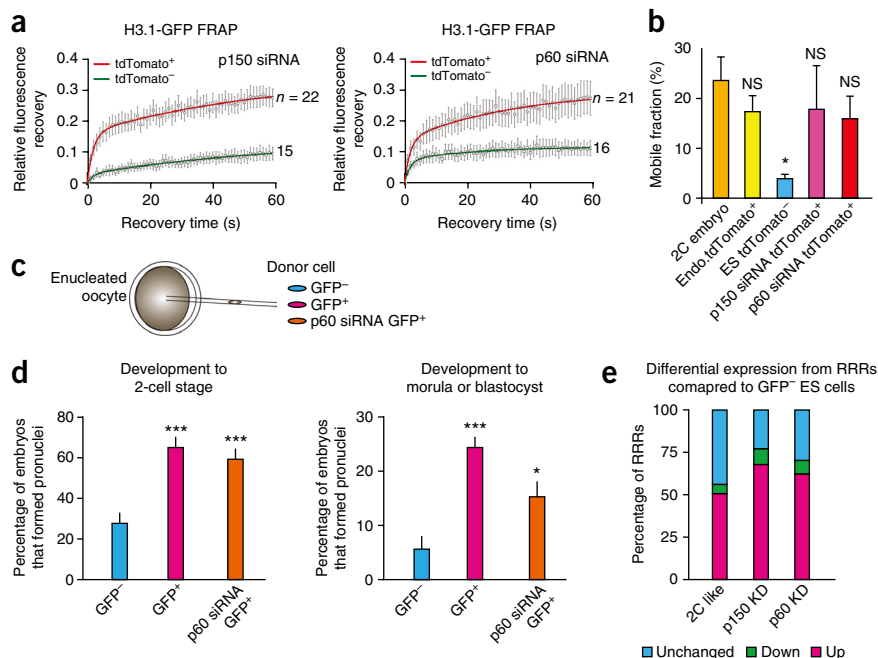
Fig. 7). The rate of development to the morula or blastocyst stage for SCNT embryos derived from endogenous 2C-like and p60 KD–induced 2C-like cells was 4- and 2.5-fold higher, respectively, than those derived from ES cells, thus indicating that endogenous 2C-like cells as well as 2C-like cells induced upon p60 KD have a greater reprogrammability than ES cells (Fig. 5d). These results prompted us to investigate whether reprogramming-resistant regions (RRRs) normally transcribed in 2-cell-embryos²⁴ are differentially expressed in 2C-like cells compared to ES cells. RRRs are transcribed in normal 2-cell and successful SCNT embryos but not in SCNT embryos that fail to undergo complete reprogramming²⁴. Analysis of RRRs in endogenous 2C-like, p60-KD and p150-KD induced 2C-like cells revealed that a large proportion of RRRs are significantly upregulated in endogenous and induced 2C-like cells compared to ES cells (Fig. 5e; $P = 1 \times 10^{-4}$ by one-sided permutation test, $n = 10,000$ permutations), thus indicating that high transcriptional activity at RRRs is common between 2C-like cells and 2-cell embryos. These data indicate that, similarly to 2-cell-stage embryos, endogenous and induced 2C-like cells have higher reprogrammability than ES cells, suggesting a broader developmental capacity.

p60 levels anticorrelate with MERVL expression

Finally, we asked whether changes in CAF-1 levels correlate with spontaneously arising endogenous 2C-like cells. For this, we used a 2C reporter that drives the expression of a destabilized turboGFP (2C::3XturboGFP-PEST) that reflects the dynamics of MERVL activation and repression, and therefore the emergence of 2C-like cells. Immunostaining with antibody to p60 revealed that spontaneously arising 2C-like cells tend to display virtually undetectable levels of p60 (Fig. 6a,b), thus suggesting that endogenous levels of p60 might fluctuate in a small proportion of ES cells largely corresponding to the 2C-like cells. To address whether this is also the case *in vivo*, we

Figure 5 CAF-1 depletion generates cells with similar chromatin features to 2-cell-stage embryos and higher reprogrammability upon SCNT.

(a) FRAP recovery curves of H3.1-GFP in tdTomato-positive and tdTomato-negative cells after KD of p60 or p150. Recovery was quantified in the bleached area over a 60-s period, and the curves were normalized to zero to account for differences in bleach depth between experiments. Individual points are mean \pm s.e.m., and mean values were fit into an exponential curve. The number of nuclei analyzed for each condition is indicated. (b) Estimated mobile fractions (\pm s.e.m.) of H3.1-GFP in induced 2C-like cells upon CAF-1 KD (p150 KD and p60 KD) compared to H3.1-GFP mobile fraction in spontaneously arising 2C-like cells (Endo.tdTomato⁺), 2-cell-stage embryos and ES cells (ES tdTomato⁻). NS, not significant; * $P < 0.05$ by two-sided unpaired *t* test. n , number of nuclei analyzed for each condition, derived from experiments performed on at least three different days: $n = 20$ for 2C and Endo.tdTomato⁺; $n = 16$ for ES tdTomato⁻ and p60 tdTomato⁺; $n = 22$ for p150tdTomato⁺. (c) Schematic representation of the nuclear transfer (NT) experiments performed in this study. (d) Development of NT embryos to 2-cell stage and to morula or blastocyst. The percentage of NT embryos developed to 2-cell (left) and morula or blastocyst (right) is shown. * $P < 0.05$, *** $P < 0.001$, compared to GFP⁻ cells (two-sided Student's *t* test) (raw data in Supplementary Fig. 7). Data are mean \pm s.e.m. from independent nuclear transfer experiments performed four times on different days. (e) Differential transcription from RRRs in CAF-1 KD–induced and endogenous 2C-like cells. RNA-seq data from sip150 (sorted GFP⁺), sip60 (sorted GFP⁺) and endogenous 2C-like cells (sorted GFP⁺) were compared to that from GFP⁻ ES cells. The percentage of RRRs with upregulated (up), downregulated (down) and unchanged transcription is shown. $P = 1 \times 10^{-4}$ by one-sided permutation test; $n = 10,000$ permutations.



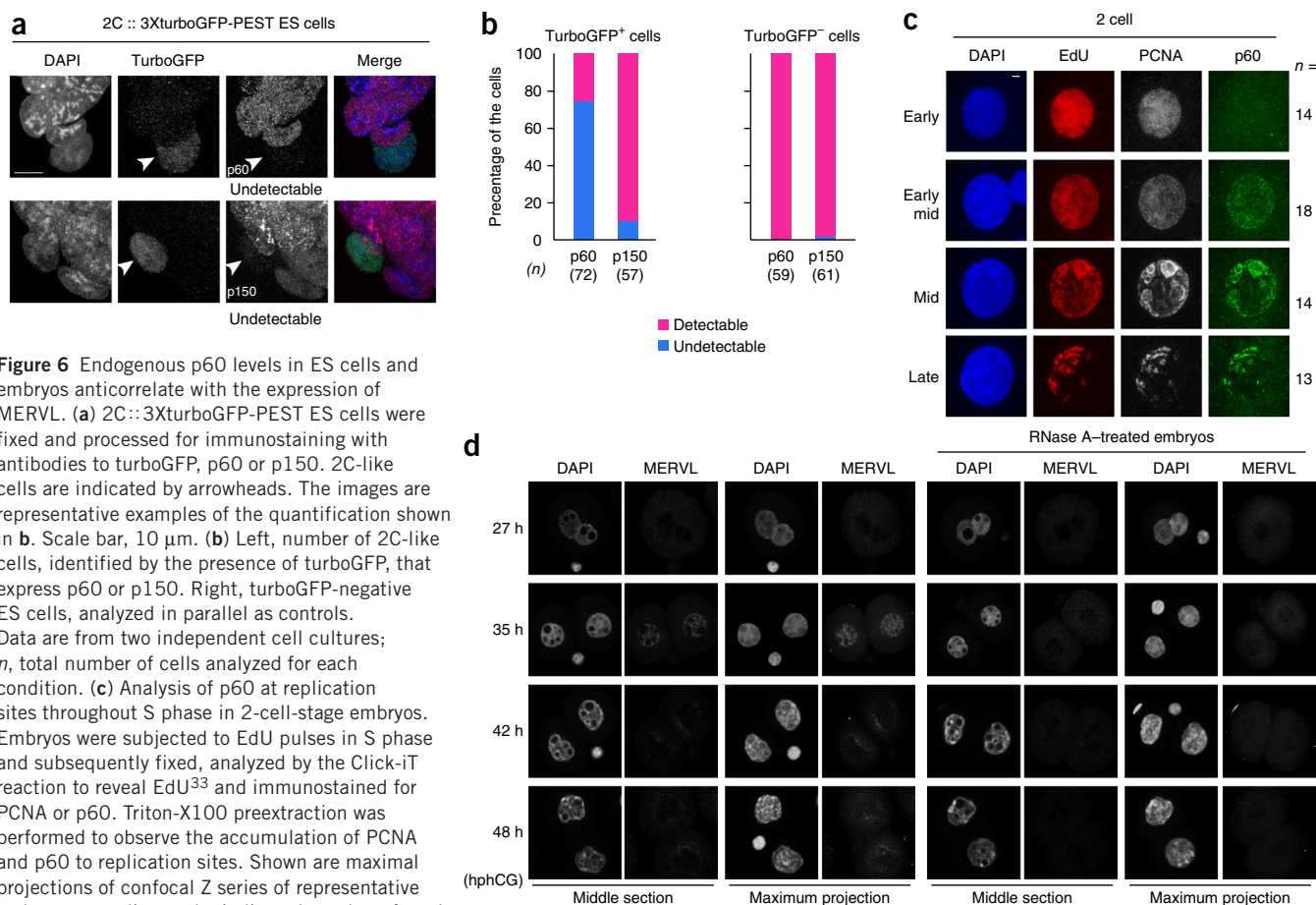


Figure 6 Endogenous p60 levels in ES cells and embryos anticorrelate with the expression of MERVL. **(a)** 2C::3XturboGFP-PEST ES cells were fixed and processed for immunostaining with antibodies to turboGFP, p60 or p150. 2C-like cells are indicated by arrowheads. The images are representative examples of the quantification shown in **b**. Scale bar, 10 μ m. **(b)** Left, number of 2C-like cells, identified by the presence of turboGFP, that express p60 or p150. Right, turboGFP-negative ES cells, analyzed in parallel as controls. Data are from two independent cell cultures; *n*, total number of cells analyzed for each condition. **(c)** Analysis of p60 at replication sites throughout S phase in 2-cell-stage embryos. Embryos were subjected to EdU pulses in S phase and subsequently fixed, analyzed by the Click-iT reaction to reveal EdU³³ and immunostained for PCNA or p60. Triton-X100 preextraction was performed to observe the accumulation of PCNA and p60 to replication sites. Shown are maximal projections of confocal Z series of representative embryos according to the indicated number of total embryos analyzed (*n*). Scale bar, 2 μ m. **(d)** RNA-FISH analysis using a MERVL probe in freshly collected zygotes (27 h), early 2-cell-stage (35 h), mid 2-cell-stage (42 h) and late 2-cell-stage embryos (48 h) showing transcriptional activation of MERVL exclusively in early 2-cell-stage embryos. Right, negative control embryos treated with RNase A. Representatives of at least 15 embryos analyzed per condition are shown.

analyzed p60 in 2-cell-stage embryos at replication sites throughout S phase, using 5-ethynyl-2'-deoxyuridine (EdU) and subsequent immunostaining for PCNA and p60. Remarkably, we found that p60 was undetectable on embryonic chromatin during a short time window in early S phase, in spite of robust EdU and PCNA staining (**Fig. 6c**), thus suggesting a transient uncoupling between DNA synthesis and chromatin assembly in the early embryo. Importantly, RNA-FISH with the MERVL probe revealed that this time window in the early 2-cell-stage embryo corresponds to the time when MERVL elements are transcriptionally activated (**Fig. 6d**).

DISCUSSION

The molecular features of totipotent cells include absence of OCT4, robust reactivation of retrotransposons and major satellite repeats, lack of chromocenters and higher chromatin mobility. The chromatin of totipotent cells thus has very distinctive features from that of pluripotent cells. Molecularly and developmentally, totipotent cells must be regarded as being in a unique cellular state that supports a high level of cellular plasticity. Our results indicate that the modulation of histone deposition, and not a DNA-synthesis defect, is the basis for the emergence of 2C-like cells induced after CAF-1 depletion. Importantly, our results showing that a large proportion of 2C-like cells have undetectable levels of p60 suggest that changes in chromatin assembly are indeed a key mechanism

underlying the emergence of 2C-like cells. CAF-1 performs the first step of the chromatin-assembly process, which consists in bringing H3 and H4 to the daughter DNA strands²⁵. Rapid nucleosome assembly is retarded in the absence of functional CAF-1 (ref. 26). We propose that less efficient and/or delayed chromatin assembly generates a chromatin state that promotes totipotency (**Fig. 7**). This is in agreement with our observations documenting a lack of detectable p60 on embryonic chromatin during early S phase in early 2-cell-stage embryos, coincident with the time of transcriptional activation of MERVL elements.

It remains to be addressed whether the bipotentiality to contribute to inner cell mass and trophoblast documented for endogenous 2C-like cells¹ is a genuine indicator as an assay for totipotency and whether chimera experiments would show a higher bipotentiality for induced 2C-like cells than for ES cells. Indeed, it should be noted that ES cells also show a degree of contribution to the trophoblast when aggregated in morulas or injected into blastocysts^{27–29}. In the two cases in which bipotentiality has been addressed conclusively with markers of trophoblast derivatives in the post-implantation placenta^{28,30}, bipotential cells, unlike 2C-like cells, expressed pluripotency markers and a number of genes expressed at the morula stage. Thus, in these cases, bipotential cells share transcriptional features with morulas and therefore represent a different developmental time frame than 2-cell-stage embryos.

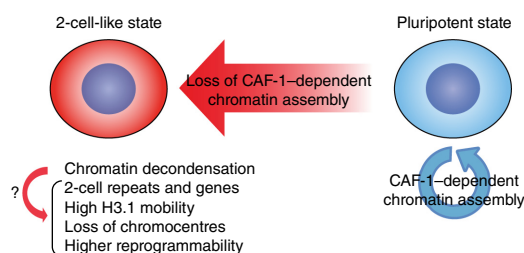


Figure 7 Working model for the induction of totipotent-like features through the modulation of chromatin assembly. The effects of downregulation of CAF-1 in ES cells generates cells with molecular features that are found *in vivo* in totipotent early 2-cell-stage embryos and *in vitro* in 2C-like cells but are absent in pluripotent cells. These features include (i) a highly mobile chromatin, (ii) lack of chromocenters, (iii) activation of pericentromeric heterochromatin, (iv) reduced levels of pluripotency factors such as OCT4 and (v) activation of MERVL.

We argue that SCNT provides an alternative and complementary approach to test for the plasticity status of the donor nucleus. Indeed, early embryos are much better donors than ES cells^{23,31} and can therefore be used to test for differences in the plasticity status of donor nuclei. Most importantly, enucleated zygotes efficiently develop to the blastocyst stage only when an early 2-cell donor nucleus is transferred but very rarely if the donor nucleus is derived from a late 2-cell, early 4-cell or mid 8-cell embryo³². Interestingly, there is a drastic difference in reprogramming ability upon SCNT when early 2-cell embryos are used as donors, as compared to late 2-cell embryos. Importantly, whereas early 2-cell embryos activate MERVL, have no detectable p60 at replication sites and have no chromocenters, late 2-cell embryos possess none of these three characteristics. Thus, although the recipient oocyte or zygote used for SCNT is potentially totipotent, the capacity to reprogram a donor nucleus also depends on the potency state of the donor nucleus itself. Therefore using SCNT to address the state of the donor nucleus would be one valid approach. However, interpretation of these results should be complemented with a careful and comprehensive molecular analysis. Our SCNT data demonstrate that the rate of development for embryos derived upon transfer from endogenous 2C-like and CAF-1-induced 2C-like cells is 4- and 2.5-fold higher, respectively, than for ES cells, thus indicating that endogenous and induced 2C-like cells have higher reprogrammability than ES cells and pointing toward a broader developmental capacity.

Until now, resetting to totipotency has been possible only through SCNT. Although transcriptional reprogramming effectively reverts differentiated cells to the pluripotent state, it appears that chromatin reprogramming may be a key process in acquisition of totipotency. Indeed, ES cells acquire higher reprogrammability upon CAF-1 depletion. Understanding the molecular mechanisms that drive and can induce totipotent features *in vitro* is essential to understanding of how a maximum degree of cellular plasticity can be achieved and maintained, thereby providing more options for efficient reprogramming and potential therapeutic avenues. Our results suggest that an atypical chromatin-assembly pathway that results in global changes in chromatin accessibility and histone mobility lies at the nuclear and molecular foundations of totipotency.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. RNA-seq data have been deposited in the ArrayExpress database under accession code [E-MTAB-2684](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank G. Almouzni and J.P. Quivy (Institut Curie) for providing antibodies to p150 and p60 and for helpful discussions, M. Takeichi (RIKEN Center for Developmental Biology) for pCAG-IRES-Hygro vector, S. Matoba for sharing the list of RRRs, M. Oginuma for help with embryo analysis, B. Jost, M. Philips and S. Vicaire from the Sequencing facility of the Institut Génétique Biologie Moléculaire Cellulaire, C. Ebel for support with FACS, M. Koch for advice on imaging, C. Noll for help in western blotting analysis, R. Diaz-Uriarte for statistical advice, C. Hug for help with code testing, P. André and M. Wattenhofer-Donze from the Institut Clinique de la Souris (ICS) for help with ES-cell work and A.J. Bannister for critical reading of the manuscript. M.-E.T.-P. acknowledges funding from EpiGeneSys NoE, ERC-Stg 'NuclearPotency', the FP7 Marie-Curie Actions ITN Nucleosome4D, the European Molecular Biology Organization Young Investigator Programme and the Fondation Schlumberger pour l'Education et la Recherche. J.M.V. acknowledges funding from EpiGeneSys NoE, Deutsche Forschungsgemeinschaft Cells-in-Motion Cluster of Excellence (EXC 1003-CiM), University of Münster and the Max Planck Society. Work in T.W.'s laboratory is funded through KAKENHI 23248048 and the Takeda Science Foundation. T.I. is supported as a recipient of postdoctoral fellowships from Uehara Memorial Foundation and Human Frontier Science Programme, and A.B. was supported as a recipient of a fellowship from the Association pour la Recherche Contre le Cancer. D.R.-T. is supported by a Dirección General de Cooperación e Internacionalización fellowship from the National University of Mexico. R.E.-G. is a member of the graduate school International Max Planck Research School—Molecular Biomedicine, Münster, Germany. This work received partial support from ANR-10-LABX-0030-INRT.

AUTHOR CONTRIBUTIONS

T.I. designed and performed most experiments and analyzed data. A.B. performed FRAP, and C.Z.-B. performed Southern blotting. R.E.-G., D.R.-T. and J.M.V. performed computational analyses. E.M. and T.W. performed nuclear transfer. M.-E.T.-P. and J.M.V. analyzed data and directed the study. M.-E.T.-P. wrote the manuscript with input from T.I., R.E.-G. and J.M.V.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Macfarlan, T.S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Tarkowski, A.K. Experiments on the development of isolated blastomers of mouse eggs. *Nature* **184**, 1286–1287 (1959).
- Ishiyoshi, T. & Torres-Padilla, M.E. Towards an understanding of the regulatory mechanisms of totipotency. *Curr. Opin. Genet. Dev.* **23**, 512–518 (2013).
- Cahan, P. & Daley, G.Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14**, 357–368 (2013).
- Probst, A.V., Santos, F., Reik, W., Almouzni, G. & Dean, W. Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. *Chromosoma* **116**, 403–415 (2007).
- Probst, A.V. *et al.* A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev. Cell* **19**, 625–638 (2010).
- Puschendorf, M. *et al.* PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat. Genet.* **40**, 411–420 (2008).
- Santenard, A. *et al.* Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat. Cell Biol.* **12**, 853–862 (2010).
- Smith, S. & Stillman, B. Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication *in vitro*. *Cell* **58**, 15–25 (1989).
- Verreault, A., Kaufman, P.D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).
- Houlard, M. *et al.* CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. *PLoS Genet.* **2**, e181 (2006).
- Huang, H. *et al.* *Drosophila* CAF-1 regulates HP1-mediated epigenetic silencing and pericentric heterochromatin stability. *J. Cell Sci.* **123**, 2853–2861 (2010).
- Peaston, A.E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
- Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M.E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* **20**, 1321–1324 (2013).

16. Rolef Ben-Shahar, T. *et al.* Two fundamentally distinct PCNA interaction peptides contribute to chromatin assembly factor 1 function. *Mol. Cell. Biol.* **29**, 6353–6365 (2009).
17. Murzina, N., Verreault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
18. Kaufman, P.D., Kobayashi, R., Kessler, N. & Stillman, B. The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* **81**, 1105–1114 (1995).
19. Nabatiyan, A. & Krude, T. Silencing of chromatin assembly factor 1 in human cells leads to cell death and loss of chromatin assembly during DNA synthesis. *Mol. Cell. Biol.* **24**, 2853–2862 (2004).
20. Evsikov, A.V. *et al.* Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* **105**, 240–250 (2004).
21. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
22. Bošković, A. *et al.* Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* **28**, 1042–1047 (2014).
23. McGrath, J. & Solter, D. Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. *Science* **226**, 1317–1319 (1984).
24. Matoba, S. *et al.* Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell* **159**, 884–895 (2014).
25. Smith, S. & Stillman, B. Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* **10**, 971–980 (1991).
26. Takami, Y., Ono, T., Fukagawa, T., Shibahara, K. & Nakayama, T. Essential role of chromatin assembly factor-1-mediated rapid nucleosome assembly for DNA replication and cell division in vertebrate cells. *Mol. Biol. Cell* **18**, 129–141 (2007).
27. Beddington, R.S. & Robertson, E.J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733–737 (1989).
28. Morgani, S.M. *et al.* Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Reports* **3**, 1945–1957 (2013).
29. Wood, S.A. *et al.* Simple and efficient production of embryonic stem cell-embryo chimeras by coculture. *Proc. Natl. Acad. Sci. USA* **90**, 4582–4585 (1993).
30. Abad, M. *et al.* Reprogramming *in vivo* produces teratomas and iPS cells with totipotency features. *Nature* **502**, 340–345 (2013).
31. Hiiragi, T. & Solter, D. Reprogramming is essential in nuclear transfer. *Mol. Reprod. Dev.* **70**, 417–421 (2005).
32. Howlett, S.K., Barton, S.C. & Surani, M.A. Nuclear cytoplasmic interactions following nuclear transplantation in mouse embryos. *Development* **101**, 915–923 (1987).
33. Casanova, M. *et al.* Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. *Cell Reports* **4**, 1156–1167 (2013).

ONLINE METHODS

Cell culture and transfection. E14 mouse ES cells were cultured in DMEM containing GlutaMAX, 15% FCS, LIF, 0.1 mM 2-mercaptoethanol, nonessential amino acid and penicillin/streptomycin in gelatin-coated plates. All the experiments were performed with this medium unless otherwise indicated. Medium containing 2i (3 μ M CHIR99021 and 1 μ M PD0325901) was used during the drug selection to generate stable cell lines and during subsequent maintenance of the cells. Lipofectamine 2000 and Lipofectamine RNAi MAX (Life Technologies) were used to transfect plasmids and siRNA, respectively. The effect of RNAi was examined 2 d after transfection unless otherwise indicated. To examine the effect of RNAi 3 d after transfection, a second transfection was carried out 1 d after the first transfection to maintain the RNAi activity over 3d. We used Silencer Negative Control No. 1 siRNA (Life Technologies) as a negative control for siRNA treatment. We used Silencer select siRNA (Life Technologies) consisting of the following sequences for targeted gene knockdown:

Mouse p150 #1: CAGACUGUAUGAUCAUAGAtt (sense)
Mouse p150 #2: CGAGUGUGUCAUUAUCGAtt (sense)
Mouse p60 #1: CAACGAGCAUAAAAGUUAUtt (sense)
Mouse p60 #3: CACCAAAGCUGUCAAUUGUUtt (sense)
Mouse HIRA #1: GCAUACUGCUUUUAUCCAUtt (sense)
Mouse HIRA #2: GAUCGAAGUUUGAAGGUAUtt (sense)

Plasmid construction and generation of stable cell lines. To construct the plasmid that contains 2C::EGFP, a fragment of the MERVL element containing the 5' LTR and a partial gag-coding fragment of 729 bp in length (designated the 2C promoter) was amplified from genomic DNA from mouse embryonic fibroblasts according to ref. 1. This fragment was used then to replace the CMV promoter sequence in the pEGFP N2 vector (Clontech) so that EGFP is expressed by the promoter activity of MERVL LTR. Several stable cell lines carrying this reporter element were generated by G418 selection and subsequent clonal selection in ES medium containing 2i. The 2-cell-like cells were identified in ES-cell cultures by the presence of EGFP-positive cells that lack OCT4 protein as described before¹. We verified the karyotype in two different clones, and these clones were used for the experiments in this manuscript. To generate a reporter with a destabilized fluorescent protein, we further modified this 2C::EGFP vector to construct 2C::3XturboGFP-PEST. The 3XturboGFP-PEST sequence was cloned by PCR from a previously described vector³⁴, and the GFP sequence in 2C::EGFP was replaced with the sequence of 3XturboGFP-PEST. We also used 2C::tdTomato plasmid obtained from Addgene (plasmid 40281) to generate stable cell lines with hygromycin selection so that 2-cell-like cells could be identified by tdTomato expression. For the TAL effector (TALE) plasmids, the expression plasmid containing a TALE that recognizes major satellite sequences generated in our laboratory (pTALYM11-B15)¹⁵ was modified so that the TALE had a Flag tag and the VP64 activator domain at the C terminus.

For the p150 expression vectors, full-length p150 cDNA was amplified and cloned by PCR from a cDNA library from E14 mES cells. Subsequently, point mutations (silent mutations) were introduced on four nucleotides at the sequences targeted by siRNA p150 #1, and a Flag tag was added at the C-terminus to generate an RNAi-resistant form of p150 mRNA. This mutant template was subsequently used to generate the PIP1, HP1-binding, PIP2 or ED deletions by conventional PCR-based mutagenesis. All these mutants were cloned into the pCAG-IRES-Hygro vector (a kind gift from M. Takeichi), and the plasmids were transfected into the 2C::EGFP-reporter ES cells. ES-cell clones homogeneously expressing full-length or mutant p150 were selected by immunofluorescence with an antibody to Flag after the hygromycin selection.

Immunofluorescence. Cells were cultured on coverslips coated with gelatin and fixed with 2% or 4% paraformaldehyde in PBS for 10 min at room temperature. After the fixation, cells were permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature and incubated in blocking buffer (3% BSA in PBS) for 30 min at room temperature. Cells were incubated with primary antibodies diluted in the blocking buffer for 1–1.5 h at room temperature and washed with PBS-T (PBS containing 0.02% Triton-X100) three times. Cells were then incubated with Alexa-conjugated secondary antibodies (Life Technologies) diluted in the blocking buffer for 1–1.5 h at room temperature, washed with PBS-T three times and mounted in Vectashield solution (Vector Labs). Images were

collected on an inverted TCS SP5 confocal microscope (Leica). Antibodies to the following proteins were used: Oct4 (BD Laboratories 611203), Zscan4 (Millipore AB4340), Gapdh (Millipore MAB374-clone 6C5), turboGFP (Origene TA150042-clone 2H8), turboGFP (Thermo PA5-22688), Hira (Millipore 04-1488-clone WC119), RFP (MBL M155-3 and MBL PM005), PCNA (Santa Cruz sc-56), GFP (Abcam ab13970), Flag (Wako 018-22381 and Sigma F7425), and p60 and p150 (refs. 35,36). Secondary antibodies were Alexa 488 goat anti-mouse, A-11001; Alexa 488 goat anti-rabbit A-11008; Alexa 555 goat anti-rabbit A-21428; and Alexa 555 goat anti-mouse A-21422 (Life Technologies). For all commercial antibodies, validation and characterization can be found in the technical data sheets provided by the manufacturer.

Antibodies. Antibodies to the following proteins were used in this work: Rabbit p150 and p60 (refs. 35,36) (kind gifts from G. Almouzni and J.-P. Quivy), mouse Oct4 (611203, BD Pharmingen), rabbit and mouse RFP (PM005 and M155-3, MBL), rabbit GFP (598, MBL), chicken GFP (ab13970, Abcam), mouse turboGFP (2H8, Origene), rabbit turboGFP (PA5-22688, Thermo Scientific), mouse Flag (1E6, Wako), rabbit Flag (F7425, Sigma), mouse HIRA (WC119, Millipore), rabbit Zscan4 (AB4340, Millipore) and mouse GAPDH (MAB374, Millipore). Original images of gels and blots used in this study can be found in **Supplementary Data Set 1**.

RNA-FISH. RNA-FISH was carried out as described elsewhere³⁴. Cells were cultured on coverslips coated with laminin-511 (BioLamina). Cot1 DNA was omitted in the hybridization buffer. Probes for MERVL corresponding to the 250-bp fragment in the gag-coding region and probes for major satellite repeats were prepared by PCR with Alexa488- or TAMRA-labeled dATP, respectively. RNase A (1 mg/mL in PBS) treatment was performed at 37 °C for 30 min after the permeabilization step, as a negative control. To confirm the correlation between endogenous MERVL activation and EGFP-reporter expression, RNA-FISH for MERVL was carried out in wild-type E14 mES cells after transfection with control or p150 siRNA.

MNase Southern blotting. 2×10^5 to 5×10^5 FACS-sorted GFP-positive cells after p60 knockdown or unsorted control knockdown cells (which were also passed through the cell sorter) were pelleted at 1,000 r.p.m. at 4 °C. The cell pellet was resuspended in ice-cold NP-40 lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 0.5% NP-40, 0.15 mM spermine and 0.5 mM spermidine) and incubated on ice for 5 min. The samples were centrifuged at 1,000 r.p.m. for 3 min at 4 °C, and then washed in a wash buffer (10 mM Tris-HCl, pH 7.4, 15 mM NaCl, 60 mM KCl, 0.15 mM spermine and 0.5 mM spermidine) and centrifuged again at 1,000 r.p.m. for 3 min at 4 °C. Pellets were then resuspended in 800 μ L of ice-cold MNase digestion buffer (10 mM Tris-HCl, pH 7.4, 15 mM NaCl, 60 mM KCl, 0.15 mM spermine, 0.5 mM spermidine and 1 mM CaCl₂). A 200- μ L aliquot of each nuclei sample was digested with 1 U (5 U/mL) MNase (Roche) at 25 °C for 1 min, 5 min, 10 min and 20 min. The reaction was stopped by the addition of an equal volume of stop buffer (0.1 M Tris-HCl, pH 8.5, 20 mM EDTA and 2% SDS). The samples were treated with 0.5 mg/ml proteinase K at 37 °C overnight. DNA was extracted with phenol/chloroform and treated with RNase A (0.1 mg/mL) for 2 h at 37 °C. DNA was purified again with phenol/chloroform and concentrated by ethanol precipitation. The amount of DNA was quantified with Qubit 2.0 (Life Technologies), and an equal amount of DNA was used for subsequent Southern blotting assay. DNA fragment size was analyzed with a Bioanalyzer (Agilent Technologies) with a high-sensitivity DNA analysis kit. DNA templates to prepare the probes for Southern blotting for the MERVL LTR were generated by amplification of 200-bp MERVL LTR fragment from the 2C::EGFP reporter plasmid by PCR and subsequent DNA fragment purification. The probes were prepared by PCR with [³²P]dCTP, by a random-primer protocol.

Fluorescence-activated cell sorting. FACS Calibur (BD Biosciences) was used to quantify the population of GFP-positive cells. We used FACSAria II (BD Biosciences) to collect GFP-positive cells for MNase Southern blotting assays.

Fluorescence recovery after photobleaching (FRAP). Control, p150 or p60 siRNA was transfected into the 2C::tdTomato E14 stable ES-cell line in glass-bottomed dishes coated with gelatin. 1 d after the siRNA transfection, H3.1-GFP

expression vector was transfected into the cells. FRAP experiments were performed 1 d after the plasmid transfection, as described before²².

Real-time PCR. Total RNA was extracted from ES cells with a GenElute total RNA miniprep kit (Sigma) and treated with turbo DNase (Life Technologies) to remove genomic DNA. Reverse transcription was performed with SuperScript III (Life Technologies) with random hexamers. Real-time PCR was performed with Lightcycler 480 SYBR Green I Master Mix (Roche) on a LightCycler 480 Real-time PCR System (Roche). The relative expression level of each gene was normalized to *Gapdh* expression level. The primers used in this study^{30,37–40} are as follows (forward and reverse): MERVL (5'-CTCTACCCTGGACCATATGAC and 5'-GAGGCTCCAAACAGCATCTCTA); IAP (IAPEZI) (5'-AAGCAGCA ATCACCCACTTTGG and 5'-CAATCATTAGATGCGGCTGCCAAG); LINE1 (orf1) (5'-GGACCAGAAAAGAAATTCCTCCCG and 5'-CTCTTCTGGCTT TCATAGTCTCTGG); Zscan4 (5'-GAGATTCATGGAGAGTCTGACTGAT GAGTG and 5'-GCTGTTGTTCAAAGCTTGATGACTTC); Eif1a(-like) (5'-AACAGGCGCAGAGGTAAGAA and 5'-CTTATATGGCAGCCTCCT); Cdx2 (5'-AGGCTGAGCCATGAGGAGTA and 5'-TGAGTCCATAATTC CACTCA); Gm6763 (5'-CTGGTGGGAAGCTCTTCTTG and 5'-TCAACG TTCCAAATTCAGCA); TALE (5'-CATGGATGCAGTGAAGAAAGG and 5'-ACGTGCGTTCGCCAATAC); major satellite (5'-GCACACTGAAGGAC CTGGAATATG and 5'-GATTTCGTCAATTTTCAAGTCGTC); GAPDH (5'-CA TGGCCTTCGGTTCCTA and 5'-GCCTGCTTACCACCTTCTT).

RNA sequencing. 2 d after transfection of siRNA control, p150 #2 and p60 #1, total RNA was extracted from wild-type E14 mES cells with a GenElute total RNA miniprep kit (Sigma) and treated with turbo DNase (Life Technologies). DNase-treated total RNA was purified with an RNeasy MinElute Cleanup kit (Qiagen). Two biological replicates were prepared for each sample. Libraries for strand-specific sequencing were created with a TruSeq Stranded Total RNA with Ribo-Zero Gold Prep Kit (Illumina). Briefly, starting with 0.3 µg of total RNA, the first step involved the removal of cytoplasmic and mitochondrial ribosomal RNA (rRNA) with biotinylated, target-specific oligonucleotides combined with Ribo-Zero rRNA-removal beads. The purified RNA was fragmented into small pieces with divalent cations under elevated temperature. The cleaved RNA fragments were used to generate first-strand cDNA with reverse transcriptase and random primers, and this was followed by second-strand cDNA synthesis with DNA polymerase I and RNase H. The double-stranded cDNA fragments were blunted with T4 DNA polymerase, Klenow DNA polymerase and T4 PNK. A single A nucleotide was added to the 3' ends of the blunt DNA fragments with a Klenow fragment (3' to 5' exo minus) enzyme. The cDNA fragments were ligated to double-stranded adaptors with T4 DNA ligase. The ligated products were enriched by PCR amplification (30 s at 98 °C; 12 cycles of 10 s at 98 °C, 30 s at 60 °C and 30 s at 72 °C; 5 min at 72 °C). Excess primers were then removed by purification with AMPure XP beads (Agencourt Biosciences Corporation). Final cDNA libraries were checked for quality and quantified with a 2100 Bioanalyzer (Agilent). The libraries were loaded in the flow cell at 8-pM concentration, and clusters were generated in the Cbot. Sequencing was carried out on an Illumina HiSeq 2500 (Illumina) with a 50-bp paired-end protocol, according to Illumina's instructions. Image analysis and base calling were performed with RTA 1.17.21.3 and CASAVA 75.

RNA-sequencing analysis. Two biological replicates for p150, p60 and scrambled RNAi controls were sequenced on the Illumina HiSeq 2500 platform as paired-end 50-base reads, according to the manufacturer's standard protocols. Image analysis and base calling were performed with RTA 1.17.21.3 and CASAVA 75.

Read preprocessing. Reads were trimmed with Sickle (<https://github.com/najoshi/sickle/>). Stretches of bases with Phred scores below 20 were trimmed from both ends to a minimum length of 30 bp. Read pairs in which at least one read failed to pass the quality filters were discarded from further analysis.

Alignment to reference sequence. Reference sequence and annotation for the mouse genome (v GRCm38) were downloaded from the Illumina iGenome portal (http://support.illumina.com/sequencing/sequencing_software/igenome.ilmn). Reads were aligned to the reference genome with TopHat2 (ref. 41), taking into consideration the minimum and maximum intron size of the mouse genome. The total number of sequenced reads can be found in **Supplementary Table 9**.

Differential gene-expression analysis. Differential gene-expression analysis was performed with DeSeq2 (ref. 42). Both knockdown samples were compared to scrambled RNAi controls. Data from Macfarlan *et al.*¹ were analyzed in the same way. The HTSeq framework⁴³ was used for read counting against gene annotation with the reverse strand. This resulted in a within-replicate correlation ≥ 0.97 for all pairs (Spearman's rho) (**Supplementary Note**). Genes with an FDR-adjusted *P* value ≤ 0.05 and a \log_2 fold change of at least 1 were considered to be differentially expressed. This resulted in a total of 2,517 and 1,676 upregulated genes and 96 and 31 downregulated genes in samples treated with p150 siRNA and p60 siRNA, respectively.

Differential repeat-expression analysis. The annotation for repeat elements in the mouse genome was obtained with RepeatMasker (<http://www.repeatmasker.org/>) with RMBlast as the search engine. For differential repeat expression, reads were separated into uniquely mapping and nonuniquely mapping. Nonuniquely mapping reads were used for mapping against all occurrences of a given repeat element, with an added padding of 25 bp to both repeat-sequence ends in order to increase the likelihood of mapping reads uniquely; this was done for all repeat elements found in the mouse genome. Nonuniquely mapping reads that mapped to more than one type of repeat were excluded from further analysis. The total number of reads used for this analysis can be found in **Supplementary Table 9**. Read counts for each element, including uniquely mapping reads and reads mapping to a single element, were used as input for DeSeq2. Individual repeat elements with an FDR-adjusted *P* value ≤ 0.05 and a \log_2 fold change in read counts ≥ 1 were considered to be differentially expressed.

Distance analysis. In order to calculate the distance between genes and repeat elements, the repeat annotation from RepeatMasker was used for comparison against gene annotation. Custom-made scripts were used to count the total number of genes found in 500-bp bins located at increasing distances from each repeat occurrence or directly overlapping them. Whole-gene annotation was retrieved from Biomart (GRCm38.p2). Overall, 513 genes from Biomart did not match the iGenomes annotation used to map reads and were therefore discarded from further analysis. Custom-made scripts were used to obtain the distance from each gene to the closest repeat occurrence for either of the MERVL LTR repeats mt2, mt2b1 or mt2c. The distance was calculated as the number of base pairs between the annotated transcription start site of a gene to the closest end of a repeat without taking into account the orientation of the repeat element. The distance distributions to repeats for genes that had previously been identified as being upregulated in either p150 or p60 siRNA samples or 2-cell embryo samples from data from Macfarlan *et al.* or both (subsets described in **Fig. 4c** and **Supplementary Fig. 5d**) were compared in a pairwise fashion with two-sided Wilcoxon rank-sum tests with Bonferroni correction for multiple testing. Individual gene subsets were further compared to the distance distribution of their respective gene-group complement with two-sided Wilcoxon rank-sum tests with Bonferroni correction.

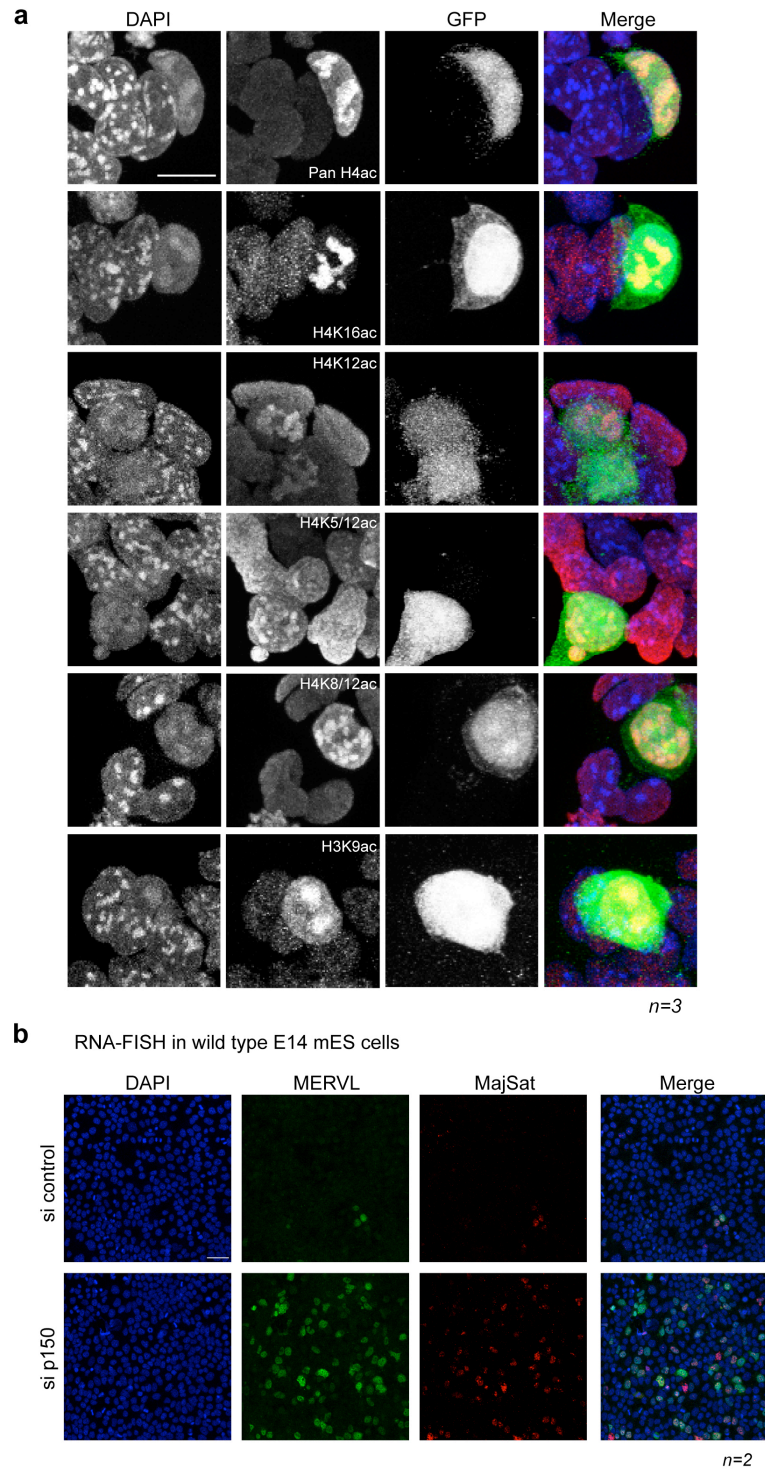
EdU labeling for 2-cell embryos. 2-cell-stage embryos were collected at 33–34 hphCG, and embryos were cultured in KSOM medium for at least 1 h before being transferred to KSOM medium containing 50 µM EdU (Life Technologies). After embryos were incubated for 45 min with EdU (which are at 34.5–36 hphCG), embryos were washed in PBS, pre-extracted with 0.1% Triton X-100 in PBS for 4 min and fixed with 4% PFA in PBS for 15 min. After that, embryos were permeabilized with 0.5% TritonX-100 in PBS for 15 min and blocked with 3% BSA in PBS. After the blocking, click-it reaction was performed for 1 h, and this was followed by washing with blocking buffer and immunostaining with antibodies to PCNA and p60. Because the cell cycle was not synchronized by hCG injection, different stages of replication was classified according to the EdU staining patterns. All experiments with mouse embryos were approved by the Ethics Committee of the Université de Strasbourg (Cometh) and performed under the authorization of French legislation. Size, age and strain information are included in ref. 34. For these experiments, no statistical method was used to predetermine sample size. The experiments were not randomized and were not performed with blinding to the conditions of the experiments.

34. Miyanari, Y. & Torres-Padilla, M.E. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* **483**, 470–473 (2012).

35. Quivy, J.P., Gerard, A., Cook, A.J., Roche, D. & Almouzni, G. The HP1-p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. *Nat. Struct. Mol. Biol.* **15**, 972–979 (2008).

36. Quivy, J.P. *et al.* A CAF-1 dependent pool of HP1 during heterochromatin duplication. *EMBO J.* **23**, 3516–3526 (2004).
37. Terranova, R., Sauer, S., Merckenschlager, M. & Fisher, A.G. The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. *Exp. Cell Res.* **310**, 344–356 (2005).
38. Maksakova, I.A. *et al.* H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. *Epigenetics Chromatin* **4**, 12 (2011).
39. Macfarlan, T.S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
40. Inoue, A., Matoba, S. & Zhang, Y. Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. *Cell Res.* **22**, 1640–1649 (2012).
41. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
42. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
43. Anders, S., Pyl, P.T. & Huber, W. HTSeq: a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

Supplementary figure 1



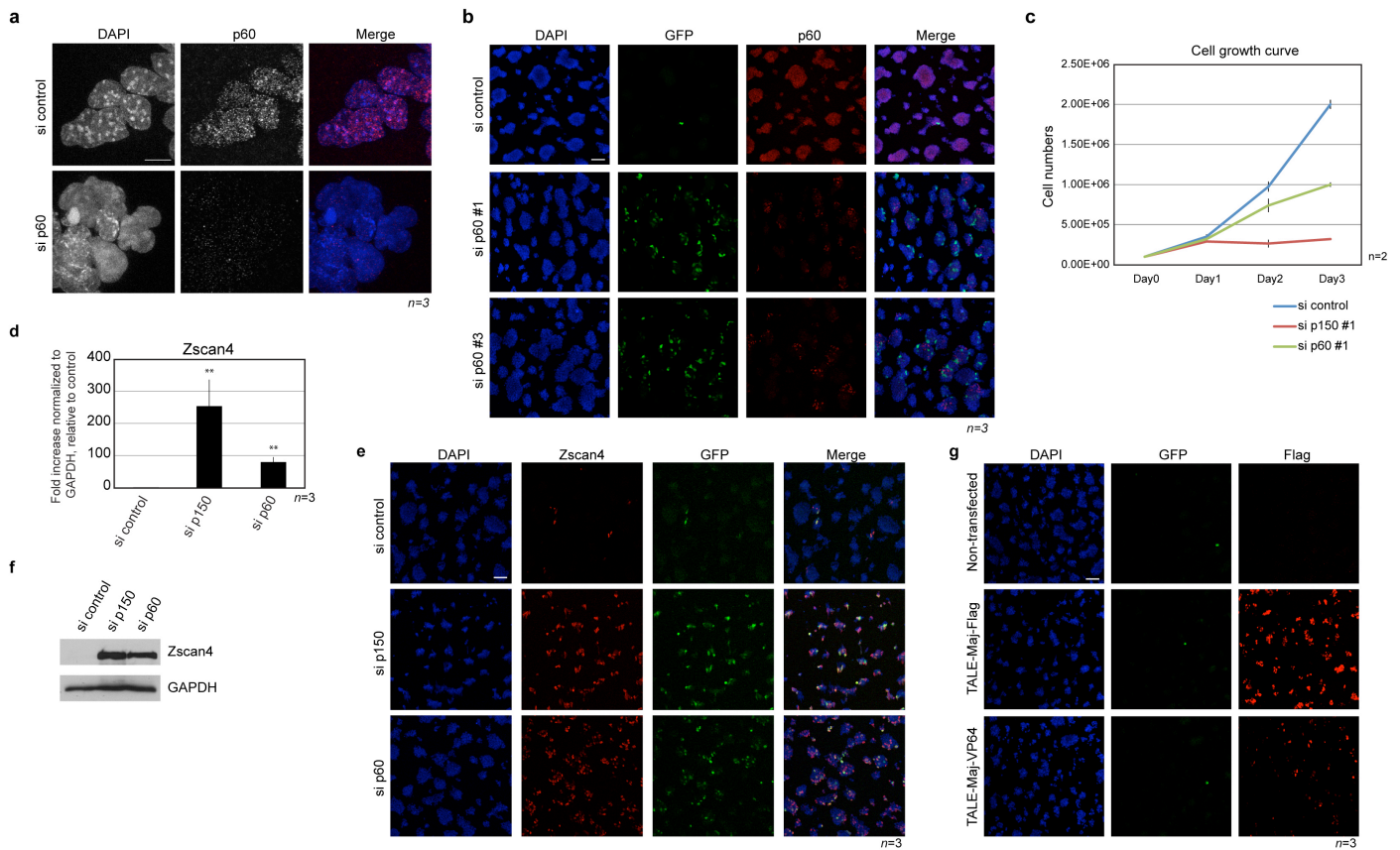
Supplementary Figure 1

Global histone acetylation in 2C-like cells and RNA-FISH for endogenous MERVL and major satellites after p150 depletion.

a. 2C::EGFP ES cells were immunostained for GFP and pan-acetylated H4, H4K16ac, H4K12ac, H4K5/12ac, H4K8/12ac or H3K9ac as indicated. The antibody for pan-acetylated H4 recognizes acetylated H4 K5, K8, K12 or K16. More than 50 cells were analysed in 3 biological replicates. Scale bar, 10 μ m.

b. RNA-FISH for MERVL and MajSat was performed in wild type E14 ES cells after transfection of siRNA for control or p150. Scale bar, 50 μ m.

Supplementary figure 2

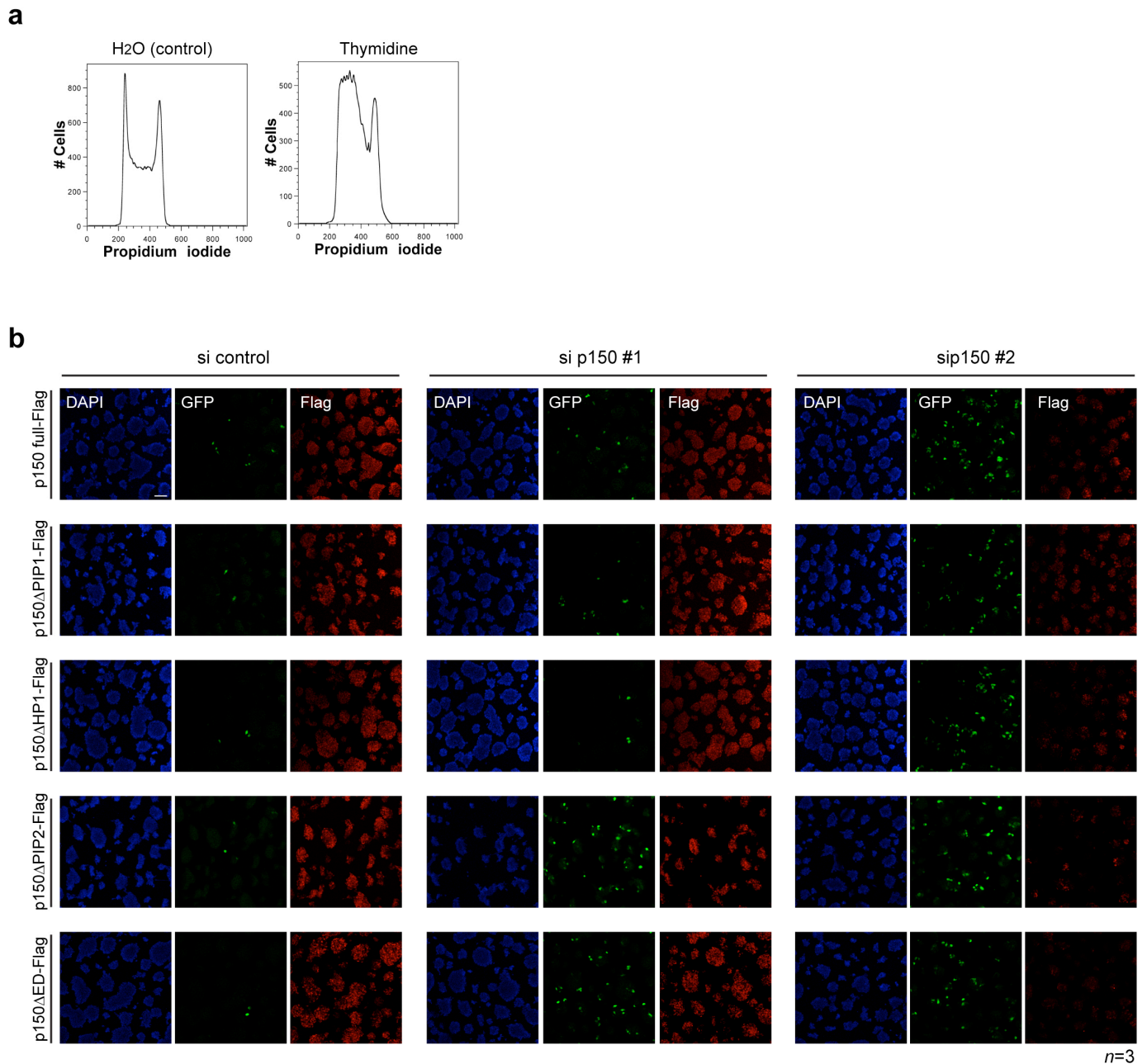


Supplementary Figure 2

Effect of p60 KD, Zscan4 activation by CAF-1 KD and use of targeted TALEs to activate major satellite transcription.

- ES cells transfected with control siRNA or p60 siRNA were immunostained for p60. Scale bar, 10 μ m. Note the absence of chromocentre after p60 loss as revealed by DAPI staining.
- As in a, but showing several 2C::EGFP ES cell colonies at a lower magnification after p60 RNAi using two different siRNAs.
- Growth curves of control or p150- or p60-depleted ES cells. Cell growth was analyzed after transfection of siRNA for control, p150 or p60 at the indicated days after transfection. The same number of cells was plated for each condition and cell numbers were counted in parallel. Shown are the mean \pm s.d. of two biological replicates.
- RT-qPCR analysis for Zscan4 expression after control, p150 or p60 RNAi in ES cells. ** $p < 0.01$ (Student's t-test, compared to si control).
- 2C::EGFP ES cells transfected with control, p150 or p60 siRNA as above and were immunostained for GFP and Zscan4. Scale bar, 100 μ m.
- Upregulation of ZSCAN4 protein levels was assessed by western blot after the transfection of control, p150 or p60 siRNA in ES cells. GAPDH was used as a loading control.
- 2C::EGFP ES cells were transfected with TALE-Maj-Flag or TALE-Maj-VP64 or mock-transfected and subsequently processed for immunostaining using an antibody for GFP and Flag. Scale bar, 100 μ m.

Supplementary figure 3



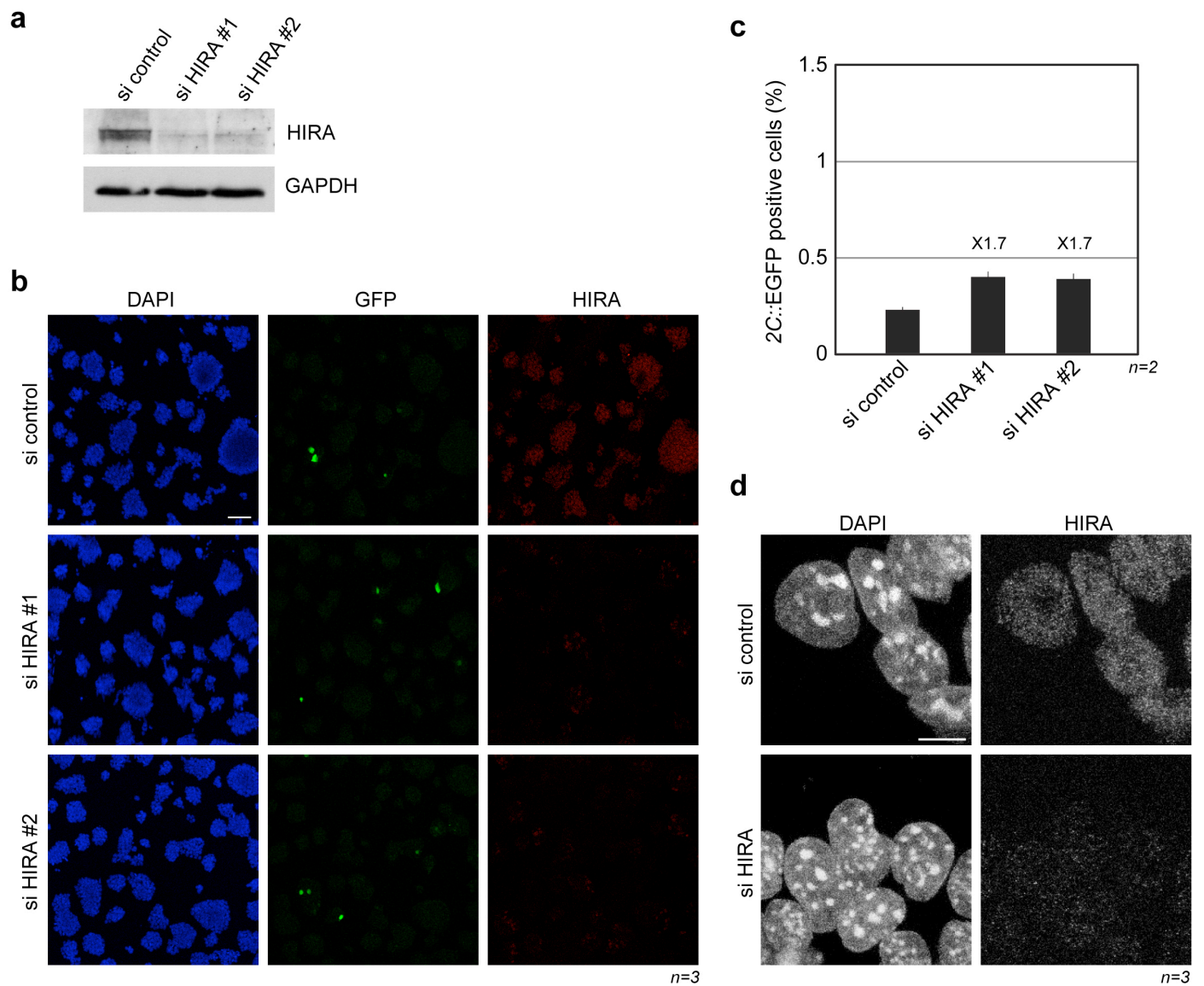
Supplementary Figure 3

G1-S cell-cycle arrest prevents the emergence of 2C-like cells after p150 RNAi and complementation assay for p150.

a. The cell cycle distribution of 2C::EGFP ES cells was analyzed by propidium iodide staining after 24 h treatment of water control or 5mM thymidine.

b. 2C::EGFP ES cells stably expressing full-length p150 or p150 mutants, as indicated on the left of each row, were immunostained for Flag and GFP after the transfection of siRNA control, si p150 #1 or si p150#2. Note that while sip150#2 targets both the endogenous p150 and the exogenous mRNAs, sip150#1 targets exclusively the endogenous p150. Scale bar, 100 μ m.

Supplementary figure 4



Supplementary Figure 4

RNAi for HIRA in ES cells does not elicit a chromocenter defect or a dramatic upregulation of the 2C::EGFP reporter.

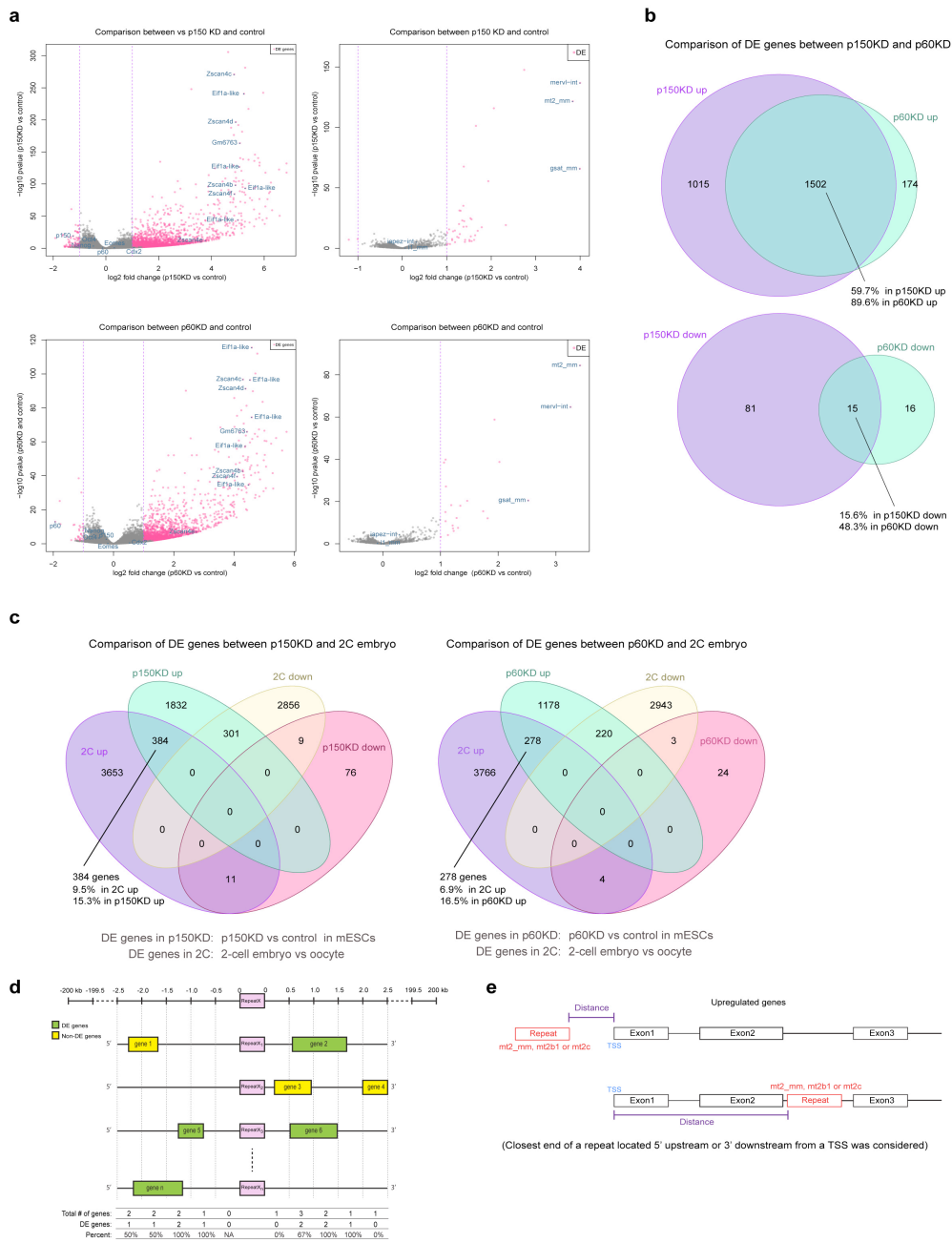
a. Knockdown efficiency for HIRA in 2C::EGFP ES cells was confirmed by western blot. Two different siRNAs were used. GAPDH was used as a loading control.

b. 2C::EGFP ES cells transfected with control siRNA or two different siRNA for HIRA were immunostained for GFP and HIRA. Scale bar, 100 μ m. Shown are representative images of two independent biological replicates.

c. Quantification of the percentage of EGFP-positive 2C-like cells by FACS following transfection of siRNA control or two different siRNA for HIRA. Shown are the mean \pm s.e.m. of two independent biological replicates. The numbers above the columns indicate the fold change of EGFP-positive cells compared to si control.

d. ES cells transfected with control or HIRA siRNA were immunostained for HIRA. The DNA was stained with DAPI. Scale bar, 10 μ m.

Supplementary figure 5



Supplementary Figure 5

Analysis of differentially expressed genes and repeats after RNAi for CAF-1 subunits p150 or p60.

a. Volcano plots of differentially expressed genes (left panels) and repetitive elements (right panels) in p150- (upper panels) and p60- (lower panels) depleted ES cells are shown. The data analysed derive from two independent biological replicates for all three conditions (within replicate correlation ≥ 0.97 for each replicate pair; Spearman's rho). An MA representation of the same data is depicted in Figure 4a.

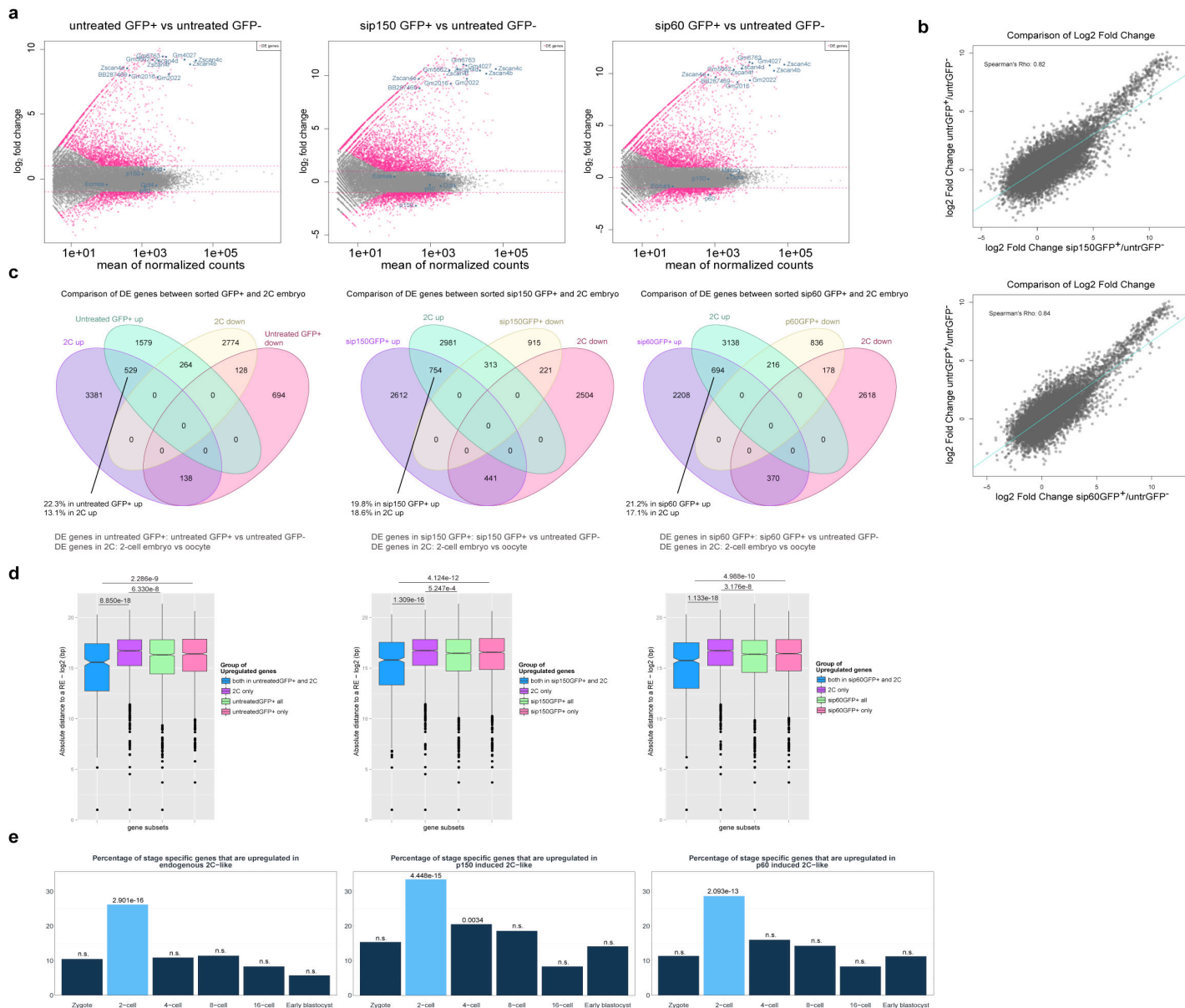
b. Venn diagram comparing up- and down-regulated genes in p150- or p60-depleted ES cells.

c. Venn diagram showing the comparison between differentially expressed genes in p150- or p60-depleted ES cells and those in 2-cell embryo. The differentially expressed genes in 2-cell stage embryo were obtained employing the same differential gene expression analysis as for the p150- and p60-depleted cells, comparing RNA-seq data from oocyte and 2-cell embryos from Macfarlan and colleagues (Macfarlan, A. *et al.*, *Nature*. **487**, 57-63, 2012).

d. Diagram showing the strategy for the distance analysis performed for Figure 4c. The genomic coordinates of a specific repeat type (Repeat X) were determined genome-wide, and windows of 500 bp upstream and downstream of each occurrence of the repeat X were used to record the genes located up to a +/- 200 kb distance from each repeat occurrence (different instances of Repeat X are depicted with subscripts). We then calculated the percentage of upregulated genes (green boxes) in p150- or p60-depleted cells among the total number of genes in a given interval.

e. Schematic representation for the distance analysis performed in Figure 4d. The genomic coordinates for the transcription start site of each set of upregulated genes were determined, and the distance to the closest end of either mt2_mm, mt2b1 or mt2c repeat was determined. Note that, in comparison to d, in which the analysis is 'repeat centric', the diagram depicted in e is 'gene centric'.

Supplementary figure 6



Supplementary Figure 6

RNA-seq analyses on FACS-sorted cells.

a. MA plots displaying differentially expressed genes in untreated GFP-positive cells, GFP-positive cells after p150 KD and GFP-positive cells after p60 KD compared to untreated GFP-negative ES cells. Data is from 2 biological replicates (Spearman's rho ≥ 0.95 for each pair of replicates; Supp. Tables 5 - 6).

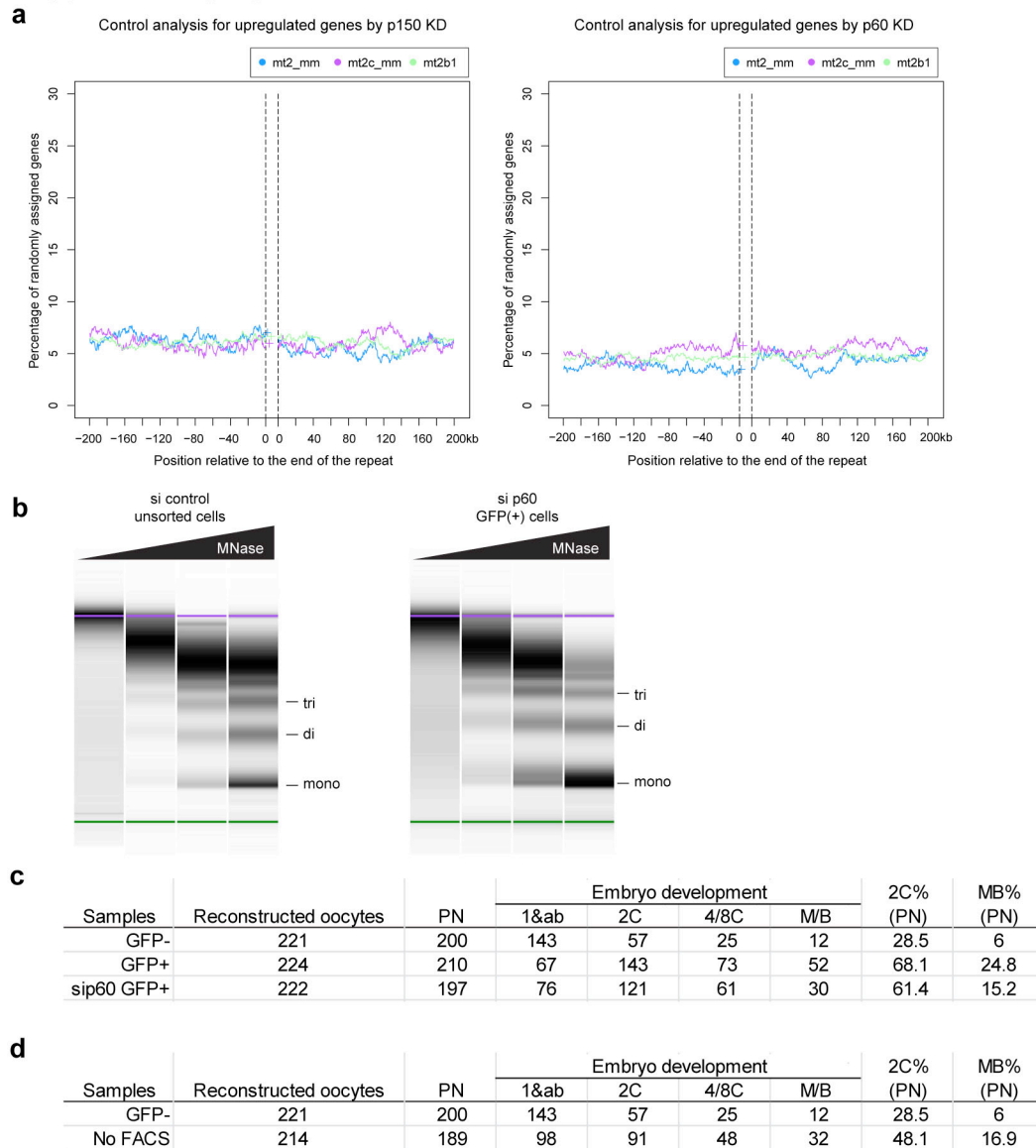
b. Comparison of fold-change differential gene expression levels between endogenous 2C-like cells (untreated GFP-positive) and p150KD 2C-like cells (left) or p60KD 2C-like cells (right). To allow a direct comparison between the samples, all differential gene expression analyses in this figure are done against a common reference of FACS sorted untreated GFP-negative cells.

c. Venn diagram showing the comparison between differentially expressed genes in untreated GFP-positive cells, GFP-positive cells after p150 KD or GFP-positive cells after p60 KD and those in 2-cell embryo. Differentially expressed genes in 2-cell embryo obtained as in Supp. Fig.7c.

d. Distribution of distances (log₂) between differentially expressed genes in untreated GFP-positive cells, GFP-positive cells after p150/p60 KD and/or genes upregulated in 2-cell embryos to 2-cell specific repeats. The absolute distance between the TSS of a gene and the end of the closest 2-cell specific MERVL family of repeat (either mt2_mm, mt2b1 or mt2c_mm) was measured for each gene within the dataset analysed. Box plots indicate the median log₂ bp distance distribution and a confidence interval around it (notches). p-values calculated with Wilcoxon tests with Bonferroni correction for multiple testing.

e. Proportion of genes specific to each embryonic stage upregulated in endogenous 2-cell like, p60 KD induced 2-cell like and p150 KD induced 2-cell like cells. Genes specific to each stage were determined by comparison of datasets in Deng et al., *Science* **343**, 193-196 (2014). A gene was considered expressed when it showed an RPKM value higher than 10 in at least 20% of the cells analyzed for its respective stage (similar results were obtained for RPKM values of 1 and 5, and cell proportions between 10-50%). A gene was considered exclusive to a stage when it was found to be expressed in only a single stage out of: Zygote, mid-2-cell, 4-cell, 8-cell, 16-cell or early blastocyst. Sets of upregulated genes in endogenous 2-cell like and p150 KD induced 2-cell like cells were selected as in Supp. Fig. 6c. Because the number of genes specific to each stage is different, we calculated the percentage of genes upregulated in each of the three 2C-like cells in relation to the number of gene subsets specific for each developmental stage. Statistical significance was assessed through Fisher's exact tests with Bonferroni correction for multiple testing. p-values are as follows: in the comparison of endogenous 2C-like cells: with zygote p=0.724; with 2-cell p=2.901e-16; with 4-cell p=0.730; with 8-cell p=1; with 16-cell p=1; with early blastocyst p=0.373. In the comparison of p60KD-induced 2C-like cells: with zygote p=1; with 2-cell p=2.092e-13; with 4-cell p=0.051; with 8-cell p=1; with 16-cell p=1; with early blastocyst p=1. In the comparison of p150KD-induced 2C-like cells: with zygote p=1; with 2-cell p=4.448e-15; with 4-cell p=0.003; with 8-cell p=1; with 16-cell p=1; with early blastocyst p=1.

Supplementary figure 7



Supplementary Figure 7

Repeat analysis randomized, global chromatin accessibility in p60-induced 2C-like cells and SCNT raw data.

a. These data correspond to the same analysis strategy as in Figure 4d but with a random assignment of differentially expressed genes the same size as in the original groups; p150 KD (left), p60 KD (right). Coloured lines represent the percentage of randomly selected differentially expressed genes among all overlapping genes up to a distance of 200kb surrounding each repeat type, mt2_mm (blue), mt2b1 (green) and mt2c_mm (purple).

b. DNA fragment distribution after MNase digestion over increasing incubation timings in control, unsorted cells and EGFP-positive cells after p60 RNAi. After purification, the DNA was analyzed using a bioanalyzer. The purple and green lines indicate the position of the molecular weight marker in the chip. The DNA size corresponding to mono-, di- or tri-nucleosome is indicated on the right of each panel.

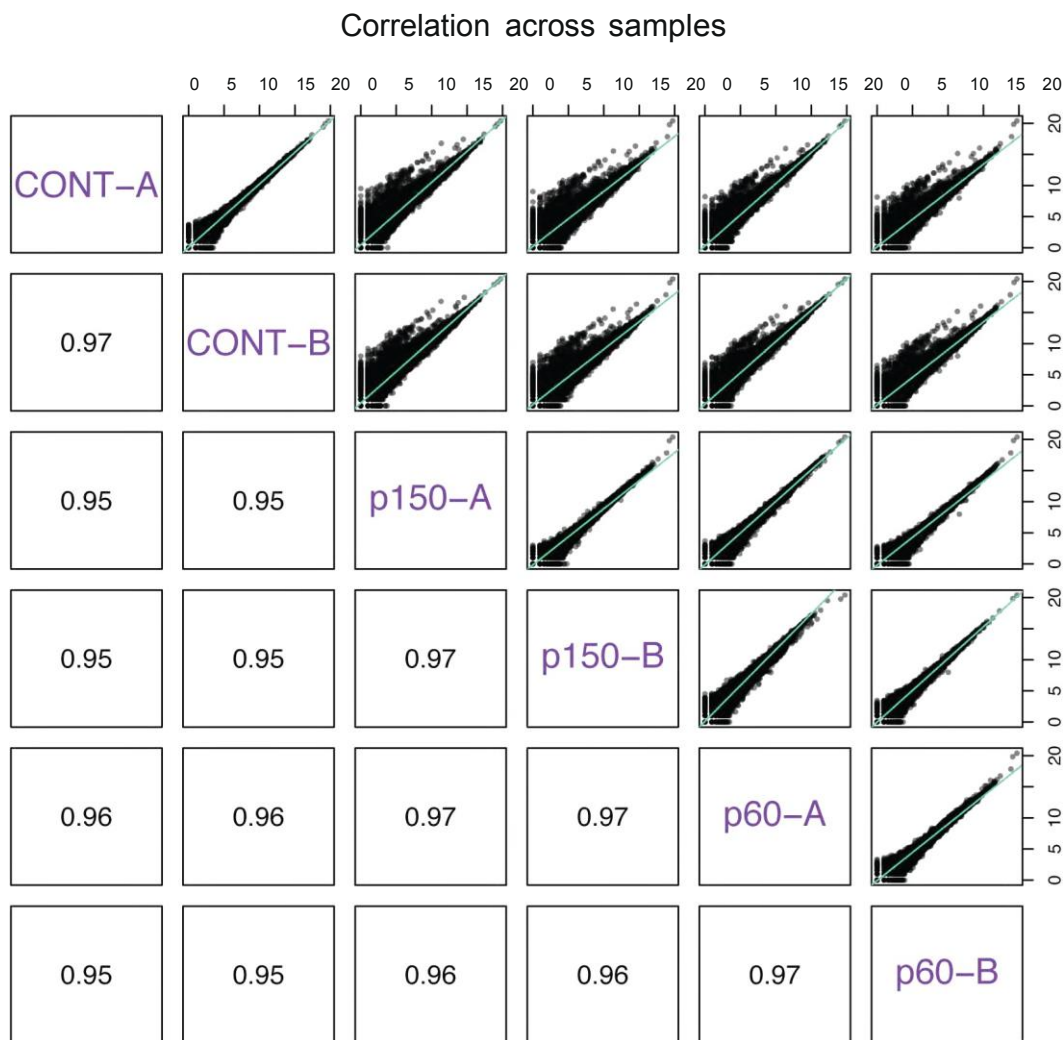
c. Raw data from nuclear transfer experiments with untreated GFP-negative (GFP-), untreated GFP-positive (GFP+) and GFP-positive cells after p60 KD (sip60 GFP+) are summarized in the table.

d. Comparison of SCNT efficiency using unsorted ES cells (no FACS) and GFP- negative ES cells. The results of nuclear transfer experiment with FACS-sorted GFP-negative cells (GFP-) and unsorted cells (No FACS) are summarized in the table. Given that the population of GFP-negative cells is ~99% in ES cells, the difference between these two groups is due to the effect of FACS procedure itself. For c and d: PN, number of NT-embryos with pronuclear formation; 1&ab, number of NT-embryos arrested at 1-cell or showing abnormal morphology; 2C, number of NT-embryos developed to 2-cell stage; 4/8C, number of NT-embryos developed to 4- or 8-cell stage; M/B, number of NT-embryos developed to morula or blastocyst. The percent of the development to 2-cell (2cell%), 4- or 8-cell (4/8C%), and morula or blastocyst (M/B%) was calculated using the number of NT-embryos that formed visible pronuclei.

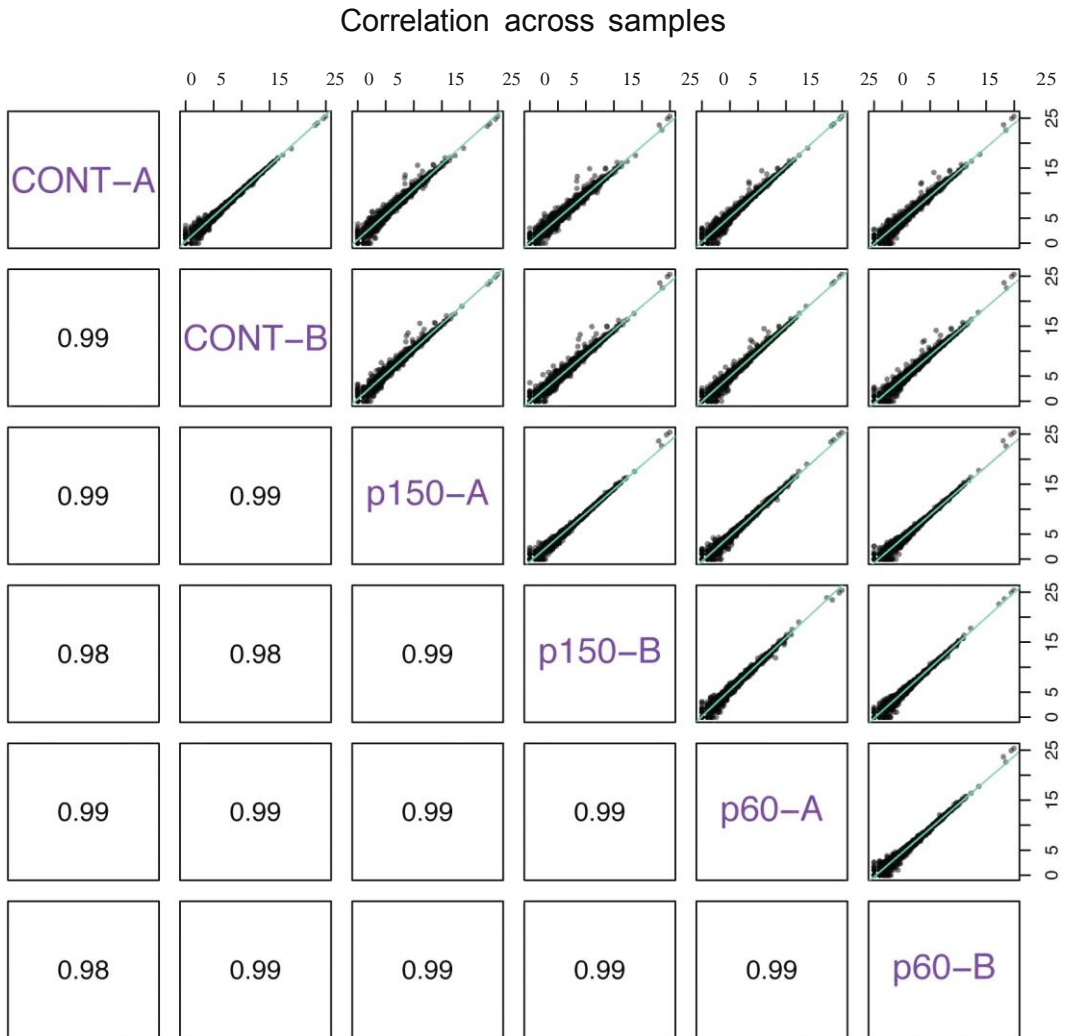
Supplementary Note.

The Supplementary Note contains the controls and explanations corresponding to all the RNAseq datasets generated in this work and is distributed as follows:

A. Correlation analysis (Spearman's rho) for single copy genes for each replicate as indicated.

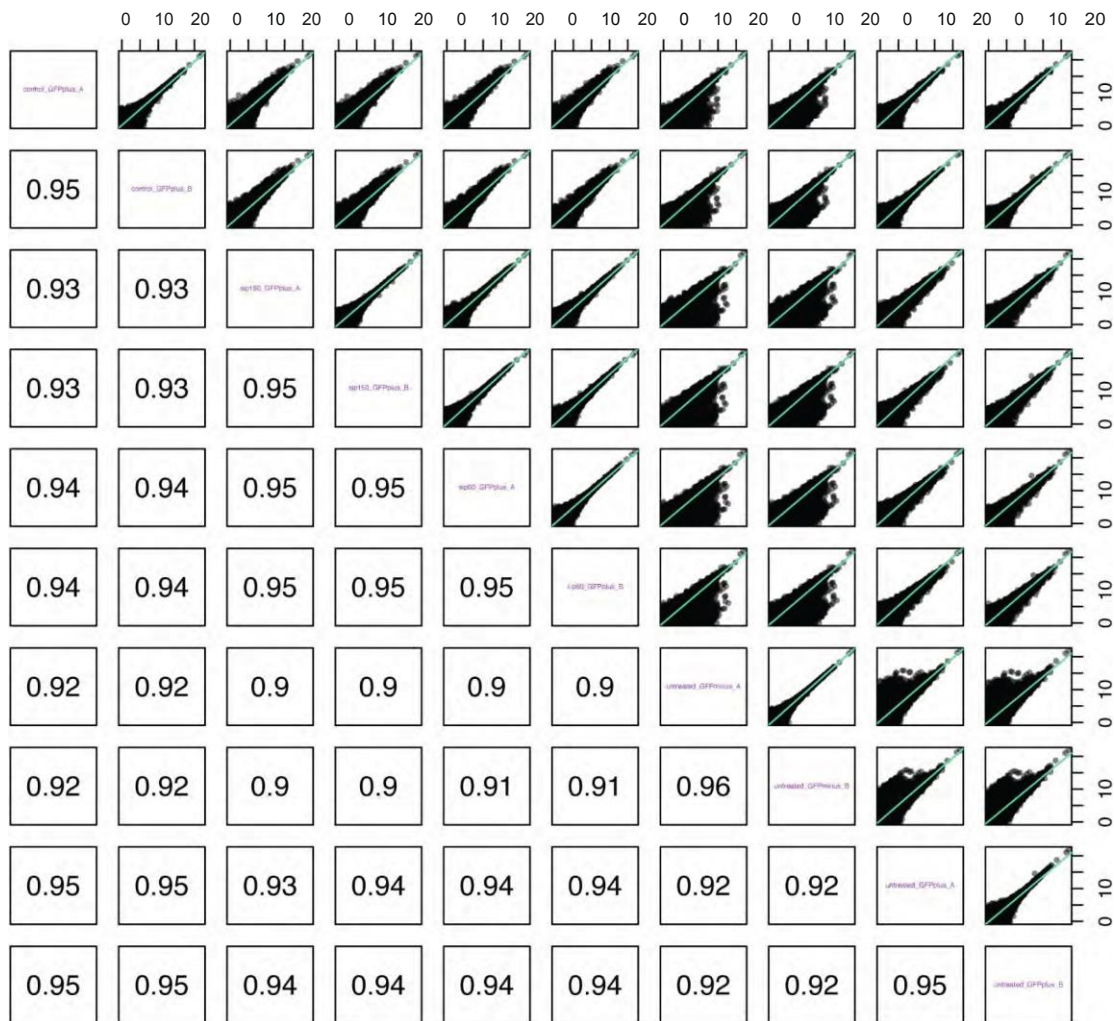


B. Correlation analysis (Spearman's rho) for repeats for each replicate as indicated.



C. Correlation analysis (Spearman's rho) for single copy genes for each replicate using sorted cells as indicated.

Correlation across samples



Supplementary references

- 1 Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63, doi:10.1038/nature11244 (2012).
- 2 Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196, doi:10.1126/science.1245316 (2014).
- 3 Casanova, M. *et al.* Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. *Cell Rep* **4**, 1156-1167, doi:10.1016/j.celrep.2013.08.015 (2013).

A molecular roadmap for the emergence of early-embryonic-like cells in culture

Statement of contribution

I hereby state that my contribution to the publication:

Rodriguez-Terrones, D.*, Gaume, X.*, Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. Nat. Genet. 50, 106–119.

consisted in the generation of the single-cell gene expression profiles, the execution of live cell microscopy experiments, the immunofluorescence imaging of pluripotency factors, the completion of the Rex1-reporter flow cytometry experiments and the implementation of all single-cell, ChIP-seq and ATAC-seq bioinformatics analyses with the exception of those included in Fig. 8b, which were carried out by Dr. Kai Kruse and Dr. Juan M. Vaquerizas. In addition, Dr. Xavier Gaume carried out the high-throughput screen, the manipulations of epigenetic factors and all of their associated experiments. Finally, Dr. Takashi Ishiuchi contributed to the experimental design of the single-cell experiments and generated most reporter cell lines.

Diego Rodriguez-Terrones

München, December 12th, 2018

Confirmation of contribution

We hereby confirm that the statement of contribution reproduced above is both truthful and accurate and represents a substantial enough contribution to warrant a first-co-authorship.

Prof. Dr. Maria-Elena Torres-Padilla

Dr. Xavier Gaume

A molecular roadmap for the emergence of early-embryonic-like cells in culture

Diego Rodriguez-Terrones^{1,2}, Xavier Gaume^{1,2}, Takashi Ishiuchi^{1,5}, Amélie Weiss², Arnaud Kopp², Kai Kruse³, Audrey Penning¹, Juan M. Vaquerizas³, Laurent Brino² and Maria-Elena Torres-Padilla^{1,4*}

Unlike pluripotent cells, which generate only embryonic tissues, totipotent cells can generate a full organism, including extra-embryonic tissues. A rare population of cells resembling 2-cell-stage embryos arises in pluripotent embryonic stem (ES) cell cultures. These 2-cell-like cells display molecular features of totipotency and broader developmental plasticity. However, their specific nature and the process through which they arise remain outstanding questions. Here we identified intermediate cellular states and molecular determinants during the emergence of 2-cell-like cells. By deploying a quantitative single-cell expression approach, we identified an intermediate population characterized by expression of the transcription factor ZSCAN4 as a precursor of 2-cell-like cells. By using a small interfering RNA (siRNA) screen, we identified epigenetic regulators of 2-cell-like cell emergence, including the non-canonical PRC1 complex PRC1.6 and the EP400-TIP60 complex. Our data shed light on the mechanisms that underlie exit from the ES cell state toward the formation of early-embryonic-like cells in culture and identify key epigenetic pathways that promote this transition.

Cellular plasticity, the ability to give rise to different cellular fates, is essential for multicellularity. Multicellular organisms derive from a single cell, the 1-cell embryo, which forms at fertilization and has the capacity to generate a full organism. This capacity is referred to as totipotency^{1–3}. Pluripotency emerges later in development and relates to the ability to form all germ layers of the embryo proper, but not the extra-embryonic annexes⁴. In the mouse, only the zygote and individual 2-cell-stage blastomeres are, strictly speaking, totipotent, as they can generate a full organism on their own^{5–7}.

Pluripotent ES cells derived from the inner cell mass of the blastocyst^{8,9} recapitulate some molecular features of the preimplantation epiblast, including expression of transcription factors such as NANOG and OCT4 (also known as POU5F1)^{10–12}. ES cell cultures are heterogeneous, with subpopulations of cells differing in gene expression in a dynamic equilibrium^{13–18}. Much of this heterogeneity results from changes in expression of pluripotency-associated transcription factors (TFs), which are part of the core regulatory network of ES cells¹⁹. In addition, the developmental potentials of ES cells grown under different conditions are not equivalent^{20–23}.

Unlike pluripotency, the molecular features of totipotency remain largely unknown. Cells resembling 2-cell-stage embryos arise spontaneously in ES cell cultures, constituting less than 1% of the culture²⁴. These 2-cell-like cells display a transcriptome that is highly similar to that of 2-cell-stage embryos, including expression of genes like members of the *Zscan4* family and the MERVL retrotransposon^{24,25}, and they have molecular features that are distinct from those of ES cells, including downregulation of the OCT4 protein²⁴, greater histone mobility²⁶ and dispersed chromocenters²⁷. Two-cell-like cells seem to have greater developmental plasticity and greater nuclear reprogrammability than ES cells^{24,27}.

The molecular regulatory networks underpinning the molecular identity and emergence of 2-cell-like cells have not been established.

Also, it is unclear whether these cells can self-renew; however, given their higher plasticity, knowledge of the underlying biology of mouse 2-cell-like cells could potentially be applied to expand the potency of existing human pluripotent cells.

Results

Single-cell analysis reconstructs the transition to the 2-cell-like state. To shed light on the molecular mechanisms underlying the emergence of embryonic-like features, we set out to identify intermediate steps in the transition from ES to 2-cell-like cells. We used transcriptional profiling at the single-cell level, which can highlight dynamic cellular states and thereby identify cell fate determinants^{28,29}. We focused mainly on chromatin modifiers and TFs, which generally display low expression levels. Because of the need for highly precise and sensitive gene expression measurements, we used a qPCR-based microfluidics approach based on the Biomark Fluidigm platform, as opposed to a poly(A)-based RNA-seq method^{30,31}.

We performed quantitative gene expression analysis in single cells using a mouse ES cell line containing the 2C::turboGFP reporter, in which turboGFP expression is driven by the MERVL promoter, thereby marking the 2-cell-like state²⁷. We identified a representative set of genes to profile on the basis of bulk RNA-seq data that compared ES cells to endogenous 2-cell-like cells and to 2-cell-like cells induced by knockdown of *Chaf1a* and *Chaf1b* (encoding CAF-1 subunits)²⁷. We selected 93 genes on the basis of their differential expression between ES and 2-cell-like cells and their functional significance (Fig. 1a, Supplementary Fig. 1a and Supplementary Table 1). Using an experimental setup designed to enrich for cells at an early stage of the transition toward the 2-cell-like state (Fig. 1b and Supplementary Fig. 1b,c), we analyzed gene expression in 136 individual turboGFP⁺ and turboGFP⁻ cells (Supplementary Table 2). Two spike-in controls allowed us to

¹Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum München, Munich, Germany. ²Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS-INSERM U964, Strasbourg, France. ³Max Planck Institute for Molecular Biomedicine, Münster, Germany. ⁴Faculty of Biology, Ludwig Maximilians Universität, Munich, Germany. Present address: ⁵Division of Epigenetics and Development, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan. Diego Rodriguez-Terrones and Xavier Gaume contributed equally to this work. *e-mail: torres-padilla@helmholtz-muenchen.de

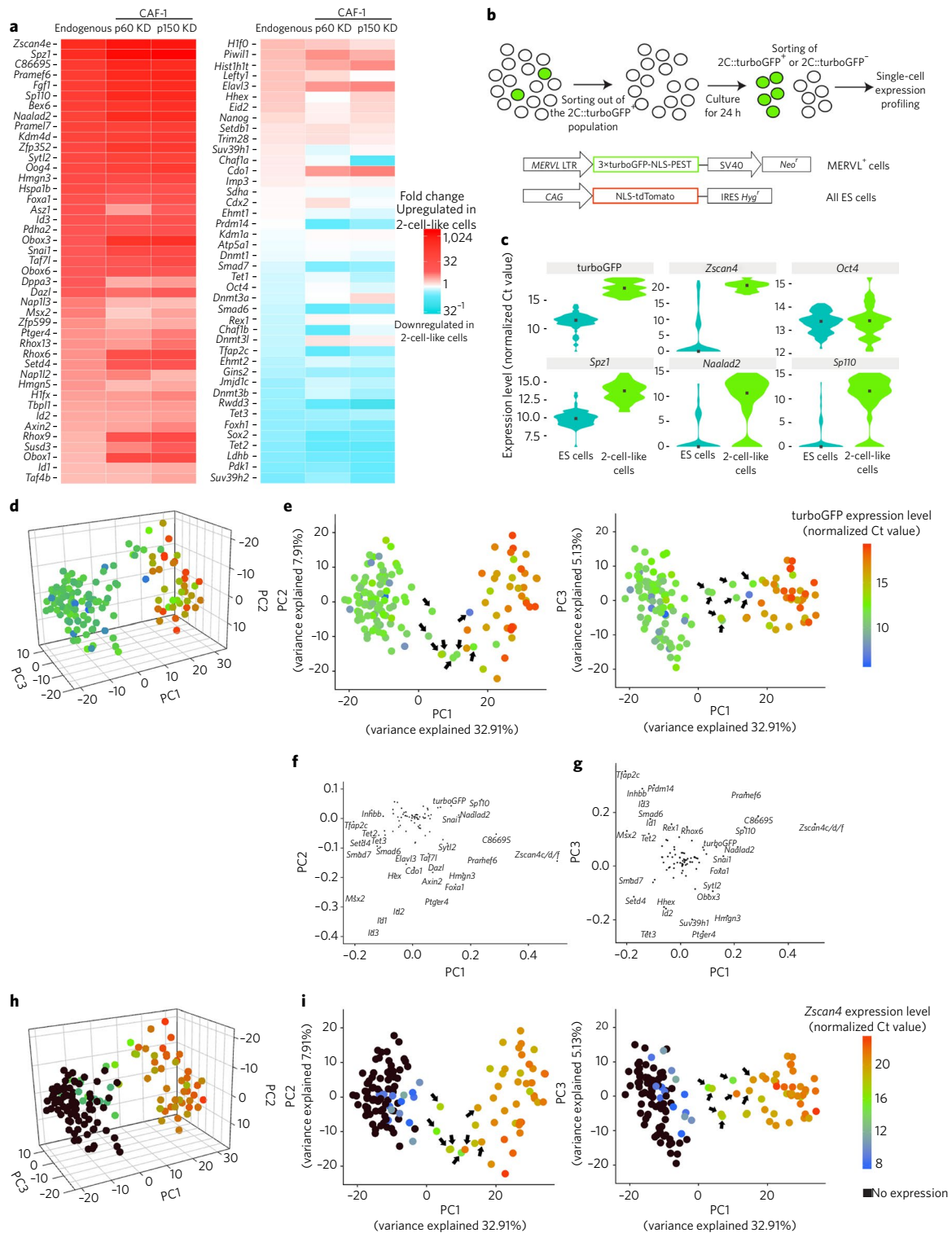


Fig. 1 | *Zscan4*⁺ cells exhibit an expression profile intermediate to those of ES and 2-cell-like cells. **a, Heat map showing changes in expression levels for the genes selected for single-cell analysis in endogenous, *Chaf1b*-knockdown-induced (p60 KD) and *Chaf1a*-knockdown-induced (p150 KD) 2-cell-like cells. Fold changes were calculated on the basis of bulk RNA-seq data²⁷ and are color-coded relative to the values in ES cells. **b**, Experimental design for the single-cell expression profiling experiments. In total, 92 turboGFP⁻ and 44 turboGFP⁺ cells were profiled. **c**, Violin plots showing the distribution of the expression levels of individual cells for the indicated genes. Higher values correspond to higher expression levels, and a Ct value of 0 indicates that no amplification was detected. The median is indicated by a square. Note that the turboGFP and *Spz1* transcripts possess no exon junctions; their readings therefore to some extent reflect the genomic background. **d, e**, PCA of the single-cell expression dataset showing all three principal components (**d**) or comparing two components at a time (**e**). Each point corresponds to a single cell, which is colored according to the expression level of turboGFP. PC1 separates ES cells from 2-cell-like cells. Black arrows indicate turboGFP⁻ cells with an expression profile intermediate to those of ES and 2-cell-like cells. **f, g**, PC2 (**f**) and PC3 (**g**) projections of the 93 genes used for the analysis, showing the contribution of each gene to the first of the three principal components. Only the most influential genes are labeled. **h, i**, PCA as in **d, e**, but individual cells (dots) are colored according to their levels of *Zscan4* expression. Black dots correspond to no expression.**

assess technical noise (Supplementary Fig. 1d). TurboGFP mRNA expression was observed in cells that were selected as turboGFP⁺ on the basis of their fluorescence, but not in cells lacking turboGFP protein expression (Fig. 1c). Two-cell-like (turboGFP⁺) cells maintained *Oct4* mRNA levels and displayed high expression of *Zscan4* transcripts and of the MT2_mm and MT2B/C long terminal repeat (LTR)-driven chimeric transcripts *Spz1*, *Naalad2* and *Sp110*, thereby validating our dataset (Fig. 1c).

To identify transitional cellular states, we performed principal-component analysis (PCA), which showed three main sources of variability in the data. Principal component 1 (PC1) separated turboGFP⁻ from turboGFP⁺ cells, and therefore defined the ES to 2-cell-like cell transition (Fig. 1d,e and Supplementary Fig. 1e). Expression of *Zscan4* transcripts and of the chimeric transcripts, in addition to that of turboGFP mRNA, provided the strongest identity to PC1 (Fig. 1f,g). PC2 highlighted heterogeneity within the 2-cell-like cell population (Fig. 1d,e). For example, 2-cell-like cells displayed graded expression of *Id1*, *Id2* and *Id3* (Fig. 1f). PC3 contrasted naive with primed pluripotency (Fig. 1g and Supplementary Fig. 1e)^{32–34}. Notably, computing the same PCA but excluding turboGFP mRNA expression produced the same results (data not shown), suggesting that the dispersion of the data is inherent to the global expression profile of individual cells and is not determined by expression from the 2C::turboGFP reporter. Thus, our single-cell analysis allows us to model acquisition of the 2-cell-like state.

***Zscan4*-expressing cells exhibit an expression profile intermediate to those of ES and 2-cell-like cells.** We next interrogated the single-cell dataset for transcripts that showed graded expression between the ES and 2-cell-like cells. PCA identified a set of cells with an intermediate expression profile that was located between the two clusters of cells, along PC1 (Fig. 1e, black arrows). These cells lacked expression of turboGFP (Fig. 1d,e), indicating that they had not yet entered the 2-cell-like state. However, they expressed the TF-encoding *Zscan4c*, *Zscan4d* and/or *Zscan4f* (*Zscan4c/d/f*) (Fig. 1h,i). The intermediate clustering of these cells was maintained even when omitting expression data for *Zscan4c/d/f* and turboGFP transcripts (Supplementary Fig. 1f,g). Thus, although expression of *Zscan4c/d/f* delineates an intermediate cellular state, this state is not solely determined by changes in expression of these genes, suggesting that cells undergo broader transcriptional reprogramming concomitantly with *Zscan4c/d/f* induction but before the acquisition of a 2-cell-like identity.

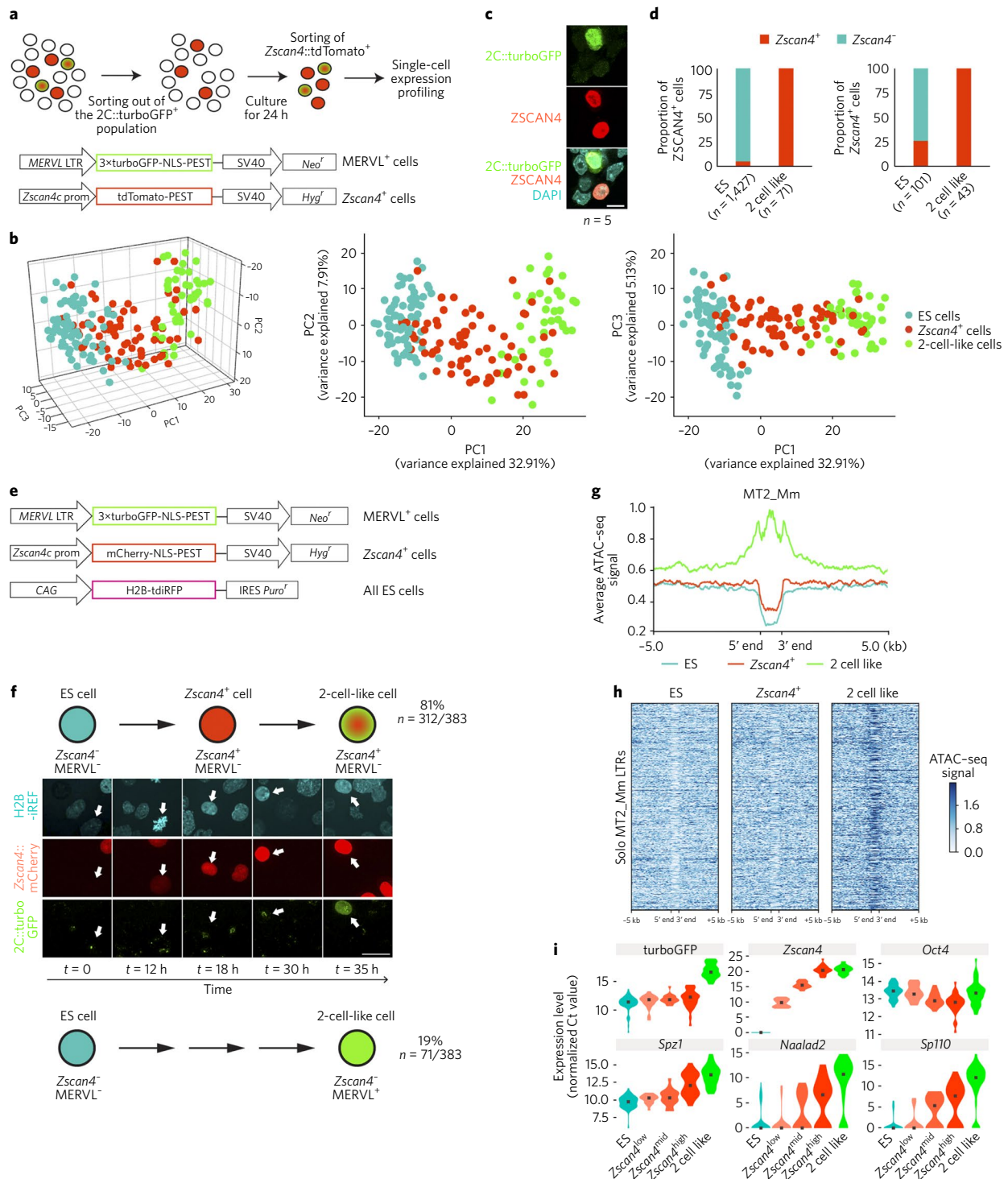
Two-cell-like cells arise primarily from *Zscan4*-expressing cells. We next addressed whether cells expressing *Zscan4c/d/f* (hereafter referred to as *Zscan4*⁺ cells) constituted an intermediate state during the transition from ES to 2-cell-like cells. Whereas knock-down of *Zscan4* expression in ES cells results in genome instability³⁵, its ectopic overexpression can induce a limited part of the 2-cell-specific transcriptome^{36,37}. We reasoned that, if 2-cell-like cells arose from *Zscan4*⁺ cells, we could formulate four predictions. First, profiling of single cells expressing *Zscan4* but not turboGFP from the 2C::turboGFP reporter should identify a cluster of cells that occupy an intermediate position along the original PC1. Indeed, additional single-cell expression profiling of *Zscan4*⁺ cells that were sorted on the basis of fluorescence from expression of the *Zscan4c::tdTomato* reporter (Fig. 2a, Supplementary Fig. 2a and Supplementary Note) and projection of the data into the previous dataset indicated that *Zscan4*⁺ turboGFP⁻ cells localized in the middle along PC1 ($n=189$ cells) (Fig. 2b, red dots). This intermediate character persisted when computing the two datasets together and when data for turboGFP and *Zscan4* were excluded from the analysis (Supplementary Fig. 2b,c). Thus, *Zscan4* expression is a marker of an intermediate cell population between the ES and 2-cell-like states, and *Zscan4*⁺ cells are not necessarily 2-cell-like cells.

Second, we reasoned that the majority of—if not all—2-cell-like cells should also express *Zscan4*. We found that all 2-cell-like cells expressed both endogenous ZSCAN4 protein (Fig. 2c,d and Supplementary Note) and *Zscan4* mRNA (Fig. 2d), consistent with a recent report³⁸. Thus, the 2-cell-like cell population is contained within the *Zscan4*⁺ population. Third, we predicted that *Zscan4*⁺ cells should convert more frequently to 2-cell-like cells than *Zscan4*⁻ cells. By using a *Zscan4c::mCherry* reporter encoding destabilized mCherry (Fig. 2e and Supplementary Fig. 2d,e), we found that *Zscan4*⁺ cells gave rise to significantly more ($\sim 9\times$) 2-cell-like cells than *Zscan4*⁻ cells ($4\pm 2.6\%$ versus $0.5\pm 0.13\%$; $P=0.04$) (Supplementary Fig. 2f,g). And, fourth, time-lapse microscopy should reveal 2-cell-like cells arising from *Zscan4*⁺ intermediates (Fig. 2e, Supplementary Fig. 2d,e and Supplementary Note). We imaged 383 emerging 2-cell-like cells and tracked them individually. We found that the majority (81%; 312/383) emerged from *Zscan4*⁺ cells (Fig. 2f and Supplementary Videos 1–3). The remaining $\sim 19\%$ (71/383) appeared directly from *Zscan4*⁻ cells (Fig. 2f), suggesting that 2-cell-like cells may arise through different pathways. Although we cannot formally rule out the possibility that these 2-cell-like cells were also derived from *Zscan4*⁺ cells because of incomplete *Zscan4c::mCherry* reporter penetrance (Supplementary Fig. 2e), our results suggest that 2-cell-like cells arise primarily from *Zscan4*⁺ precursors, thereby identifying *Zscan4* transcription as the first molecular marker in the transition from ES cells toward 2-cell-like cells.

We analyzed publicly available transposase-accessible chromatin using sequencing (ATAC-seq) profiles of these populations³⁸ and further demonstrated that *Zscan4*⁺ cells display a pattern of chromatin accessibility intermediate to those of ES and 2-cell-like cells (Supplementary Fig. 2h). Notably, ATAC-seq analysis³⁸ also showed that, although 2-cell-like cells have an open chromatin structure at MERVL sites, *Zscan4*⁺ cells do not (Fig. 2g,h). This suggests that the chromatin landscape of *Zscan4*⁺ cells differs from that of 2-cell-like cells and that *Zscan4* activation precedes chromatin opening at MERVL sites.

Next, we investigated the dynamics of *Zscan4* expression. ZSCAN4 proteins are encoded by six genes (*Zscan4a* through *Zscan4f*)²⁵. Our single-cell dataset showed that three distinct cell clusters are defined according to *Zscan4* transcript abundance: low, mid and high (Supplementary Fig. 3a). Their positions along PC1, from lowest to highest mRNA content (Supplementary Fig. 3b,c), suggest gradual activation of *Zscan4* after exit toward the 2-cell-like state (Supplementary Fig. 3a and Supplementary Note). We classified cells into five groups on the basis of their combined expression of turboGFP from the 2C::turboGFP reporter and *Zscan4*: (i) ES cells, (ii) *Zscan4*^{low} (turboGFP⁻) cells, (iii) *Zscan4*^{mid} (turboGFP⁻) cells, (iv) *Zscan4*^{high} (turboGFP⁻) cells and (v) 2-cell-like (turboGFP⁺) cells. *Oct4* transcript levels did not vary across the five groups (Fig. 2i (note Ct values) and Supplementary Table 3). *Zscan4*^{low} cells showed no difference in expression of the chimeric transcripts as compared to ES cells; however, the chimeric transcripts showed a bimodal expression distribution in *Zscan4*^{mid} and *Zscan4*^{high} cells (Fig. 2i). This suggests that LTRs are activated in different subpopulations of cellular intermediates, pointing toward a sequential model of transcriptional changes whereby *Zscan4* activation precedes activation of chimeric LTRs and MERVL itself, consistent with the results of the ATAC-seq analysis (Fig. 2g,h).

To address whether the conclusions we derived above reflect sequential transcriptional changes across the genome, we interrogated available single-cell RNA-seq datasets³⁸. We modeled single-cell RNA-seq data according to the abundance of *Zscan4* transcripts; this approach confirmed the presence of subpopulations of cells expressing various levels of *Zscan4*, which we classified into ES cells, *Zscan4*^{low}, *Zscan4*^{mid} and *Zscan4*^{high} cells, and 2-cell-like cells (Supplementary Fig. 3d,e). Differential expression analysis identified unique transcriptional changes between each transitional



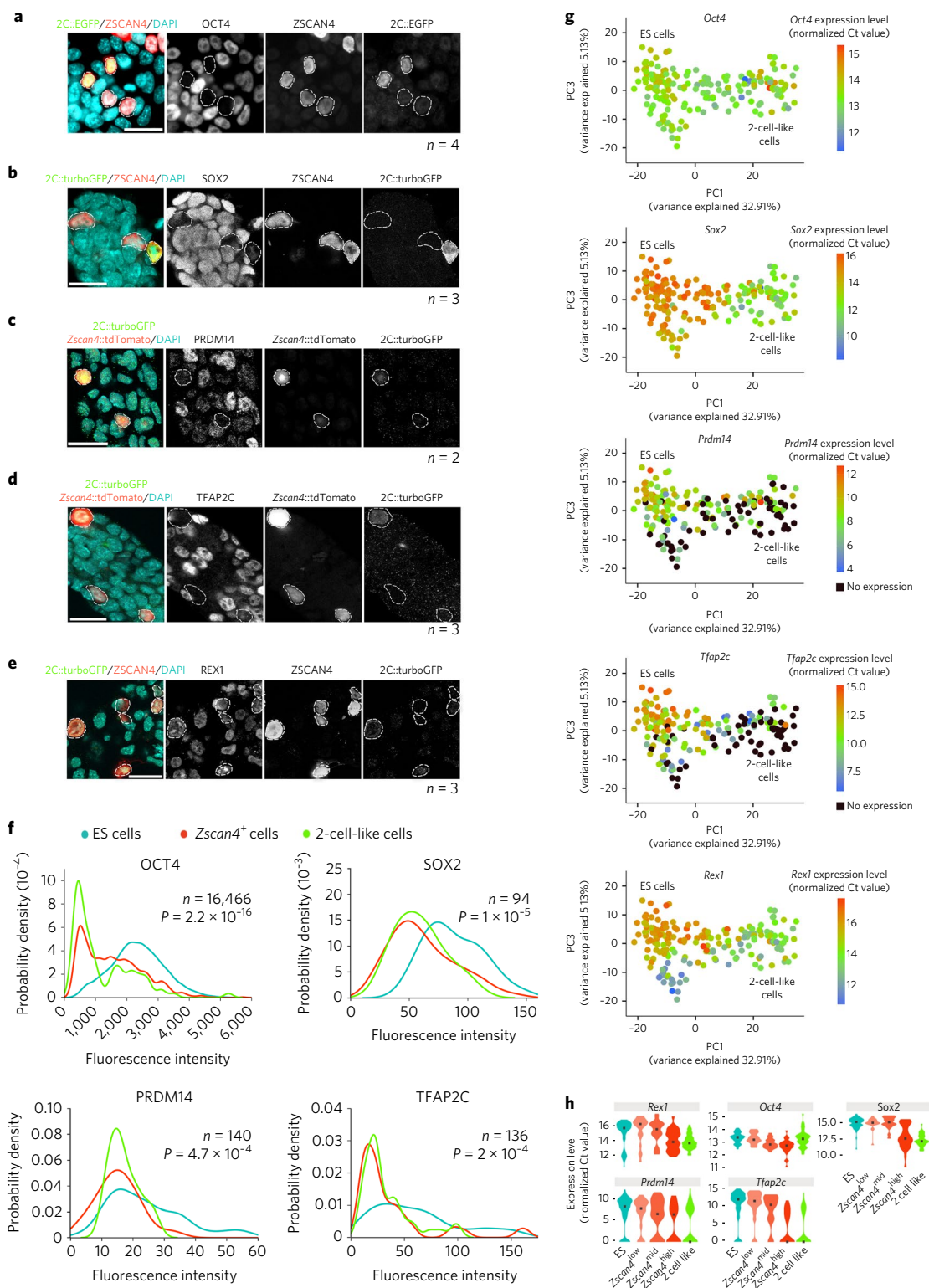


Fig. 3 | *Zscan4*⁺ cells downregulate pluripotency factors. a–e, Immunofluorescence staining using antibodies for OCT4 (**a**), SOX2 (**b**), PRDM14 (**c**), TFAP2C (**d**) and REX1 (**e**) together with 2C::turboGFP or 2C::EGFP and ZSCAN4 or *Zscan4*::tdTomato, as indicated (*n* indicates the number of independent cultures). The merged image (left) shows the DAPI (cyan), ZSCAN4 (red) and 2C reporter (green) expression. Grayscale images of the respective TF, ZSCAN4 and GFP reporter channels are shown (right). Dashed lines contour *Zscan4*⁺ and 2-cell-like cells. Scale bars, 20 μm. **f**, Quantification of the data in **a–e** (*n* indicates the number of cells). Density plots of the distribution of mean fluorescence intensities for the indicated TFs. The number of cells quantified for each graph is indicated. *P* values were calculated using the Mann–Whitney *U* test. **g**, PCA of the single-cell dataset from Fig. 2b, with single cells color-coded according to their expression level of the indicated TF. Black dots signify no expression. As in Figs. 1 and 2, PC1 separates ES from 2-cell-like cells, whereas PC3 segregates naive from primed pluripotency. **h**, Violin plots showing the distribution of the expression levels of individual cells for the indicated genes. Higher values correspond to higher expression levels, and a Ct value of 0 indicates that no amplification was detected. The median is indicated by a square. ES, *n* = 75; *Zscan4*^{low}, *n* = 21; *Zscan4*^{mid}, *n* = 20; *Zscan4*^{high}, *n* = 30; 2 cell like, *n* = 43.

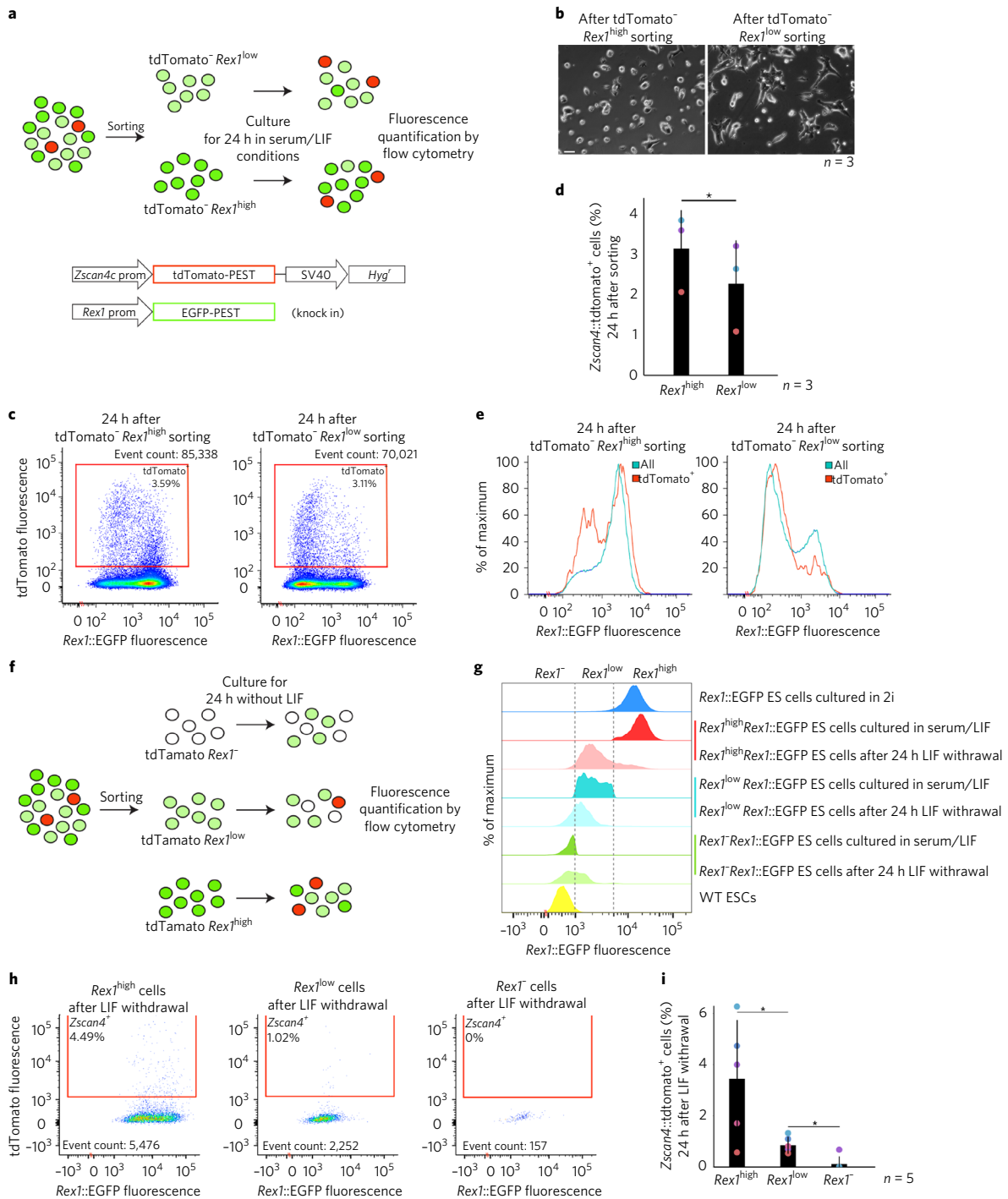


Fig. 4 | Entry into the $Zscan4^{+}$ state and downregulation of pluripotency factors are unrelated to differentiation. **a**, Experimental design for the pluripotency exit experiments. $Rex1^{high}$ and $Rex1^{low}$ $tdTomato^{-}$ cells were sorted on the basis of $Rex1::EGFP$ fluorescence and cultured for 24 h in serum- and LIF-containing medium over feeder cells. After this culture period, the number of $Zscan4^{+}$ cells was determined by FACS. **b**, Phase-contrast images of $Rex1^{high}$ and $Rex1^{low}$ cells 24 h after sorting (n indicates the number of independent cultures). Scale bar, 10 μm . **c,d**, Representative scatterplots showing $Zscan4::tdTomato$ and $Rex1::EGFP$ fluorescence measurements of individual cells as assayed by FACS (**c**) and quantification of the $Zscan4::tdTomato^{+}$ population in the $Rex1^{high}$ and $Rex1^{low}$ fractions (**d**). Error bars indicate the s.d. of three independent experiments from different cultures. $*P < 0.05$ by paired Student's t test. **e**, Density plots of $Rex1::EGFP$ fluorescence for the $Rex1^{high}$ and $Rex1^{low}$ populations after 24 h of culture in serum- and LIF-containing medium. The entire population is indicated in blue, and $Zscan4^{+}$ cells are indicated in red. Plots are representative of three independent cultures. **f**, Experimental design. Following 1 week of culture with serum- and LIF-containing medium on feeder cells, $tdTomato^{+}$ cells were removed and $Rex1^{high}$, $Rex1^{low}$ and $Rex1^{-}$ cells were replated without LIF over gelatin for 24 h, after which the proportion of $Zscan4^{+}$ cells was determined by FACS. **g**, Density plots of $Rex1::EGFP$ fluorescence for the $Rex1^{high}$, $Rex1^{low}$ and $Rex1^{-}$ populations after sorting and 24 h after LIF withdrawal. Plots are representative of five independent experiments from different cell cultures. WT, wild type. **h,i**, Representative scatterplots showing $Zscan4::tdTomato$ and $Rex1::EGFP$ fluorescence FACS measurements of individual cells (**h**) and quantification of the data (**i**). Error bars indicate s.d. of five independent experiments from different cell cultures. $*P < 0.05$ by paired Student's t test.

state and confirmed a stepwise transition from ES cells toward the 2-cell-like state globally, with the *Zscan4*⁺ intermediates showing an intermediate expression profile (Supplementary Fig. 3f,g and Supplementary Table 9).

***Zscan4*-expressing cells downregulate protein levels of pluripotency factors.** ES cell identity relies primarily on a gene regulatory network in which OCT4, NANOG and SOX2 are essential components³⁹. Two-cell-like cells downregulate OCT4 protein levels^{24,27}. To dissect the mechanism by which ES cells transition toward the 2-cell-like state, we determined at which stage of the transition the downregulation of pluripotency-associated TFs occurs. We addressed whether other pluripotency-associated factors were also affected in 2-cell-like cells (Fig. 3). OCT4 was downregulated in 2-cell-like cells, as well as in *Zscan4*⁺ intermediates (Fig. 3a,f, Supplementary Fig. 4a and Supplementary Note). SOX2, PRDM14 and AP2γ were also downregulated in *Zscan4*⁺ cells (Fig. 3b–d,f). REX1 localization changed in 2-cell-like cells, suggesting an alteration of REX1 function during the emergence of 2-cell-like cells (Fig. 3e). In addition to changes in protein levels, the single-cell expression dataset highlighted a gradient in the expression of *Sox2*, *Prdm14*, *Tfap2c* (encoding AP2γ) and *Zfp42* (also known as *Rex1*), but not of *Oct4*, along PC1 (Fig. 3g). *Sox2*, *Tfap2c* and *Rex1* transcript levels were downregulated in *Zscan4*^{high} intermediates, to levels similar to those seen in 2-cell-like cells (Fig. 3h and Supplementary Table 3). We conclude that the pluripotency regulatory network is downregulated in 2-cell-like cells, consistent with an exit from the ES cell state, but that changes in the levels of pluripotency TFs are already apparent in the *Zscan4*⁺ 2C::turboGFP⁻ state. Therefore, upregulation of *Zscan4* precedes changes in the pluripotency core regulatory network in the transition toward 2-cell-like cells.

Exit of ES cells toward the 2-cell-like state is unrelated to differentiation. The results above prompted us to investigate whether 2-cell-like cells emerge through a mechanism similar to differentiation, as both fates entail the eventual downregulation of pluripotency TFs. To address whether 2-cell-like cells exit through an early differentiation intermediate or rather exit independently of differentiation cues, we used a *Rex1* knock-in reporter⁴⁰, as *Rex1* expression faithfully reflects the earliest exit toward differentiation^{17,22}. *Rex1*⁻ cells, in contrast to *Rex1*⁺ cells, lose clonogenic self-renewal capacity when plated in serum-containing medium²¹. In addition, *Rex1*^{low} and *Rex1*⁻ cells differentiate more efficiently than *Rex1*^{high} cells^{17,22}. We reasoned that, if 2-cell-like cells emerge from differentiating ES cells, then *Rex1*^{low} cells should generate *Zscan4*⁺ cells with higher propensity *Rex1*^{high}. Using a double-reporter cell line (Fig. 4a and Supplementary Fig. 4b,c), we removed *Zscan4*⁺ cells from ES cells grown in serum- and leukemia inhibitory factor (LIF)-containing medium and replated *Rex1*^{low} and *Rex1*^{high} cells separately (Supplementary Fig. 4d). *Rex1*^{low} cells showed the expected flattened morphology after 24 h of culture (Fig. 4b)¹⁷. FACS analysis showed that *Rex1*^{low} cells generated fewer *Zscan4*⁺ cells than *Rex1*^{high} cells (Fig. 4c,d), consistent with the single-cell data showing an enrichment of *Zscan4*⁺ cells in *Rex1*^{high} cells (Supplementary Fig. 4e,f). To rule out the possibility that *Zscan4*⁺ cells within the *Rex1*^{low} fraction were derived from a population that interconverted to the *Rex1*^{high} state^{17,21,40} (Fig. 4e), we performed the same experiment as described above, but included *Rex1*⁻ cells, which cannot convert to a *Rex1*⁺ state²¹ (Fig. 4f,g). This time, the cells were plated without LIF because, after 24 h of LIF withdrawal, *Rex1*^{low} cells irreversibly lose ES cell identity and are closer to lineage specification²¹. FACS analysis confirmed that *Rex1*^{low} cells had decreased ability to generate *Zscan4*⁺ cells (Fig. 4h,i). In addition, the number of *Zscan4*⁺ cells in the *Rex1*⁻ population was significantly lower than that in the *Rex1*⁺ fraction. Thus, *Zscan4*⁺ cells—and therefore 2-cell-like cells—do not emerge from a differentiating precursor. In agree-

ment, treatment with siRNAs specific for *Oct4*, *Nanog*, *Sox2* or *Rex1* did not affect the percentage of *Zscan4*⁺ and 2C::turboGFP⁺ cells (Supplementary Fig. 4g–i). We conclude that 2-cell-like cells exit the ES cell state through a path that is different from differentiation.

Defining a molecular roadmap during the emergence of the 2-cell-like state. We next sought to define a molecular roadmap for 2-cell-like cell emergence on the basis of sequential changes in gene expression across the five cellular states described above (Supplementary Fig. 5a,b and Supplementary Table 3). Specific TFs showed substantial changes in expression mostly among the *Zscan4*⁺ intermediates (Fig. 5a). For example, the expression of *Foxa1*, which encodes a pioneer TF, was sharply upregulated at the *Zscan4*^{low} to *Zscan4*^{mid} transition. Among the 22 chromatin modifiers profiled, a few displayed marked changes accompanying each of the transitions between ES and 2-cell-like cells (Fig. 5a). The transition with the most changes in the chromatin factors analyzed was between the *Zscan4*^{mid} and *Zscan4*^{high} states. These observations prompted us to undertake a broad functional analysis to identify epigenetic factors that promote 2-cell-like fate.

Identification of epigenetic pathways regulating emergence of the 2-cell-like state. We performed a screen using an siRNA library targeting 1,167 genes⁴¹ (Fig. 5b and Supplementary Table 4). On the basis of our previous results, we used three readouts to identify bona fide 2-cell-like cells: 2C::EGFP expression, ZSCAN4 expression and loss of OCT4 protein (Fig. 5c). The screen achieved single-cell resolution through nuclear segmentation (Supplementary Fig. 6a). We included siRNAs that targeted *Chaf1a*, which encodes the p150 subunit of CAF-1, as a positive control²⁷ (Fig. 5c and Supplementary Fig. 6b) and applied stringent thresholds based on individual and combined *z* scores for hit selection (Fig. 5d,e and Supplementary Fig. 6c). Some hits led to reduced cell numbers (Supplementary Fig. 6d), suggesting toxicity and/or cell proliferation effects. Chromatin modifiers that have previously been shown to affect emergence of 2-cell-like cells, such as KAP1 and LSD1, were also in our screen; however, their downregulation had a mild effect on 2-cell-like cell induction and a lower *z* score. We extracted the top 50 hits (for which knockdown resulted in the highest percentages of 2-cell-like cells) (Fig. 5f) and validated them using a secondary screen (Supplementary Fig. 7a,b). Because the top hits of the primary screen included Polycomb proteins and histone chaperones, we included additional siRNAs to investigate all known components of these pathways. We validated all top 50 hits, with the exception of *Dnmt3b* (Supplementary Fig. 7c). In addition, treatment with siRNAs targeting transcripts encoding the Polycomb-related proteins MGA, MAX and RYBP^{42,43} and the histone chaperone DAXX induced 2-cell-like cells from 3- to 15-fold relative to cells transfected with negative-control siRNA (Supplementary Fig. 7d,e). The combined top 50 hits from both screens varied in their relative impact on 2C::EGFP, ZSCAN4 and OCT4 expression (Fig. 5f). Quantification of 2-cell-like cells in the top 50 hits showed induction that ranged between 0.28% and 2.52%, which represents a fold induction of 3.5 to 32, as compared to the controls (Fig. 5g and Supplementary Table 4).

Analysis of the protein networks of the validated top 49 hits highlighted five major complexes that seem to regulate 2-cell-like cell emergence, including 23 components of the spliceosome, 4 major members of the EP400–TIP60 (KAT5) complex, 7 members of Polycomb repressor complex 1 (PRC1) and proteins involved in DNA replication (Fig. 6a). The members of the PRC1 identified belong to a non-canonical PRC1 complex^{44,45} that is characterized by the presence of RYBP and PCGF6^{46,47}, suggesting specificity among the Polycomb group proteins in their ability to regulate 2-cell-like cell emergence. Notably, our screen identified the second CAF-1 subunit, p60 (encoded by *Chaf1b*), in agree-

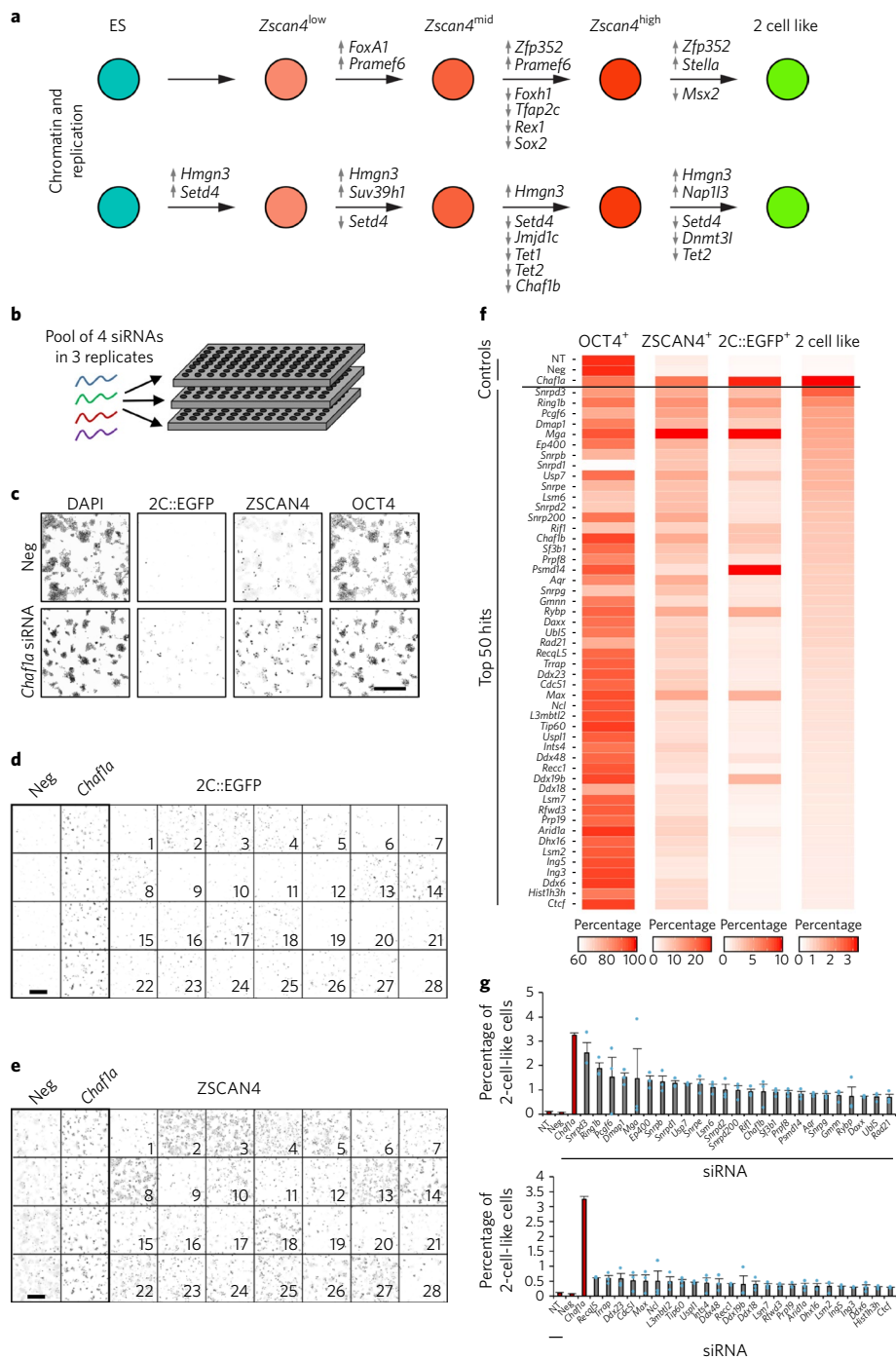


Fig. 5 | An siRNA screen for chromatin modifiers identifies regulators of 2-cell-like cell emergence. a, Schematic of significantly and differentially expressed genes between individual stages of the transition from the ES to 2-cell-like state. Changes were considered significant if they were at least twofold across cells in different individual states and had $P < 0.05$ (Mann-Whitney U test). **b**, Design of the siRNA screen. **c**, Representative inverted dynamics images for the negative (cells transfected with a scrambled siRNA; $n = 270$ wells) and positive (cells transfected with a *Chaf1a*-specific siRNA; $n = 270$ wells) controls of the screen are shown. Nine random images from one well were combined. Scale bar, 500 μm . **d,e**, Representative images ($n = 3$ technical replicate wells) of the EGFP (**d**) and ZSCAN4 (**e**) immunostaining for selected hits from the screen are shown in inverted dynamics. Scale bar, 500 μm . Images correspond to cells transfected with siRNAs to the following transcripts as indicated: *Snrpd3* (1), *Ring1b* (2), *Pcgf6* (3), *Dmap1* (4), *Ep400* (5), *Snrpb* (6), *Snrpd1* (7), *Usp7* (8), *Snrpe* (9), *Lsm6* (10), *Snrpd2* (11), *Snrp200* (12), *Rif1* (13), *Chaf1b* (14), *Sf3b1* (15), *Prpf8* (16), *Psmc14* (17), *Aqr* (18), *Snrpg* (19), *Gmnn* (20), *Dnmt3b* (21), *Ubl5* (22), *Rad21* (23), *Recq15* (24), *Trrap* (25), *Ddx23* (26), *Cdcl5* (27) and *Ncl* (28). **f**, Heat map showing the effect on OCT4, ZSCAN4 and 2C::EGFP expression of the top 50 hits of the screen, ranked according to the percentage of 2-cell-like cells induced after siRNA treatment. Data are combined from the primary and secondary validation screens and include additional hits that were tested during the secondary screen. Two-cell-like cells are defined as cells that are positive for EGFP and ZSCAN4 immunostaining but negative for OCT4 immunostaining. **g**, Quantification of 2-cell-like cells (for example, 2C::EGFP⁺, ZSCAN4⁺ and OCT4⁻) induced after transfection with siRNA for the top 50 targets (gray) from the primary and secondary screens. The mean values \pm s.d. from triplicate wells are shown. Each dot corresponds to measurements of independent wells. Negative (NT, non-transfected; Neg, scrambled siRNA) and positive (*Chaf1a*) controls are shown in red.

ment with our previous work²⁷. The mRNA expression profiles of the identified hits in endogenous and *Chaf1a*- and *Chaf1b*-knockdown-induced 2-cell-like cells suggest that many but not all of the siRNA targets analyzed are downregulated in 2-cell-like cells (Supplementary Fig. 8a). In the embryo, the expression level of all the hits obtained varied across developmental stages, with most displaying sharp downregulation or upregulation at the 4-cell stage (Supplementary Fig. 8b).

We selected 11 hits from the major chromatin pathways identified: PRC1 complexes, the EP400–TIP60 complex and the replication factor RIF1, which we validated individually with single siRNAs, first by assessing the efficiency of knockdown (Fig. 6b). Because RNA interference (RNAi) of transcripts encoding spliceosome proteins resulted in increased cell death (Supplementary Fig. 6d), we did not focus on these factors for further analysis. Instead, our subsequent analysis concentrated on *Ep400*, *Dmap1*, *Rnf2* (also known as *Ring1b*), *Pcgf6*, *Rif1*, *L3mbtl2*, *Usp7*, *Kat5* (encoding TIP60), *Mga*, *Max* and *Rybp*, all of which were effectively and specifically downregulated after transfection with specific siRNA (Fig. 6b and data not shown). We included siRNAs directed at *Chaf1a* and *Chaf1b* in all of our subsequent experiments as positive controls. All 2C::EGFP⁺ induced cells displayed robust expression of ZSCAN4 protein, similarly to endogenous and *Chaf1a*- and *Chaf1b*-knockdown-induced 2-cell-like cells (Fig. 6c, Supplementary Fig. 9). In addition, these cells expressed the GAG protein, reflecting expression of endogenous MERVL loci (Fig. 6d), and indicating that the 2C::EGFP reporter faithfully recapitulates MERVL transcription (Supplementary Fig. 9). RT-qPCR analysis demonstrated that siRNAs for all 11 hits provoked a strong increase in expression of MERVL loci, major satellites, *Zscan4* and transcripts from chimeric genes (Fig. 6e and Supplementary Fig. 10a). OCT4 protein was undetectable in 2C::EGFP⁺ cells that emerged after knockdown of the 11 hits we analyzed (Supplementary Fig. 9c). All of these results together indicate that the 2-cell-like cells induced after downregulation of the expression of our 11 new candidates are bona fide 2-cell-like cells, as they display the known molecular characteristics of endogenous 2-cell-like cells^{24,27}. FACS quantification of 2C::EGFP reporter fluorescence showed an induction of 2-cell-like cells of between 5- and 30-fold as compared to controls (Fig. 6f). *Ep400*, *Dmap1* and *Mga* were the most effective hits: downregulation of their expression by siRNA treatment resulted in an increase in the 2-cell-like cell population, which made up 5–6% of the culture, as compared with the only ~0.2% seen in controls (Supplementary Table 4). Thus, our screen effectively led to the identification of new regulators of 2-cell-like cell emergence.

We next addressed whether selected subunits of the PRC1.6 and EP400–TIP60 complexes function in the maintenance and/or induction of 2-cell-like cells by sorting out *Zscan4*⁺ cells and 2-cell-like cells followed by transfection with siRNAs targeting *Pcgf6*, *Ring1b*, *Ep400* and *Dmap1*. The percentage of 2-cell-like cells after transfection with siRNAs targeting all four subunits was the same in cultures in which *Zscan4*⁺ and 2-cell-like cells had been removed as in cultures with preexisting 2-cell-like cells (Fig. 7a), suggesting that the primary function of the corresponding proteins is in the induction rather than the maintenance of the 2-cell-like state. In addition, although siRNA targeting of most of the hits identified affected the percentages of *Zscan4*⁺ and 2-cell-like cells similarly, some of the factors identified had a bigger effect on 2-cell-like cell induction (Fig. 7b and Supplementary Fig. 10b). We conclude that components of the EP400–TIP60 complex, as well as the PRC1 complexes and RIF1, act as inhibitors of the emergence of 2-cell-like cells from ES cell cultures.

To understand the molecular nature of PRC1 activity in 2-cell-like cell formation, we determined which Polycomb complex inhibited 2-cell-like emergence by performing siRNA-mediated knockdown of the transcripts encoding all known PRC1 subunits and associated polypeptides (Supplementary Fig. 11a). Nine of the

28 siRNAs tested led to a significant increase in the 2-cell-like cell population (Supplementary Fig. 11b; protein names highlighted in red in Supplementary Fig. 11a), with subunits specific to the non-canonical PRC1 (ncPRC1) complex PRC1.6, such as PCGF6 and L3MBTL2, acting as the main gatekeepers for 2-cell-like cell emergence (Supplementary Fig. 11b). Consistent with this, downregulation of the expression of MGA and MAX, known interactors of MYC also known to assemble into PRC1.6^{46,48,49}, robustly induced 2-cell-like cells (Supplementary Fig. 11b). In addition to identifying PRC1.6 as a regulator for 2-cell-like cells, our results suggest specificity in the ncPRC1 subunits involved. Downregulation of RYBP but not of YAF-2, which is mutually exclusive with RYBP and defines different complexes⁴⁶, induced emergence of 2-cell-like cells (Supplementary Fig. 11c–f). Likewise, downregulation of RING1B, but not of RING1A, induced 2-cell-like cells (Supplementary Fig. 11c–f). In agreement with this observation, downregulation of USP7 and SKP1, which are known interactors of RING1B but for which specific allocation to a given ncPRC1 subcomplex is unknown⁵⁰, efficiently induced 2-cell-like cells. Notably, siRNAs targeting transcripts encoding the PRC2 components EED and EZH2 had no effect on 2-cell-like cell emergence or MERVL transcription (Supplementary Fig. 11g–j).

Because PRC1 catalyzes ubiquitination of histone H2A on lysine 119 (H2AK119Ub), we assessed H2AK119Ub levels in 2-cell-like cells. Immunostaining for H2AK119Ub showed that endogenous and induced 2-cell-like cells have substantially lower levels of H2AK119Ub than their neighboring ES cells (Supplementary Fig. 12a,b). This is relevant considering that reduced H2AK119Ub levels occurred not only after *Ring1b* and *Pcgf6* downregulation—for which this is expected—but also after downregulation of *Ep400* and, although to a lesser extent, *Dmap1*, suggesting that emergence of 2-cell-like cells, regardless of the molecular pathway involved, entails a decrease in H2AK119Ub levels.

We next asked whether the epigenetic pathways identified function in parallel in regulating 2-cell-like cell fate. siRNAs targeting transcripts for combinations of EP400–TIP60 and PRC1.6 subunits seemed to have an additive effect (Fig. 7c,e,f and Supplementary Note), indicating that EP400–TIP60 and PRC1.6 function through different pathways to induce 2-cell-like cells. In agreement, treatment with *Rex1* and *Nanog*-specific siRNAs did not affect the relative change in number of 2-cell-like cells after *Ep400* and *Dmap1* knockdown, but did slightly affect the extent to which downregulation of *Pcgf6* and *Ring1b* expression induced 2-cell-like cells (Supplementary Fig. 12c). This suggests that part of the effect of PRC1.6—but not of EP400 or TIP60—in inducing 2-cell-like cells is dependent on REX1 and NANOG function. In addition, culture in the presence of the inhibitors CHIR99021 and PD0325901 (hereafter 2i), which stabilize the naive pluripotent state, decreased the number of 2-cell-like cells emerging after knockdown of *Pcgf6*, *Ring1b*, *Ep400* and *Dmap1* (Fig. 7d), and overexpression of *Nanog* had no effect (Supplementary Fig. 12d).

Lastly, to investigate the potential mechanism of action of PRC1.6 and EP400–TIP60 complexes in 2-cell-like cell emergence, we asked whether they occupy genes that are differentially expressed in 2-cell-like cells relative to ES cells and analyzed ChIP-seq data for PRC1.6 and EP400–TIP60 subunits in ES cells. We classified differentially expressed genes as those that were upregulated or downregulated²⁷ and analyzed enrichment of the PRC1.6 subunits RING1B, RYBP and MAX, H2AK119Ub, and EP400 and histone H3 trimethylated on lysine 4 (H3K4me3) over their transcription start sites (TSSs). Because the transcriptional changes observed in 2-cell-like cells usually involve upregulation^{27,51}, we focused on upregulated TSSs. *k*-means clustering of ChIP-seq profiles highlighted five main binding profiles (Fig. 8a). Of these, three clusters were strongly co-occupied by RING1B, RYBP and H2AK119Ub (odds ratio for enrichment = 2.84, $P = 1.24 \times 10^{-6}$), suggesting that

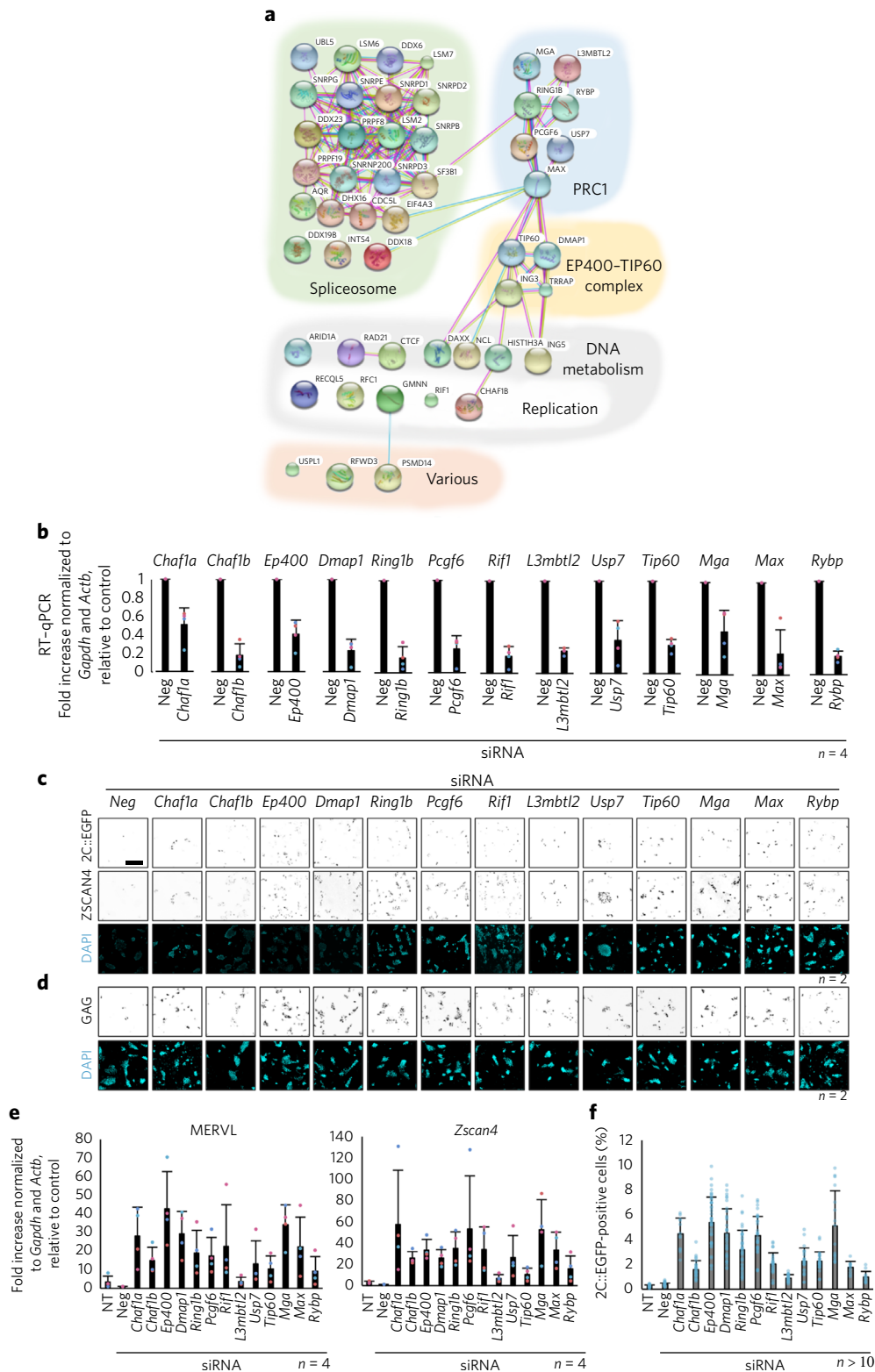


Fig. 6 | The PRC1.6 and EP400-TIP60 complexes are inhibitors of 2-cell-like cell emergence. **a**, Protein interaction network for the validated top 49 hits from the siRNA screens. **b**, RT-qPCR analysis of the indicated transcripts after transfection with the corresponding siRNAs in the 2C::EGFP reporter cell line. Shown are mean values \pm s.d. of two technical replicates from four independent cultures performed on different days. **c,d**, Immunostaining of the 2C::EGFP reporter cell line with antibodies to EGFP and ZSCAN4 (**c**) or to EGFP and the GAG protein encoded by endogenous MERVL loci (**d**) after transfection with siRNA to the indicated transcript. Representative images from two independent cell cultures are shown in inverted dynamics. Nuclei were counterstained with DAPI (cyan). Scale bar, 200 μ m. **e**, RT-qPCR analysis of MERVL and *Zscan4* expression in the 2C::EGFP reporter cell line after transfection with the indicated siRNA. Shown are mean values \pm s.d. of two technical replicates from four different cell cultures performed on different days. **f**, Quantification of EGFP⁺ cells (%) by FACS after transfection with siRNA to the indicated transcripts. The mean \pm s.d. from independent culture measurements is shown. Each dot represents measurements from independent cell cultures in **b**, **e** and **f**.

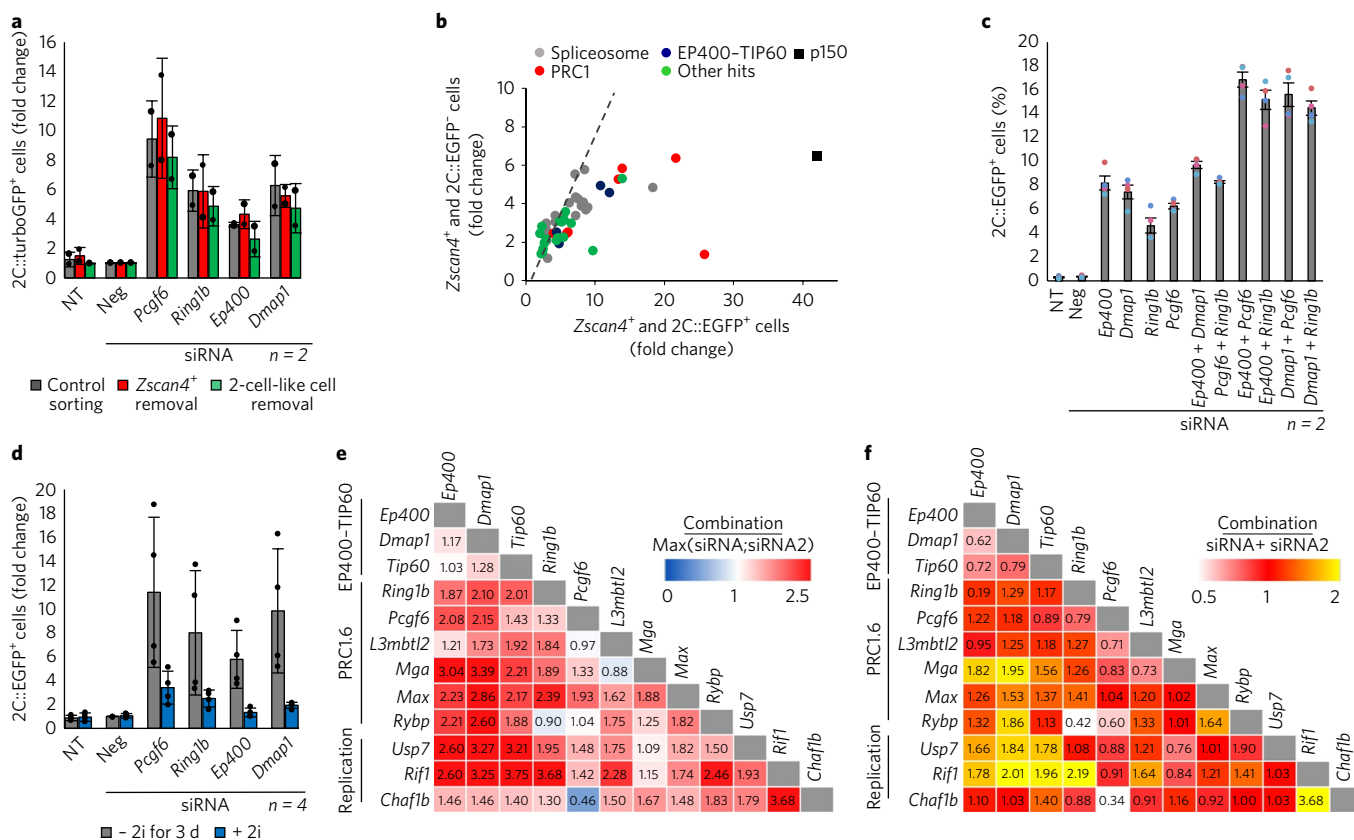


Fig. 7 | The PRC1.6 and EP400-TIP60 complexes synergistically regulate 2-cell-like cell emergence. **a**, Induction of 2-cell-like cells after transfection with siRNA for *Pcgf6*, *Ring1b*, *Ep400* or *Dmap1* in an initial population of cells devoid of 2-cell-like or of *Zscan4+* cells. In control, ‘sorted’ cells, cells were passed through the FACS sorter but *Zscan4+* and 2-cell like cells were not removed. Quantification of turboGFP+ cells by FACS was performed 2 d after siRNA transfection. The mean \pm s.d. of two different transfected wells (dots) is shown. **b**, Scatterplot of the top 50 hits of the siRNA screens showing the fold change in the percentage of ZSCAN4+ and 2C::EGFP+ cells (x axis) and of cells positive for ZSCAN4 and negative for 2C::EGFP (y axis). Each dot is color-coded according to its molecular complex. The dashed gray line indicates a 1:1 ratio. **c**, Quantification of EGFP+ cells (%) by FACS after transfection with the indicated siRNAs individually or in pairs. Shown are the mean values \pm s.d. of two independent cell cultures. **d**, Effect of 2i treatment on 2-cell-like cell induction after treatment with siRNA targeting *Pcgf6*, *Ring1b*, *Ep400* or *Dmap1*. EGFP+ cells (%) were quantified by FACS 48 h after transfection. Cells were grown in 2i and were either kept in 2i for the whole experiment or placed in medium from which 2i was removed 1 d before transfection. The mean \pm s.d. of the indicated number of independent cell cultures (shown by individual dots) is shown. **e, f**, Summary of the combinatorial siRNA analyses on 2-cell-like cell induction. Pairs of siRNAs were cotransfected into the 2C::EGFP reporter cell line, and the percentage of EGFP+ cells was measured by FACS. Combinatorial additive (**e**) or synergistic (**f**) effects were assessed.

these genes are normally repressed by PRC1 in ES cells. The TSSs and associated transcription in these three clusters were mostly silent in ES cells while expression was induced in 2-cell-like cells, where they represented ~17% of the upregulated TSSs (Fig. 8a). Two of the three PRC1-bound clusters comprised mainly ‘bivalent’ promoters marked by H3K4me3 and PRC1 occupancy (in addition to H3K27me3; data not shown). Approximately half of the upregulated TSSs (~57%) did not show binding of any of the chromatin modifiers or modifications analyzed, presumably owing to the low mappability of some of these regions (Fig. 8a). Of note, we found that many of these TSSs contained a MERVL (MT2_Mm) element within 50kb upstream (Fig. 8a), contrary to downregulated genes. The upregulation of the genes within this cluster is therefore presumably explained by proximity to MERVL^{1,51,52}. Notably, we did not observe significant binding of EP400 or PRC1 subunits RING1B or RYBP at MERVL, except at a minor fraction of complete MERVL elements that were bound by MAX (Fig. 8b and data not shown; see Methods). Because a large proportion of the downregulated genes in 2-cell-like cells were bound by EP400 (Fig. 8c) (odds ratio for enrichment = 3.70, ~60% of downregulated genes, $P < 2.2 \times 10^{-16}$), downregulation of the latter may help to repress part of the ES cell transcriptome in 2-cell-like cells. This also supports the observa-

tions that simultaneous downregulation of these two complexes synergistically induces 2-cell-like cells. Taken together, these data suggest that the transition toward the 2-cell-like cell transcriptome involves a combination of molecular pathways, including regulation through the binding and action of the identified chromatin modifiers and activation of a proximal MERVL element, presumably through binding of the TF DUX^{53,54}.

Discussion

We have identified an intermediate cellular state during the transition to the 2-cell-like state that is characterized by a transcriptional profile distinctive from those of ES and 2-cell-like cells. We note that, in contrast to previous findings in which ZSCAN4+ and MERVL+ populations have been considered interchangeable³⁸, our approach of classifying individual cells on the basis of their single-cell transcriptome rather than by their fluorescence or population-wide transcriptome allowed us to uncover differences between the populations in a more robust and quantitative manner. ZSCAN4 expression is a bona fide marker for this intermediate population. Although activation of *Zscan4* expression demarcates the first molecular change that we detect, the transcriptional changes in *Zscan4+* cells are not limited to *Zscan4*. This anticipates that

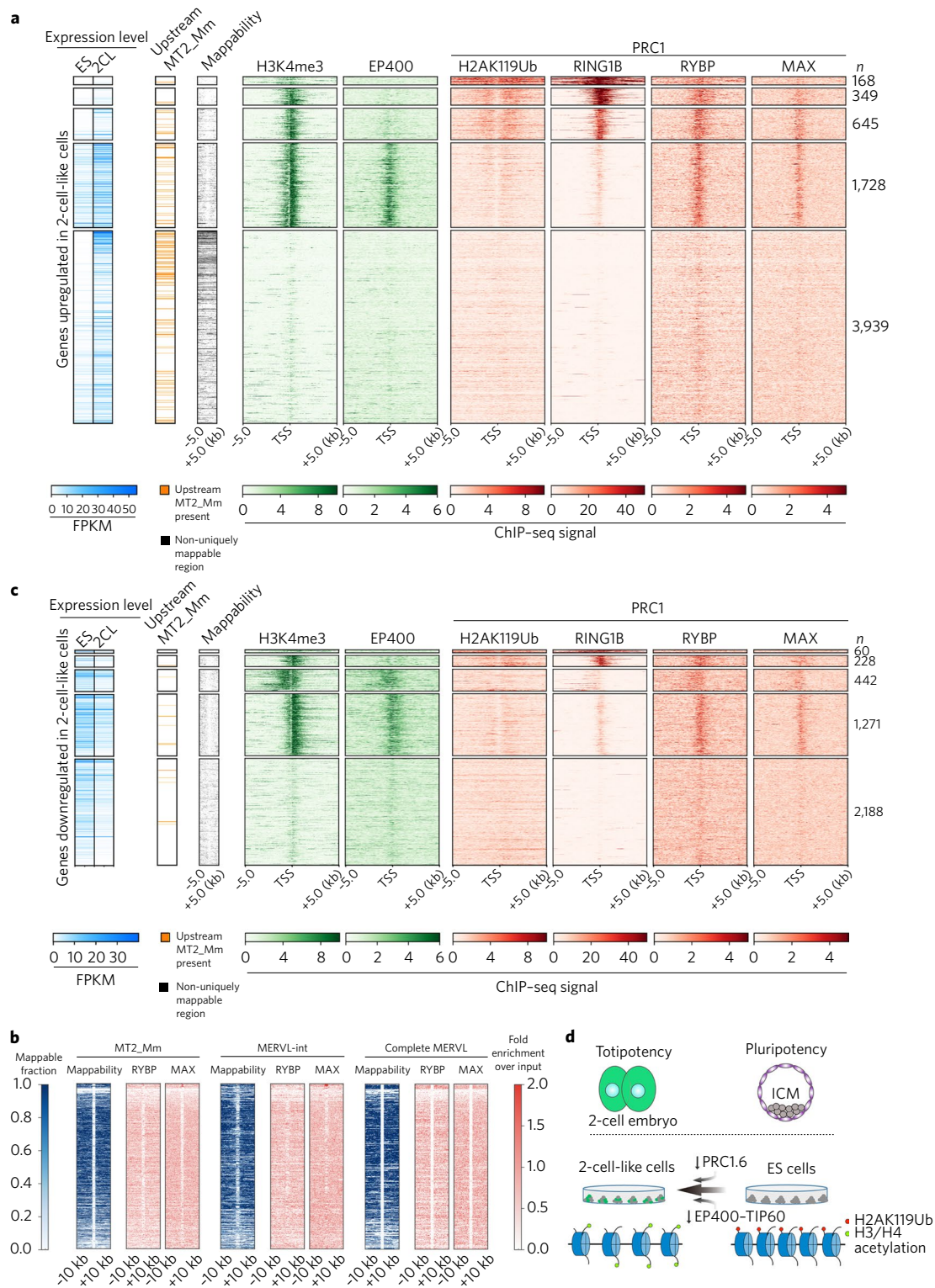


Fig. 8 | The PRC1.6 and EP400-TIP60 complexes occupy genes differentially expressed in 2-cell-like cells relative to ES cells. a, Heat map of the ChIP-seq profiles (for references, see Methods) of the indicated chromatin modifiers or histone modifications over all TSSs of the 2,313 upregulated (6,829 TSSs) genes in ES cells grown in serum- and LIF-containing medium. Regions of low mappability are depicted in black, and occurrence of an MT2_Mm site within the 50 kb upstream of the respective TSS is indicated in orange; expression levels in ES and 2-cell-like cells (2CL) are shown in blue. The number of TSSs per cluster is indicated by *n*. **b**, Heat map of RYBP and MAX enrichment (fold enrichment over input) in 20-kb windows encompassing complete, internal or solo LTR MERVL elements as indicated. The mappability score for the corresponding regions is depicted. **c**, Heat map of the ChIP-seq profiles of the indicated chromatin modifiers or histone modifications over all TSSs of the 4,189 downregulated (4,189 TSSs) genes, as in **a**. **d**, Working model for 2-cell-like cell induction from mouse ES cell cultures based on the siRNA screen results. Downregulation of various subunits of the PRC1.6 and EP400-TIP60 complexes leads to induction of the 2-cell-like state, which is accompanied by a global reduction in H2AK119Ub levels and increased histone acetylation^{24,27}. The precise interplay of the two complexes, as well as the extent to which these mechanisms may operate in vivo to repress the totipotent regulatory program, remains to be established. ICM, inner cell mass.

ZSCAN4 itself may not necessarily have a directive, essential role in the emergence of 2-cell-like cells. Indeed, ZSCAN4 overexpression in ES cells seems to elicit cell death and can trigger acquisition of some but not all of the repertoire of 2-cell-like cell characteristics³⁶.

Early-differentiating cells generate fewer *Zscan4*⁺ cells than naive ES cells, indicating that exit toward the 2-cell-like cell stage demarcates a process different from the loss of pluripotency after differentiation. Our data also indicate that 2-cell-like cells themselves are heterogeneous, but the nature of their heterogeneity differs from that of ES cells. Although 2-cell-like cells have been proposed to constitute a metastable state²⁴, it remains to be seen whether such a state is also in an internal dynamic equilibrium or whether the heterogeneity that we observed represents additional transitional states.

A more open chromatin state characterizes the 2-cell state in vivo and in vitro^{26,54,55}. The EP400–TIP60 complex regulates nucleosome stability^{56,57}, and this complex could potentially provide a molecular basis for chromatin opening. Our data suggest that several activities are in play to regulate the transition from ES to 2-cell-like cells. This is supported by the additive effects of PRC1.6 and EP400–TIP60 downregulation in 2-cell-like cell emergence. In addition, PRC1.6 is the only PRC1 complex that possesses histone deacetylation activity^{46,47}, which may also affect the global histone hyperacetylation observed in 2-cell-like cells^{24,27}. Notably, specific components of PRC1.6, such as L3MBTL2 and RYBP, are either present at low levels or absent from embryonic chromatin in 2-cell-stage embryos⁵⁸. Although work on induced pluripotency indicates that reprogramming does not necessarily recapitulate developmental progression in reverse, investigating whether the factors that we identified are responsible for regulating the 2-cell-like cell transition in vivo will be an important task for the future.

Our data support a role for several biochemical activities—including chromatin assembly, nucleosome remodeling, and histone acetylation and ubiquitination—in reshaping the chromatin landscape in the transition from ES to 2-cell-like cells (Fig. 8d). Altogether, our work identifies intermediate states during the transition from ES to 2-cell-like cells and the chromatin pathways involved.

Received: 13 February 2017; Accepted: 13 November 2017;
Published online: 18 December 2017

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41588-017-0016-5>.

References

- Ishiyuchi, T. & Torres-Padilla, M. E. Towards an understanding of the regulatory mechanisms of totipotency. *Curr. Opin. Genet. Dev.* **23**, 512–518 (2013).
- Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
- Wu, G. & Schöler, H. R. Lineage segregation in the totipotent embryo. *Curr. Top. Dev. Biol.* **117**, 301–317 (2016).
- Nichols, J. & Smith, A. The origin and identity of embryonic stem cells. *Development* **138**, 3–8 (2011).
- Tarkowski, A. K. Experiments on the development of isolated blastomeres of mouse eggs. *Nature* **184**, 1286–1287 (1959).
- Tarkowski, A. K. & Wróblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *J. Embryol. Exp. Morphol.* **18**, 155–180 (1967).
- Tsunoda, Y. & McLaren, A. Effect of various procedures on the viability of mouse embryos containing half the normal number of blastomeres. *J. Reprod. Fertil.* **69**, 315–322 (1983).
- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Smith, A. G. et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).
- Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
- Schöler, H. R., Hatzopoulos, A. K., Balling, R., Suzuki, N. & Gruss, P. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J.* **8**, 2543–2550 (1989).
- Canham, M. A., Sharov, A. A., Ko, M. S. & Brickman, J. M. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* **8**, e1000379 (2010).
- Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Hayashi, K., de Sousa Lopes, S. M. C., Tang, F., Lao, K. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401 (2008).
- Kalmar, T. et al. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7**, e1000149 (2009).
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918 (2008).
- Torres-Padilla, M. E. & Chambers, I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development* **141**, 2173–2181 (2014).
- Martinez Arias, A. & Brickman, J. M. Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Curr. Opin. Cell Biol.* **23**, 650–656 (2011).
- Morgani, S. M. et al. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep.* **3**, 1945–1957 (2013).
- Marks, H. et al. The transcriptional and epigenomic foundations of ground-state pluripotency. *Cell* **149**, 590–604 (2012).
- Alexandrova, S. et al. Selection and dynamics of embryonic stem cell integration into early mouse embryos. *Development* **143**, 24–34 (2016).
- Martin Gonzalez, J. et al. Embryonic stem cell culture conditions support distinct states associated with different developmental stages and potency. *Stem Cell Rep.* **7**, 177–191 (2016).
- Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Falco, G. et al. *Zscan4*: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.* **307**, 539–550 (2007).
- Bošković, A. et al. Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* **28**, 1042–1047 (2014).
- Ishiyuchi, T. et al. Early-embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
- Grün, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810 (2015).
- Etzrodt, M., Ende, M. & Schroeder, T. Quantitative single-cell approaches to stem cell research. *Cell Stem Cell* **15**, 546–558 (2014).
- Buganim, Y. et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
- Guo, G. et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* **13**, 492–505 (2013).
- Leitch, H. G. et al. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
- Ficz, G. et al. FGF signaling inhibition in ES cells drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
- Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
- Zalzman, M. et al. *Zscan4* regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
- Amano, T. et al. *Zscan4* restores the developmental potency of embryonic stem cells. *Nat. Commun.* **4**, 1966 (2013).
- Hirata, T. et al. *Zscan4* transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. *Sci. Rep.* **2**, 208 (2012).
- Eckersley-Maslin, M. A. et al. MERVL–*Zscan4* network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
- Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14**, 357–368 (2013).
- Wray, J. 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.* **13**, 838–845 (2011).
- Fazio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60–p400 as a regulator of embryonic stem cell identity. *Cell* **134**, 162–174 (2008).
- Hisada, K. et al. RYBP represses endogenous retroviruses and preimplantation- and germ-line-specific genes in mouse embryonic stem cells. *Mol. Cell Biol.* **32**, 1139–1149 (2012).

43. Suzuki, A. et al. Loss of MAX results in meiotic entry in mouse embryonic and germline stem cells. *Nat. Commun.* **7**, 11056 (2016).
44. Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525–2534 (2013).
45. Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* **14**, 853–864 (2013).
46. Gao, Z. et al. PCGF homologs, CBX proteins and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **45**, 344–356 (2012).
47. Levine, S. S. et al. The core of the Polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol. Cell Biol.* **22**, 6070–6078 (2002).
48. Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D. M. & Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**, 1132–1136 (2002).
49. Zhao, W. et al. Essential role for Polycomb group protein Pcgf6 in embryonic stem cell maintenance and a noncanonical Polycomb repressive complex 1 (PRC1) integrity. *J. Biol. Chem.* **292**, 2773–2784 (2017).
50. Sánchez, C. et al. Proteomics analysis of Ring1B–Rnf2 interactors identifies a novel complex with the Fbxl10 (Jhdm1B) histone demethylase and the Bcl6-interacting co-repressor. *MCP* **6**, 820–834 (2007).
51. Macfarlan, T. S. et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1 (KDM1A). *Genes Dev.* **25**, 594–607 (2011).
52. Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
53. De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
54. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
55. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
56. Xu, Y. et al. The p400 ATPase regulates nucleosome stability and chromatin ubiquitination during DNA repair. *J. Cell Biol.* **191**, 31–43 (2010).
57. Pradhan, S. K. et al. EP400 deposits H3.3 into promoters and enhancers during gene activation. *Mol. Cell* **61**, 27–38 (2016).
58. Eid, A. & Torres-Padilla, M. E. Characterization of noncanonical Polycomb repressive complex 1 subunits during early mouse embryogenesis. *Epigenetics* **11**, 389–397 (2016).

Acknowledgements

We thank A. Smith (Wellcome Trust/MRC Stem Cell Institute) for providing the knock-in REX1 reporter cell line, M. Ko (Keio University) for the *Zscan4c* promoter plasmid, R. Enriquez-Gasca for providing a classification of MERVLs before publication, D. Reinberg (New York University Langone School of Medicine) for the rabbit antibody to PRDM14, A. Ettinger for time-lapse analysis, C. Ebel, D. Pich, T. Hofer and W. Hammerschmidt for help and access to FACS, the INGEMEM infrastructure for access to the IGBMC high-throughput high-content screening workstation, C. Thibault, F. Recillas-Targa and M. Zurita-Ortega for helpful discussions and A. Burton for critical reading of the manuscript. M.-E.T.-P. acknowledges funding from EpiGeneSys NoE, ERC-Stg 'NuclearPotency' (280840), the EMBO Young Investigator Programme, the Fondation Schlumberger pour l'Education et la Recherche (2016-Torres-Padilla) and the Helmholtz Association. J.M.V. acknowledges funding from the Max Planck Society and Epigenesis NoE. T.I. was a recipient of postdoctoral fellowships from the Uehara Memorial Foundation and the Human Frontier Science Programme (LT000015/2012-1). D.R.-T. was partially supported by a DGECI fellowship (2890/2014) from the National University of Mexico.

Author Contributions

D.R.-T., X.G. and T.I. designed, performed and analyzed experiments. D.R.-T. performed most of the computational analyses. A.W. performed the screen together with X.G., under the supervision of L.B. A.K. implemented the screening analysis pipeline with L.B. K.K. performed bioinformatics analysis under the supervision of J.M.V. A.P. performed experiments for screen validation. M.-E.T.-P. designed and supervised the study. All authors contributed to manuscript preparation and read, commented on and approved the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-017-0016-5>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.-E.T.-P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Cell culture. All cell lines in this study, unless otherwise stated, were grown in DMEM supplemented with Glutamax-I, 15% FCS, $2 \times$ LIF, 0.1 mM 2-mercaptoethanol, non-essential amino acids, penicillin and streptomycin over a bed of feeder cells. For the LIF withdrawal experiment, LIF and feeder cells were omitted from the culture conditions. Medium supplemented with 2i ($3 \mu\text{M}$ CHIR99021 and $1 \mu\text{M}$ PD0325901, Miltenyi Biotec) was used for the establishment of stable cell lines and for their expansion and maintenance. After removal of 2i, cells were always cultured for at least 7 d in serum and LIF conditions over a bed of feeder cells before being used for experiments (except for siRNA transfection).

siRNA transfection. Two days before transfection, cells were plated in gelatin-coated dishes. The 2i inhibitors were removed from the medium 1 d before transfection. Lipofectamine RNAiMAX (Life Technologies) was used for siRNA transfection according to the manufacturer's instructions. 75,000 cells were reverse transfected in 24-well gelatin-coated plates using a final concentration of 30 nM for each siRNA condition (the siRNAs are listed in Supplementary Table 5). We used Silenced Negative Control No.1 siRNA (Life Technologies) as a negative control for siRNA treatment. The effect of RNAi was examined 2 d after transfection. The effects of *Pcgf6*, *Ring1b*, *Ep400*, *Dmap1*, *Mga*, *Max* and *Rybp* siRNAs on 2-cell-like cell induction were validated by FACS, immunofluorescence and RT-qPCR (qPCR primers are listed in Supplementary Table 6) with an additional second siRNA sequence available upon request (data not shown). The effects of treatment with *Snrpd1*- and *Lsm6*-specific siRNAs on 2-cell-like cell induction observed in the screen were validated by FACS, immunofluorescence and RT-qPCR (data not shown). However, because of the high cell toxicity observed after treatment with siRNAs targeting transcripts encoding spliceosome components (Supplementary Fig. 5d), we did not focus on these hits for the remainder of the work. Lipofectamine 2000 (Life Technologies) was used for cotransfection with siRNA and the *Nanog* expression vector³⁵, according to the manufacturer's instructions. Cells were analyzed by RT-qPCR 2 d after transfection.

RNAi screening. Screening was performed in the high-throughput screening facility of the IGBMC, using a custom siRNA library (Dharmacon) of chromatin-factor-related siRNAs, the siGenome smartpool (four different siRNAs/pool). Controls were performed with smartpool siRNAs from Dharmacon. Transfection efficiency was optimized using a cell death siRNA, which triggers cell death when transfected into cells (transfection efficiency quantified by assessing toxicity), and the *Chaf1a*-specific siRNA (transfection efficiency quantified by measuring the induction of 2-cell-like cells). For each target, 20 nM (final concentration) of siRNA was reverse transfected in triplicate into 5,000 mouse ES cells by using INTERFERin-HTS (Polyplus-Transfection). For the primary screen, 1,167 genes were targeted (Supplementary Table 4). The screen was performed in 96-well microplates coated with gelatin. 2 d after siRNA transfection, cells were fixed with 1.5% paraformaldehyde, permeabilized with 0.1% Triton X-100, blocked with 2% BSA, and incubated with antibodies to EGFP, ZSCAN4 or OCT4 followed by incubation with Alexa-Fluor-647-, Alexa-Fluor-555- or Alexa-Fluor-488-conjugated secondary antibodies (Invitrogen). Cell nuclei were counterstained with $1 \mu\text{g}/\text{ml}$ DAPI. Screening was achieved on a Tecan Freedom EVO 150 (for cell transfection, staining and immunocytochemistry) and an Orbitor RS Microplate Mover robotic arm coupling microplate stacks to a Cellomics CellInsight NXT High-Content Screening Platform (Thermo Scientific). Images were acquired with CellInsight NXT (Thermo Scientific) and analyzed with HCS Studio Cell Analysis Software (nuclear segmentation and EGFP, ZSCAN4 and OCT4 intensities). Quantification of positive cells for each of these factors was done automatically, based on nuclear segmentation and across >5 fields (for actual cell counts, see Supplementary Tables 7 and 8). The percentage of cells positive for EGFP, ZSCAN4 or OCT4 staining was quantified for each well. Two-cell-like cells were defined as cells that were positive for EGFP and ZSCAN4 staining but negative for OCT4 staining.

To validate selected targets from the siRNA screen, a secondary screen was performed with individual siRNAs. In the secondary screen, the expression of 81 genes from the primary screen was assessed by transfecting four different individual siRNAs, in triplicate, for each target. The secondary screen showed high reproducibility with the primary screen (Supplementary Fig. 7a,b). In addition, 32 new genes were investigated in the secondary screen by transfecting with a pool of four siRNAs per target, as were 9 genes present in the primary screen as internal controls (Supplementary Table 4). The methods used for the secondary screen were as described for the primary screen. For each condition, z scores were calculated as $\text{mean} \left(\frac{x_i - \bar{x}}{s} \right)$, where x_i is the value for each triplicate experiment and \bar{x} and s are the mean value and the s.d. for the negative-control conditions, respectively.

Reporter cell lines. EGFP and turboGFP 2C reporter cell lines have previously been described²⁷. To generate the *Zscan4* reporter cell line, we replaced the emerald cassette of the reporter plasmid (kindly provided by M. Ko)³⁵ with a cassette encoding nuclear localization sequence (NLS)-tagged destabilized tdTomato or mCherry. The 2C::turboGFP reporter cell line was transfected with the respective plasmids using Lipofectamine 2000, and a single clone was then selected. In the case of the *Rex1* and *Zscan4* reporter, a stable cell line carrying an EGFP

cassette knocked into the ORF of *Rex1* was kindly provided by the laboratory of A. Smith⁴⁰, and these cells were transfected with the *Zscan4c*::tdTomato reporter construct. The reporter cell lines used in this study are summarized in Supplementary Table 10.

Analysis of RNA-seq data and selection of genes for the single-cell expression profiling experiments. RNA-seq data for 2-cell-like cells were reported previously²⁷. Genes used in the Biomark single-cell analysis were selected on the basis of their functional significance and differential expression from our previous bulk RNA-seq analysis. Chimeric genes were defined on the basis of transcription starting from an LTR belonging to the MT2 families by genome browser analysis of the RNA-seq analysis in Ishiuchi et al.²⁷. Heat maps of these transcriptomic data were generated using the ggplot2 R package.

Validation of TaqMan assays and custom designs. TaqMan assays (Thermo Fisher Scientific) were first tested on three different fivefold serial dilutions of cDNA from turboGFP⁺ cells on a LightCycler qPCR instrument. TaqMan assays that failed to amplify or that did not exhibit linearity in their measurements when the different dilutions were compared were omitted from the single-cell analysis. All TaqMan assays used are described in Supplementary Table 1. The TaqMan assay used for *Zscan4* amplifies *Zscan4c*, *Zscan4d* and *Zscan4f*. Custom primers were designed to target the turboGFP reporter and were mixed to a final concentration equal to that of the components in the TaqMan assays ($18 \mu\text{M}$ for the primer and $4 \mu\text{M}$ for the probe).

Single-cell expression profiling. After thawing, reporter cell lines were cultured for 6 d in serum- and LIF-containing medium over a bed of feeders and passaged every day, except for the second day of culture. On the sixth day of culture, cells were sorted with the help of a FACS machine, and only turboGFP⁺ cells were replated; 2-cell-like cells were discarded. 24 h later, cells were sorted once again, but this time the ES, 2-cell-like or *Zscan4*⁺ cell fraction was preserved and prepared according to the manufacturer's protocol (Fluidigm, PN1006117) for use in Fluidigm's C1 microfluidics-based single-cell sample preparation platform. Of note, $1 \mu\text{l}$ of a 1:286 dilution of ERCC spike-in mix was added to the lysis solution of each C1 run instead of water. Single-cell expression data were generated in Fluidigm's Biomark qPCR platform in technical duplicates using TaqMan probes (Thermo Fisher Scientific) according to the manufacturer's protocol (Fluidigm, PN68000130). Only cells that passed the quality control check were included in the analyses performed throughout the manuscript. In total, we profiled 189 cells across four biological replicates. A constitutive pCAG promoter driving NLS-tdTomato expression allowed us to sort feeder cells out (Fig. 1b and Supplementary Fig. 1c). In addition, we used two spike-in controls, which allowed us to assess the quality of the normalization to endogenous control genes and therefore to constrain technical noise (Supplementary Fig. 1d).

Normalization, quality control and modeling of single-cell data. Ct values obtained from the Biomark platform were processed as described previously⁴⁰. Briefly, Ct values higher than 28 or with quality scores lower than the threshold of 0.65 were substituted with Ct values of 28. Subsequently, Ct values were subtracted from a baseline value of 28, such that 0 implied no expression and 28 implied a high level of expression. For normalization purposes, the average of the *Actb* and *Gapdh* Ct values was subtracted from the values for all other genes for the same cell and the minimum dataset-wide value was subtracted obtain positive values in all assays and samples. The average of the Ct values for the two technical replicates for the same cell was used. Note that, unlike in RNA-seq approaches, higher levels of expression of specific genes relative to others is unlikely to bias gene expression data because the usage of a pool of primers specific for each of the assayed genes during preamplification provides additional robustness to a single highly expressed gene being preferentially amplified and biasing a cell's transcriptional profile. For PCA, the principal components of the dataset were computed using the svd method in R, and no scaling was performed. For projection of the *Zscan4* dataset into the principal components of the ES and 2-cell-like dataset, principal components were calculated using the svd method for the ES and 2-cell-like dataset, and its loadings matrix and the *Zscan4* dataset's matrix were then multiplied. All plotting was done in R with the ggplot2 package. To classify cells into the ES, *Zscan4*^{low}, *Zscan4*^{mid}, *Zscan4*^{high} and 2-cell-like cell categories, we first classified them into two groups based on whether they expressed *Zscan4c/d/f*. Cells that did not express *Zscan4* were termed ES cells. The remaining cells were subsequently classified based on whether the turboGFP transcript level exceeded the established cutoff, which was determined with the help of the density function of turboGFP transcript expression. Cells with expression values above the cutoff were classified as 2-cell-like cells. Note that this threshold had to be defined because the turboGFP reporter is intronless and therefore has to be distinguished from the genomic background (Fig. 1c, threshold selection not shown). The remaining cells were classified into *Zscan4*^{low}, *Zscan4*^{mid} or *Zscan4*^{high} cells according to the thresholds that were set over the density function for *Zscan4* in Supplementary Fig. 3a.

Gapdh and *Actb* were used as internal controls for normalization, as they reflect independent molecular pathways and are unrelated to pluripotency.

Indeed, their expression was stable across samples and showed a high correlation. The Ct values for ES cells (GFP⁻) and 2-cell-like cells (GFP⁺) were consistently similar for the spike-in RNAs after normalization (Supplementary Fig. 1c), supporting the robustness of the normalization. Once data processing was performed, all cells expressing outlier values for *Actb*, *Gapdh* or the spike-in RNAs were removed from the analysis.

Flow cytometry. Cells were washed with room-temperature sterile PBS, trypsinized and resuspended in ice-cold sterile 0.5% BSA–PBS solution. Sorting was performed using a BD Biosciences FACS Aria II or III. During sorting, cells were collected in culture medium and kept at 4 °C during the sort. Analysis of FACS data was performed using FlowJo software, and the same gating parameters were used for all replicates of the same experiment. Cells were not index sorted, and the purity of the sort was estimated at 96% for 2-cell-like cells and at 97% for ES cells. A control flow profile for wild-type (WT) ES cells is shown in Supplementary Fig. 1c. For the REX1 experiments, the REX1⁻ gate was defined based on the fluorescence of WT ES cells and the REX1^{high} gates were defined based on the fluorescence of *Rex1::EGFP*⁺ cells cultured in 2i medium. A FACSCalibur (BD Biosciences) was used to quantify the population of EGFP⁺ cells. Cells that had been frozen in 2i medium were thawed, and the REX1-based sorting was performed 4 d after 2i withdrawal. For data presented in Fig. 7a, the 2C::turboGFP and *Zscan4c::mCherry* cell line was FACS sorted just before transfection, and the 2-cell-like cells and *Zscan4c::mCherry*⁺ cells were removed, from the population.

Immunofluorescence, image processing and quantification. Cells were cultured over feeder-coated coverslips, fixed in paraformaldehyde for 10 min at room temperature and permeabilized with 0.2% Triton X-100 for another 10 min. Cells were incubated with primary antibodies overnight followed by three washes in PBS. Cells were then incubated with secondary antibodies for 1 h. Mounting was done in Vectashield mounting medium (Vector Labs). Image acquisition was performed using a Leica SP8 confocal microscope. For quantification of immunofluorescence, manual segmentation of the cell nucleus was performed on the DAPI channel using ImageJ, and the average fluorescence intensity was measured. Afterward, the average fluorescence intensity was measured. Only ES cells, *Zscan4*⁺ cells and 2-cell-like cells from the same coverslip and imaging session were used for each comparison. Density plots for each of these groups were computed using the kernel density estimation function in R with the Sheather and Jones bandwidth-selection method. For 2C::EGFP immunostaining, cells were fixed as described and blocked for 30 min at 37 °C in 10% FCS, 3% BSA and 0.1% Triton X-100 in PBS (blocking buffer). Cells were incubated with primary and secondary antibodies 30 min each at 37 °C in the blocking buffer solution.

Real-time PCR. Total RNA was extracted from ES cells with the RNeasy Plus mini kit (Qiagen) and treated with turbo DNase (Life Technologies) to remove genomic DNA. Reverse transcription was performed with SuperScript II (Life Technologies) using random hexamers. Real-time PCR was performed with LightCycler 480 SYBR Green I Master Mix (Roche) on a LightCycler 96 Real-Time PCR System (Roche). The relative expression level of each gene was normalized to those of *Gapdh* and *Actb*. The primers used in this study are listed in Supplementary Table 6.

Antibodies. Antibodies used in this work were the following: mouse antibody to turboGFP (OTI2H8, Origene), rabbit antibody to turboGFP (PA5-22688, Thermo Fisher), rabbit antibody to ZSCAN4 (AB4340, EMD Millipore), rabbit antibody to MuERV1–Gag (R1501-2, Hangzhou HuaAn Biotechnology), chicken antibody to EGFP (ab13970, Abcam), rabbit antibody to H2AK119Ub (8240, Cell Signaling), goat antibody to REX1 (sc-50670, Santa Cruz), goat antibody to OCT4 (sc-8628, Santa Cruz), mouse antibody to OCT4 (611203, BD Biosciences), goat antibody to SOX2 (sc-17320, Santa Cruz), rabbit antibody to PRDM14 (a gift from D. Reinberg)⁶⁵, rabbit antibody to TFAP2C (sc-8977, Santa Cruz).

Time-lapse experiments. Prior to time-lapse analysis, cells were cultured overnight on glass-bottom laminin-coated (Sigma-Aldrich) Ibidi micro-Insert cell culture dishes to allow the cells to attach. Image acquisition was carried out over the entire well with a 20×0.75 NA Plan-Apochromat objective lens every 30 min for 48 h using a Nikon Ti-E system equipped with the Bruker Opterra II multipoint confocal system. Images were recorded on an EMCCD camera using emission filters for turboGFP (BP520/40), mCherry (570LP) and iRFP (655LP) mounted on a FLI filter wheel. Spontaneously arising 2-cell-like cells were manually identified using ImageJ software and scored based on whether they arose from a *Zscan4*⁺ cell as determined by the intensity of the mCherry channel. This analysis was run independently by two different people, and results were cross-compared for accuracy.

ATAC-seq data analysis. Previously published ATAC-seq data in ES, *Zscan4*⁺ and 2-cell-like cells were obtained from GEO accession GSE75751³⁸. Paired-end reads were trimmed for adaptor sequences using trimmomatic 0.36 and mapped to the mm10 reference genome using bowtie2 with parameters --dovetail -X 2000

--no-discordant --no-mixed. The resulting bam files were then filtered for non-uniquely mapping reads using SAMtools with a MAPQ threshold of 10 and filtered for duplicates using MarkDuplicates in Picard Tools. Finally, mitochondrial reads were removed using SAMtools, and signal tracks were generated using macs2.1.1⁶³ with parameters --SPMR --nomodel --nolambda=shift --100 --extsize 200 for the combined reads of all replicates of the same population. For the MERVL/MT2_Mm analysis in Fig. 2g,h, MT2_Mm coordinates were obtained from RepeatMasker release 20140131, and plots were generated using deepTools. For clarity, only solo LTRs are shown in Fig. 2h, but the same pattern of enrichment was also observed for full-length MERVLs. For the differential accessibility analysis in Supplementary Fig. 2h, regions of differential accessibility between ES and 2-cell-like cells were called using the bdgdiff macs2 command, and plots were generated using deepTools.

Single-cell RNA-seq data analysis. Single-cell RNA-seq data for ES, *Zscan4*⁺ and 2-cell-like cells were obtained from ArrayExpress accession E-MTAB-5058³⁸. Paired-end reads were trimmed for adaptor sequences using trimmomatic 0.36 and mapped to the mm10 reference genome using STAR 2.5.3a⁶⁴. The resulting bam files were then filtered for unmapped reads and secondary alignments using SAMtools. Finally, reads intersecting repetitive elements were quantified by intersecting the RepeatMasker annotation with the aligned bam files using bedtools intersect with the --split parameter. To classify the individual cells in a comparable manner to the classification performed on the single-cell qPCR data, the counts of uniquely mapping reads that mapped to the full-length *Zscan4* isoforms *Zscan4c*, *Zscan4d* and *Zscan4f* were summed. Note that the *Zscan4* TaqMan assay used for the single-cell qPCR experiments could not discriminate between the *Zscan4c*, *Zscan4d* and *Zscan4f* isoforms, and hence we decided to pool the read counts for these isoforms. The following thresholds were used to classify individual cells: cells with log₂(MT2_Mm, CPM + 1) > 12.25 were considered 2-cell-like cells; cells with log₂(*Zscan4c*, *Zscan4d* and *Zscan4f*, CPM + 1) > 24 were considered *Zscan4*^{high}; cells with log₂(*Zscan4c*, *Zscan4d* and *Zscan4f*, CPM + 1) > 12 were considered *Zscan4*^{mid}; cells with log₂(*Zscan4c*, *Zscan4d* and *Zscan4f*, CPM + 1) > 2 were considered *Zscan4*^{low}; all remaining cells were considered ES cells. The positioning of these thresholds in the context of the population-wide distribution of expression levels is shown in Supplementary Fig. 3d. For the heat map shown in Fig. 2g, 2,313 upregulated and 951 downregulated genes were selected based on the 2-cell-like and ES cell RNA-seq data reported in Ishiuchi et al.²⁷. The mean expression level for each gene within each respective group of single-cell transcriptomes was taken, and the row-wise z score was calculated. Plotting was performed using ggplot2. The differential expression analysis in Supplementary Fig. 3f was performed using DESeq⁶⁵ by considering each individual cell as a replicate. Changes were considered statistically significant when *P* (adj.) < 0.05.

ChIP-seq data analysis. ChIP-seq data for H3K4me3 (GSE74112)⁶⁶, EP400 (GSE64825)⁶⁷, H2AK119Ub (GSE89949)⁶⁸, RING1B (GSE40860)⁶⁹, RYBP (GSE42466)⁷⁰ and MAX (GSE48175)⁷¹ from ES cells grown with serum- and LIF-containing medium were obtained from GEO. Note that none of these datasets were generated in the presence of the 2i inhibitors. Single-end reads were trimmed for adaptor sequences and low-quality bases using trimmomatic 0.36 and mapped to the mm10 reference genome using bowtie with parameters -S --best --strata -v 2 -M 1. The resulting bam files were then filtered for non-uniquely mapping reads using SAMtools with a MAPQ threshold of 10 and filtered for duplicates using MarkDuplicates in Picard Tools. Finally, fragment size was estimated using MaSC, and sequencing-depth-normalized signal tracks were generated using macs2.1.1 with the MaSC-estimated⁷² fragment size. Biomart was used to obtain 10-kb-wide genomic windows centered on all TSSs for the 2,313 upregulated (6,829 TSSs) and 951 downregulated (4,189 TSSs) genes in 2-cell-like cells relative to ES cells, as reported²⁷. Mappability tracks were generated using the GEM suite with a 36-bp window. The heat maps shown in Fig. 8a,c were plotted using deepTools⁷³ and custom scripts. A TSS was considered to be associated with MT2_Mm if an MT2_Mm copy was found within 50 kb upstream of it. For the enrichment calculation of given chromatin modifiers, peaks were called using macs2.1.1 with a *q*-value cutoff of 10⁻⁶ and peaks within 1 kb of each other were merged. Genes with at least one of their TSSs occurring within 5 kb of a ChIP-seq peak were considered to be bound by the respective factor. Fisher's exact tests for these overlaps were performed using R.

ChIP-seq enrichment at MERVL and MT2_mm. Input and treatment reads were mapped individually to the mm10 reference genome using bowtie⁷⁴ (1.1.1) with the following options: -k 201 --best --strata --m 200, allowing for up to 200 alignments of the highest quality to be reported. macs⁶³ (2.1.1) predict⁷⁵ -g mm was used to estimate the fragment size of the uniquely mapping treatment reads. Because repetitive regions were the focus of this analysis, CSEM⁷⁵ (2.4) was used to assign multi-mapping reads to their most likely origin. It was run with the option --upper-bound 500, and fragment size was estimated as described above. For each multi-mapping read, only the most likely alignment was selected. If multiple equally likely alignments existed, one was selected at random. macs bdgcmp was used to generate fold enrichment tracks from these input and

treatment alignments. Mappability tracks were generated by splitting the mm10 genome into overlapping reads of 36 bp in length, starting at each base, and aligning them back to the reference genome using bowtie²⁴ (1.1.1). A base was 'mappable' if it could be aligned to the reference genome uniquely; otherwise, it was 'unmappable'. To avoid mappability issues, for the transposable elements MERVL and MT2_Mm, enrichment *P* values were calculated using a permutation test. First, the enrichment profile was calculated as the mean of all rows in the enrichment heat map. Then, the maximum value of the enrichment profile e_{\max} was recorded around windows 1 kb upstream and 1 kb downstream of the transposable element. A permuted set of regions was then generated by selecting random genomic locations that matched the number, size and chromosome distribution of the original elements. As before, the maximum value of the enrichment profile in the random set r_{\max} was recorded. The permutation *P* value was then calculated as $P = e_{\max} / \sum_i r_{i\max}$, where *i* denotes the *i*th permutation. For every *P* value, 10,000 randomizations were performed. For the analysis, MERVLs were classified into complete MERVL elements according to the presence of contiguous MT2_Mm LTRs within 7 kb with the same 5'-to-3' orientation interspersed with an internal MERVL element as annotated by RepeatMasker (v4.0.3). MT2_Mm LTRs without a corresponding MERLV internal part or a nearby (<7 kb) LTR pair in the same orientation were classified as solo LTRs. Occurrences of MERVL internal elements without matching MT2_Mm LTRs flanking these elements were classified as internal MERVLs. The *P* values for enrichment for the following factors for the MERVL ChIP-seq analyses are as follows: for MAX: 0.0001, 1.0 and 0.9801 for complete MERVL, MT2_Mm and MERVL-int, respectively; for RYBP: 0.0537, 0.9610 and 0.5597 for complete MERVL, MT2_Mm and MERVL-int, respectively; and for H2AK119Ub (data not shown): 0.6739, 1.0 and 1.0000 for complete MERVL, MT2_Mm and MERVL-int, respectively.

Statistical analyses. Statistical tests were performed keeping in mind data distribution and the number of data points available. Details on sample sizes, in addition to the statistical tests conducted, are presented in the corresponding figure legends. Asterisks indicate *P* values below the 0.05 threshold.

Primers and TaqMan assays in this study. The list of TaqMan assays is in Supplementary Table 1, and the list of primers used is in Supplementary Table 6.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The data supporting this study are available from the authors upon reasonable request.

References

59. Miyanari, Y. & Torres-Padilla, M. E. Control of ground-state pluripotency by allelic regulation of *Nanog*. *Nature* **483**, 470–473 (2012).
60. Guo, G. et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
61. Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods—a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164–1167 (2007).
62. Burton, A. et al. Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. *Cell Rep.* **5**, 687–701 (2013).
63. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
64. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
65. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
66. Liu, Z. & Kraus, W. L. Catalytic-independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. *Mol. Cell* **65**, 589–603 (2017).
67. de Dieuleveult, M. et al. Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113–116 (2016).
68. Kundu, S. et al. Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell* **65**, 432–446 (2017).
69. Farcas, A. M. et al. KDM2B links the Polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**, e00205 (2012).
70. Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A. & Di Croce, L. RYBP and Cbx7 define specific biological functions of Polycomb complexes in mouse embryonic stem cells. *Cell Rep.* **3**, 60–69 (2013).
71. Krepelova, A., Neri, F., Maldotti, M., Rapelli, S. & Oliviero, S. Myc and Max genome-wide binding sites analysis links the Myc regulatory network with the Polycomb and the core pluripotency networks in mouse embryonic stem cells. *PLoS One* **9**, e88933 (2014).
72. Ramachandran, P., Palidwor, G. A., Porter, C. J. & Perkins, T. J. MaSC: mappability-sensitive cross-correlation for estimating mean fragment length of single-end short-read sequencing data. *Bioinformatics* **29**, 444–450 (2013).
73. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
74. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
75. Chung, D. et al. Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. *PLoS Comput. Biol.* **7**, e1002111 (2011).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

▶ Experimental design

1. Sample size

Describe how sample size was determined.

Sample size was chosen in order to ensure that it will be appropriate for statistical analysis. See Figures legends for each experiment.

2. Data exclusions

Describe any data exclusions.

See Normalization, quality control and modeling of single cell data in Online Methods.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All data was replicated at least twice and replicate number indicated in the respective panel.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

na

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

na

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted
- A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

R, excel

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* [guidance for providing algorithms and software for publication](#) may be useful for any submission.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

All antibodies used in the study are listed in the online method section: Antibodies.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

All the reporter cell lines used in the paper are listed in the online method section: reporter cell lines.

b. Describe the method of cell line authentication used.

Reporter cell lines are authenticated by immunofluorescence, transcriptional profiling.

c. Report whether the cell lines were tested for mycoplasma contamination.

Cells are frequently tested for mycoplasma contamination using MycoAlert Mycoplasma detection kit from Lonza.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

na

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

na

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

na

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

▶ Data presentation

For all flow cytometry data, confirm that:

- 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- 3. All plots are contour plots with outliers or pseudocolor plots.
- 4. A numerical value for number of cells or percentage (with statistics) is provided.





▶ Methodological details

- | | |
|--|---|
| 5. Describe the sample preparation. | Mouse ES cells were washed with PBS, trypsinized and resuspended in 3% BSA PBS |
| 6. Identify the instrument used for data collection. | FACS Aria II or IIIu or FACS Calibur |
| 7. Describe the software used to collect and analyze the flow cytometry data. | FlowJo |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | Whenever cell numbers were not an issue, fluorescence was verified after sorting and was usually 95-100%. Downstream experiments always confirmed a very high degree of sorting purity. |
| 9. Describe the gating strategy used. | Stringent gatings were always used, leaving a significant gap in between negative/postiive or low/high populations. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

In the format provided by the authors and unedited.

A molecular roadmap for the emergence of early-embryonic-like cells in culture

Diego Rodriguez-Terrones ^{1,2}, Xavier Gaume^{1,2}, Takashi Ishiuchi^{1,5}, Amélie Weiss², Arnaud Kopp², Kai Kruse ³, Audrey Penning¹, Juan M. Vaquerizas ³, Laurent Brino² and Maria-Elena Torres-Padilla ^{1,4*}

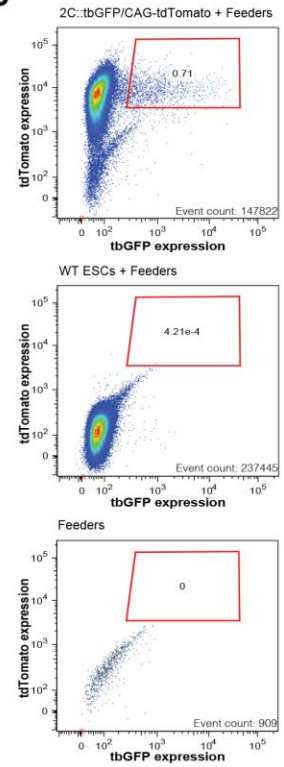
¹Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum München, Munich, Germany. ²Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS-INSERM U964, Strasbourg, France. ³Max Planck Institute for Molecular Biomedicine, Münster, Germany. ⁴Faculty of Biology, Ludwig Maximilians Universität, Munich, Germany. Present address: ⁵Division of Epigenetics and Development, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan. Diego Rodriguez-Terrones and Xavier Gaume contributed equally to this work. *e-mail: torres-padilla@helmholtz-muenchen.de

Supplementary figure 1

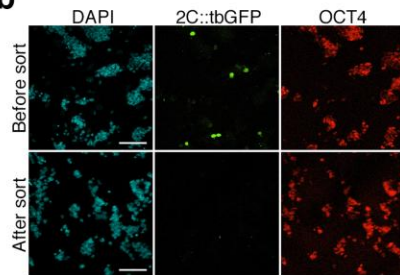
a

Genes analyzed			
Controls and reporters		Signaling pathways	
Endogenous	Gapdh, Actb, Sdhia, Atp5a1	BMP	Id1, Id2, Id3, Smad7, Msx2, Smad6
Spike-ins	ERCC-00130, ERCC-00116	WNT	Axin2
Reporters and markers	Col5a2, turboGFP	Activin/Nodal	Inhbb, Lefty1, Smad7, Foxh1
Embryonic development markers		FGF	Fgf1
Core pluripotency	Pou5f1, Sox2, Nanog	Metabolism	
Naive pluripotency	Zfp42	TCA cycle flux regulation	Pdk1, Pdha2, Ldhd
Germline markers	Dazl, Prdm14, Tfap2c	Hypoxia response	Rwdd3
Trophectoderm	Cdx2	Differentially expressed clusters	
ESC heterogeneity	Hhex, Nanog, Dppa3, Zfp42	Obox	Obox1, Obox3, Obox6
Oocyte/sperm markers	C86695, Oog4, Sytl2, Obox1, Susd3, Elavl3, Cdo1, Spz1	Rhox	Rhox6, Rhox9, Rhox13
Early embryonic markers	Snai1, Bex6, Hspa1b, Pramef7, Obox6, Foxa1, Piger4, Naalad2, Zfp599, Pramef6, Zscan4 c/d/f, Zfp352, Sp110	Chromatin	
Chimaeric genes		Chromatin assembly and accessibility	Cha11a, Cha11b, Nap112, Nap113, Hist1h1t, H1fx, H110, Hmgn3, Hmgn5
MT2_mm	Spz1, Snai1, Sp110	DNA methylation	Dnmt1, Dnmt3a, Dnmt3b, Dnmt3l, Tet1, Tet2, Tet3
MT2B1	Zfp352	H3K9 methylation	Setdb1, Ehmt1, Ehmt2, Suv39h1, Suv39h2, Kdm1a, Kdm4d, Jmjd1c
MT2B2	Bex6	Other regulatory factors	
MT2C	Naalad2	Transcription initiation variants	Taf4b, Taf7l, Tbp1l
Reprogramming resistant regions		Pw1	Pw1l1, Pw1l4, Asz1
Upregulated	Obox3, Zfp599, Pramef6, Zfp939, Zscan4	Other	Trim28, Setd4
Not-upregulated	Imp3, Eid2, Gins2		

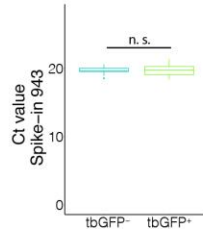
c



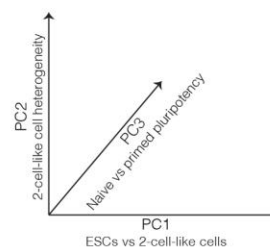
b



d

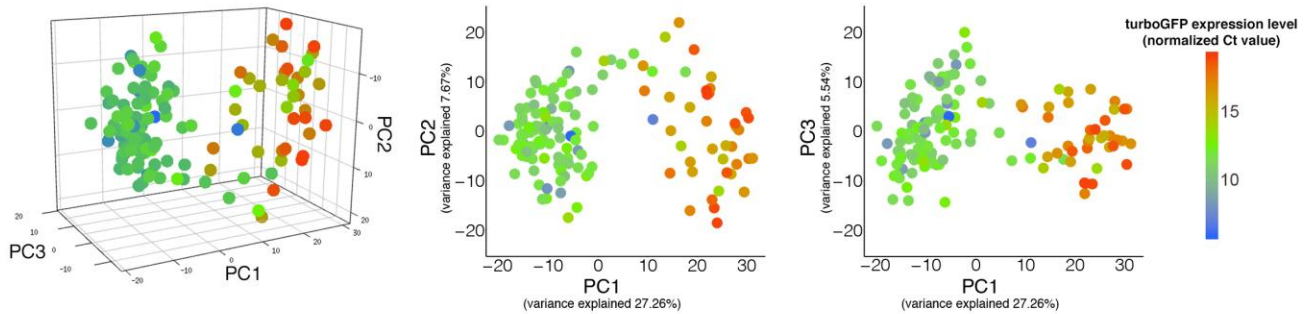


e



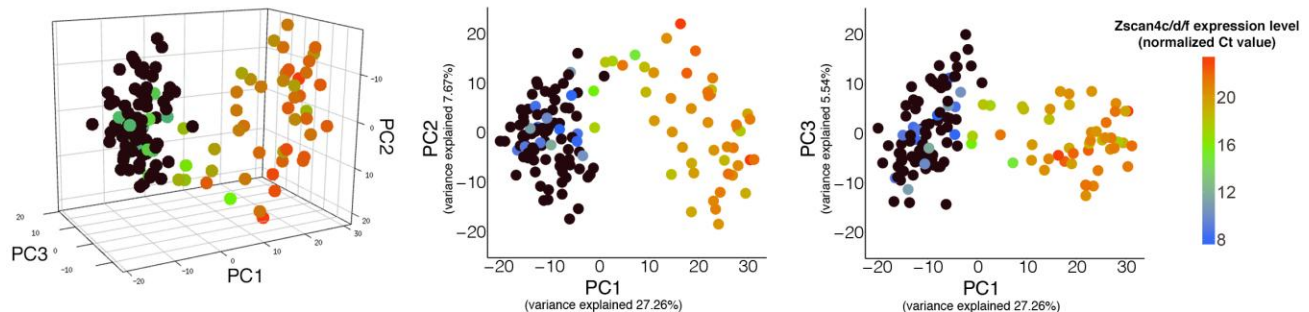
f

PCA computed on the aggregate of the ESC and the 2-cell-like datasets with 2C::tbGFP and Zscan4c/d/f expression omitted from the calculation



g

PCA computed on the aggregate of the ESC and the 2-cell-like datasets with 2C::tbGFP and Zscan4c/d/f expression omitted from the calculation

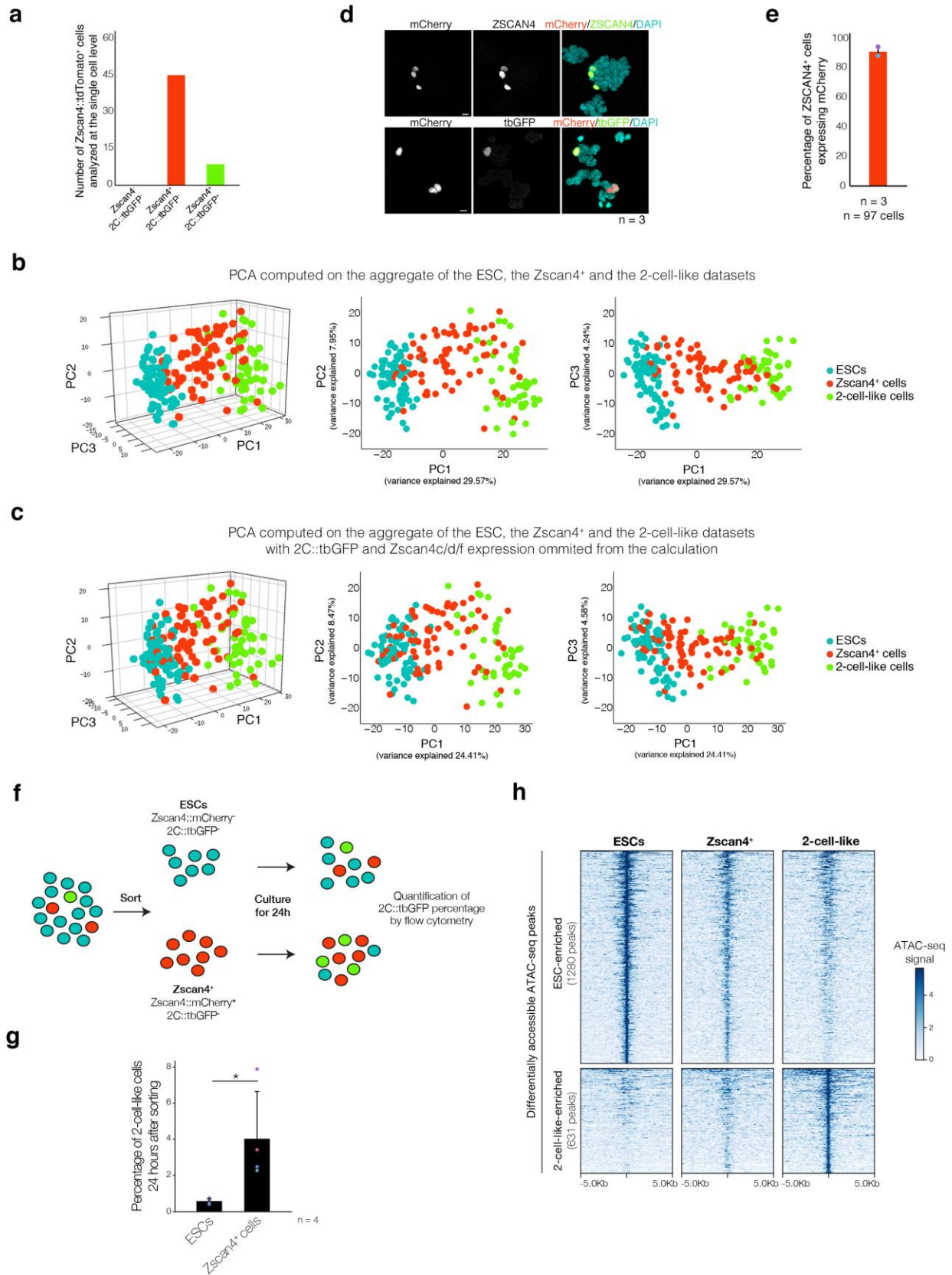


Supplementary Figure 1

Controls for the single-cell expression profiling experiments of ES and 2-cell-like cells

a, List of genes selected for the single-cell analysis classified according to their pathway or function. **b**, Immunofluorescence analysis using a turboGFP and an OCT4 antibody in the 2C::turboGFP cell line before and after sorting out 2-cell-like cells as indicated in Fig. 1b. Scale bar, 100 μm . **c**, Scatterplots of turboGFP fluorescence versus tdTomato fluorescence for feeder cells only (bottom), WT ES cells and feeder cells (middle), and the 2C::turboGFP/CAG-tdTomato reporter line with feeders (top) assayed by FACS. The presence of constitutively expressed NLS-tdTomato in the reporter line allows efficient discrimination from feeder cells. **d**, Normalized Ct values for the ERCC-943 spike-in comparing turboGFP⁻ and turboGFP⁺ cells. Note that the turboGFP⁻ and turboGFP⁺ cells analyzed in these plots come from independent sample preparation experiments but were processed on the same Biomark chip. Because the two groups both exhibit constant expression that is highly similar for ERCC-943, we conclude that they were normalized properly and that their expression levels are therefore comparable. Boxes indicate 25% and 75% quartiles, and the whiskers extend to 1.5 times the interquartile range. **e**, Graphic interpretation of the features contrasted across the first three principal components of the principal-component analyses shown in Figs. 1–3. **f,g**, Different viewpoints of the principal-component analysis of the ES and 2-cell-like single-cell dataset. This PCA was computed without the expression data from *turboGFP* and *Zscan4*. Each point corresponds to a single cell and is color-coded based on the original expression level of *turboGFP* (**f**) or *Zscan4*/d/f (**g**) as indicated on the right. Black dots indicate no expression.

Supplementary figure 2

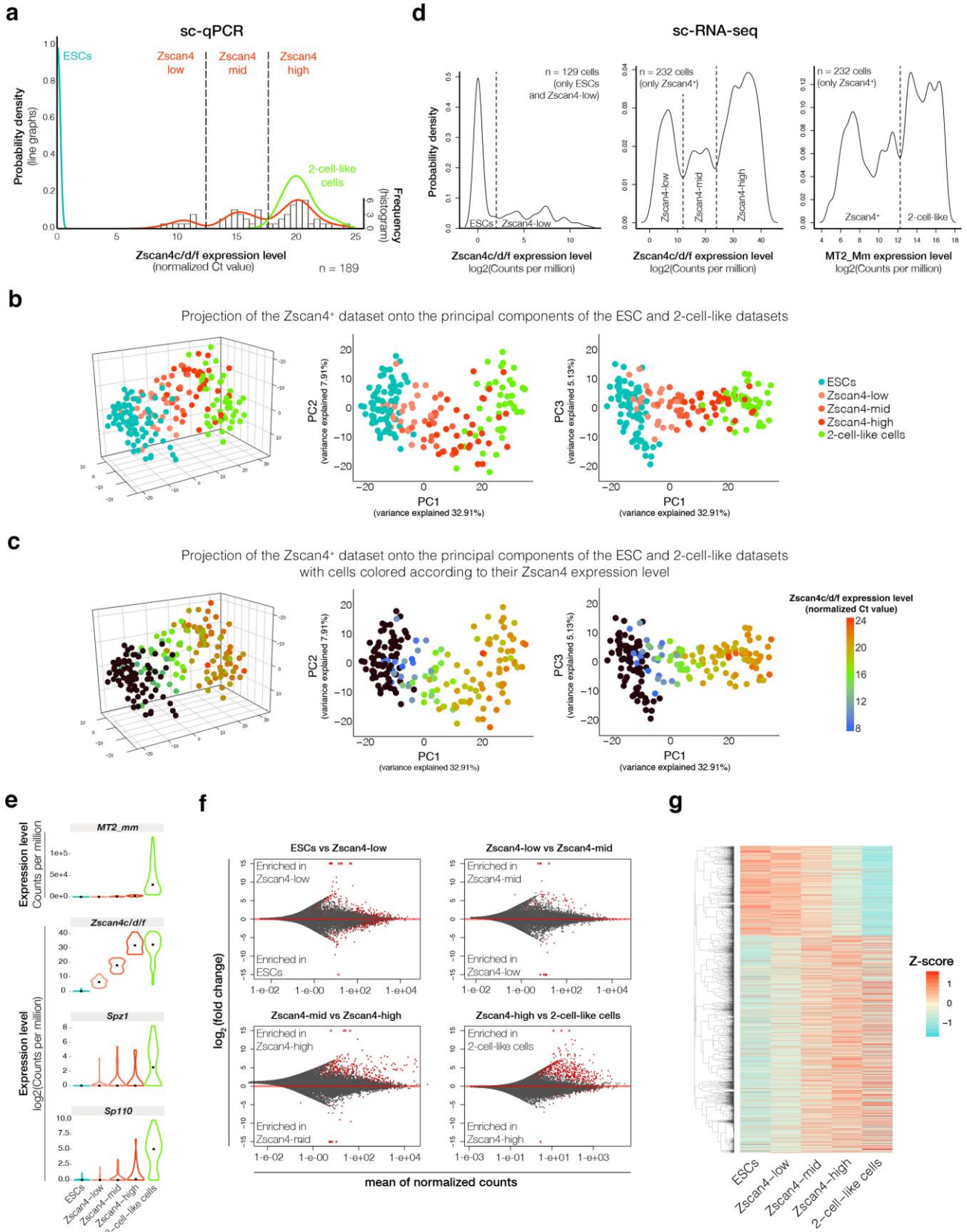


Supplementary Figure 2

Zscan4 cells are an intermediate cellular state between the ES and 2-cell-like states

a, Accuracy of the *Zscan4c*::tdTomato reporter cell line used for the single-cell profiling described in Fig. 2. The graph shows the number of tdTomato⁺ cells that scored positive as assessed by FACS in relation to whether they belong to ES cells (no *Zscan4c/d/f* transcripts detected), *Zscan4*⁺ cells (*Zscan4c/d/f* transcripts detected) or 2-cell-like cells (*Zscan4c/d/f* and *turboGFP* transcripts detected). **b,c**, Principal-component projection of all datasets combined. Principal components were calculated for the aggregate of the ES, *Zscan4*⁺ and 2-cell-like datasets (Figs. 1 and 2), unlike the analyses in Figs. 2 and 3 where the *Zscan4* dataset (from Fig. 2) was projected onto the principal components of the ES and 2-cell-like datasets. In **b**, *turboGFP* and *Zscan4c/d/f* were omitted from the calculation of the principal components. **d,e**, Validation of the *Zscan4c*::tdTomato and 2C::*turboGFP* cell line used for the time-lapse analysis in Fig. 2e. **d**, Representative immunostaining for mCherry and ZSCAN4 (top) and mCherry and turboGFP (bottom) from three independent cell cultures. **e**, Quantification of the percentage of (endogenous) ZSCAN4⁺ cells that also express mCherry. The reporter recapitulates endogenous expression of ZSCAN4 protein with ~92% accuracy. Error bars, s.d. Scale bar, 10 μm. **f,g**, *Zscan4*⁺ and *Zscan4*⁺ cells were FACS sorted based on the *Zscan4*::mCherry reporter and cultured for 24 h, after which the percentage of turboGFP⁺ cells was quantified by FACS. Shown are the means ± s.d. of four independent experiments. During the 24-h window, 4% of the *Zscan4*⁺ cell population became 2C-like cells, 63% remained *Zscan4*⁺ cells and 33% lost *Zscan4* reporter expression. **h**, Heat maps showing ATAC-seq signal intensity over 1,911 genomic regions with different accessibility in ES and 2-cell-like cells.

Supplementary figure 3



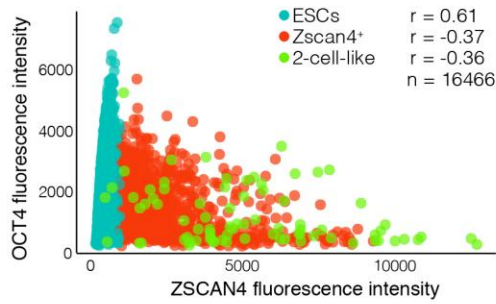
Supplementary Figure 3

Gradual transcriptional changes accompany *Zscan4* upregulation and precede entry to the 2-cell-like state

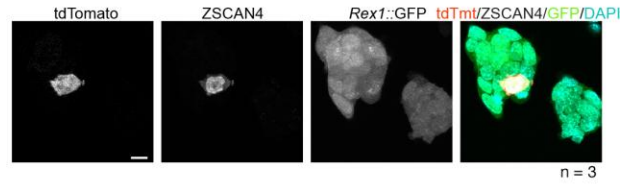
a, The graph combines two parameters: the line (left y axis) depicts probability density and the histogram under it (right y axis) refers to absolute frequency of occurrence. The probability density function of *Zscan4c/d/f* expression in ES (blue), *Zscan4*⁺ (orange) and 2-cell-like (green) cells is plotted against the normalized expression of *Zscan4c/d/f* (*x* axis) in each individual cell. These three distinct levels were classified as low, mid and high based on the histogram data, which derive from the Biomark analysis. **b,c**, Projection of the expression profiles of *Zscan4*⁺ cells onto the principal components of the ES and 2-cell-like cell dataset (Fig. 1d). Each dot represents a single cell and is color-coded according to whether it corresponds to an ES cell, a *Zscan4*^{low}, *Zscan4*^{mid} or *Zscan4*^{high} cell, or a 2-cell-like cell according to the legend on the right. In **c**, cells are colored based on their expression levels of *Zscan4c/d/f* as indicated on the right. Black indicates no expression. **d**, Density plots for *Zscan4c/d/f* and MT2_Mm based on single-cell RNA-seq data³⁹. Dotted lines represent the thresholds used to classify individual cells into ES cells, *Zscan4*^{low}, *Zscan4*^{mid} or *Zscan4*^{high} cells, and 2-cell-like cells. **e**, Violin plots for the MT2_Mm LTR, *Zscan4c/d/f* and two MERVL-driven chimeric genes in the single-cell RNA-seq dataset. **f**, MA plots showing significantly differentially expressed genes (red) for each transitional state analyzed from single-cell RNA-seq data. The list of differentially expressed genes for each transition is shown in Supplementary Table 9. **g**, Heat map showing a gradual transition in the expression profiles of cells transitioning between ES cells and 2-cell-like cells based on single-cell RNA-seq data.

Supplementary figure 4

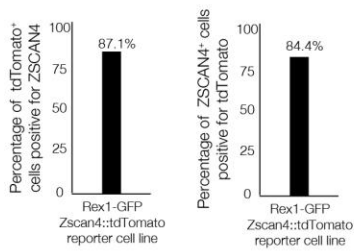
a



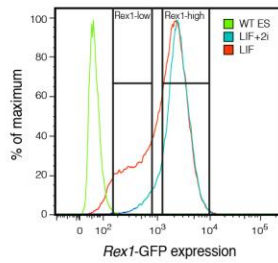
b



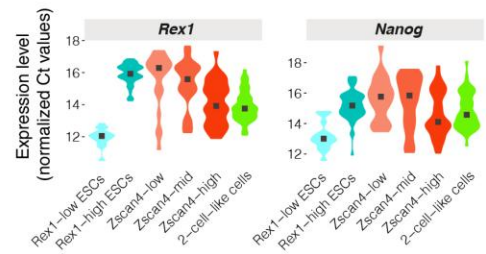
c



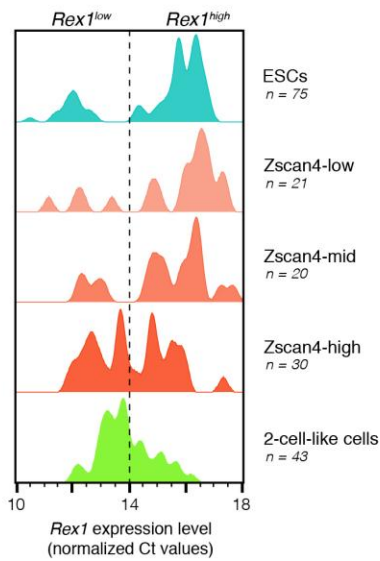
d



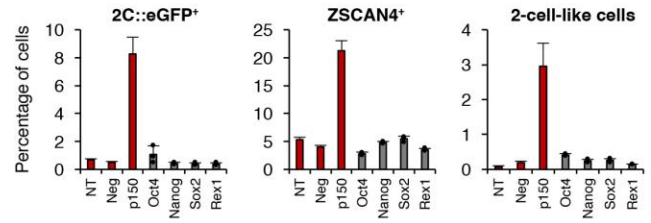
e



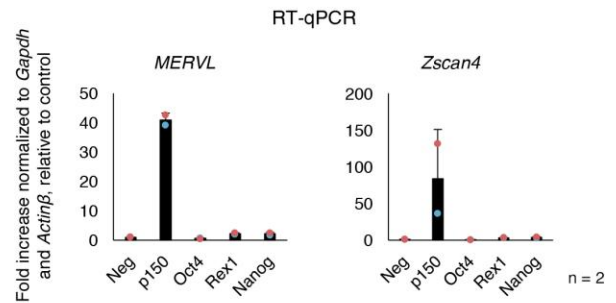
f



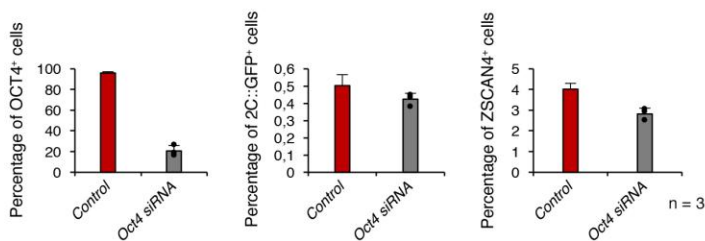
h



i



g



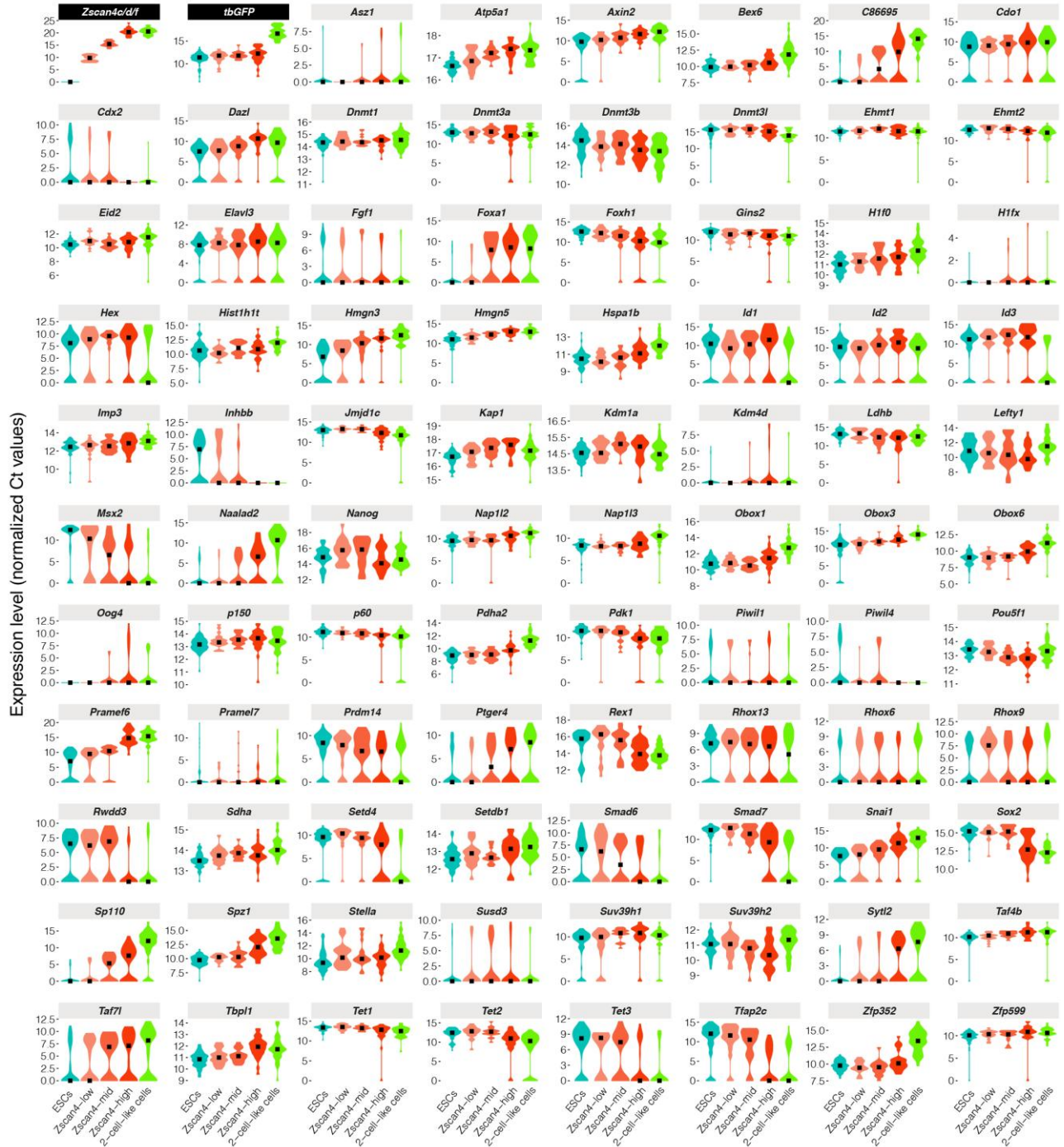
Supplementary Figure 4

Pluripotency transcription factors and the 2-cell-like state

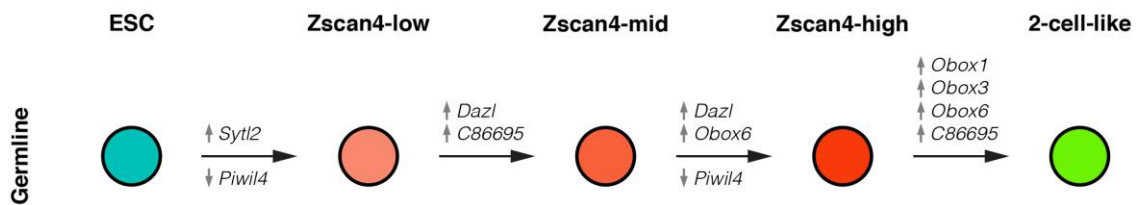
a, Scatterplot showing the fluorescence intensity measurements for OCT4 and ZSCAN4 in individual cells as judged by immunostaining. r depicts the Pearson correlation coefficient between OCT4 and ZSCAN4 expression for each group of cells, as indicated. **b,c**, Validation of the *Rex1::EGFP* and *Zscan4::tdTomato* cell line by immunofluorescence. A representative single confocal section from three independent cell cultures is shown. The *Rex1* knock-in construct was validated previously⁴⁴. **d**, Density plot showing the gating parameters used for sorting the *Rex1*^{high} and *Rex1*^{low} cells in Fig. 4a. **e,f**, Violin and density plots showing the distribution of single-cell expression for *Rex1* and *Nanog*. Note that in these plots ES cells were further classified into two groups according to whether they express high or low levels of *Rex1*, which highlights naive versus primed pluripotent states, as confirmed also by the abundance of *Nanog* transcripts in the same cells. **g**, Percentage of OCT4⁺, EGFP⁺ and ZSCAN4⁺ cells 48 h after transfection with siRNA for *Oct4* or the scrambled control. Data shown are the means \pm s.d. for three independent cell cultures. **h**, Percentage of EGFP⁺, ZSCAN4⁺ and 2-cell-like cells after transfection with *Oct4*, *Nanog*, *Sox2* or *Rex1* siRNA as compared to p150 siRNA and to the negative controls (NT and Neg). Transfection and analysis were performed as described in the Methods for the RNAi screen. Shown are the means \pm s.d. from triplicate cell cultures. **i**, RT-qPCR analysis of MERVL and *Zscan4* in the 2C::EGFP reporter cell line after transfection with the indicated siRNAs. Shown are the mean values \pm s.d. of two independent cell cultures.

Supplementary figure 5

a



b



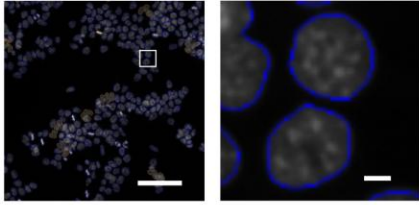
Supplementary Figure 5

Sequential gene expression changes during the transition to the 2-cell-like state

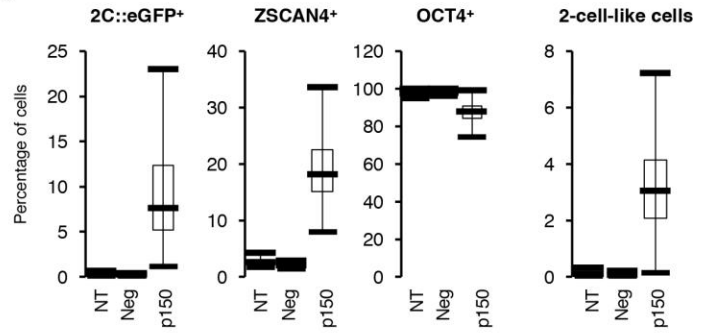
- a.** Violin plots showing the distribution of expression levels of individual cells for the indicated genes. Higher values correspond to higher expression levels, and a Ct value of 0 indicates that no amplification was detected. The median is indicated by a square.
- b.** Schematic of significantly and differentially expressed genes related to germline development between individual stages of the transition from the ES to the 2-cell-like state. Changes were considered significant if they exhibited at least 2-fold changes across cells between individual states and $P < 0.05$ (Mann–Whitney U test). The arrow indicates the direction (up or down) of the changes in gene expression.

Supplementary Figure 6

a

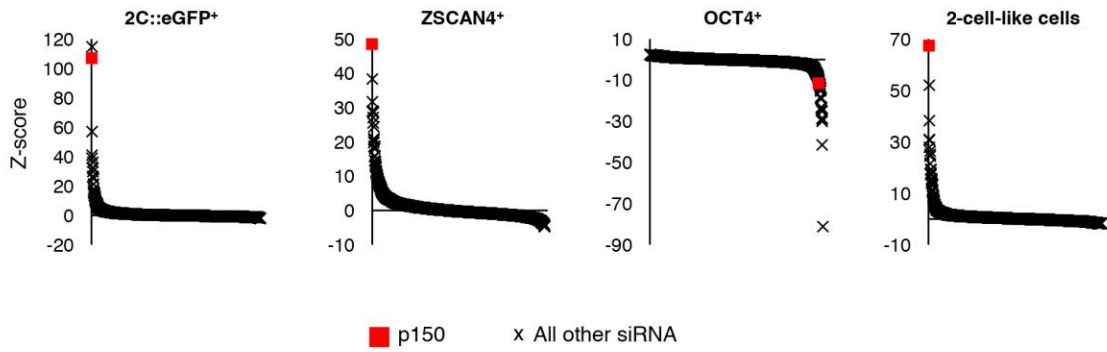


b

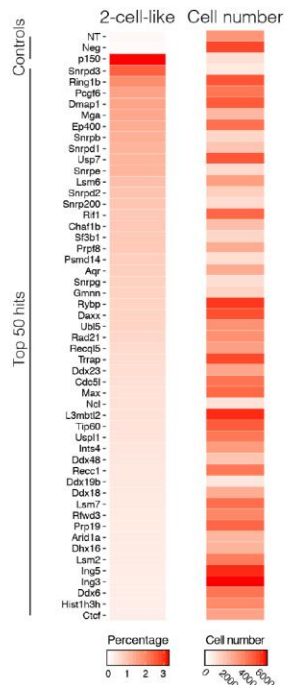


NT n= 45 ; Neg n=270 ; p150 n=270

c



d

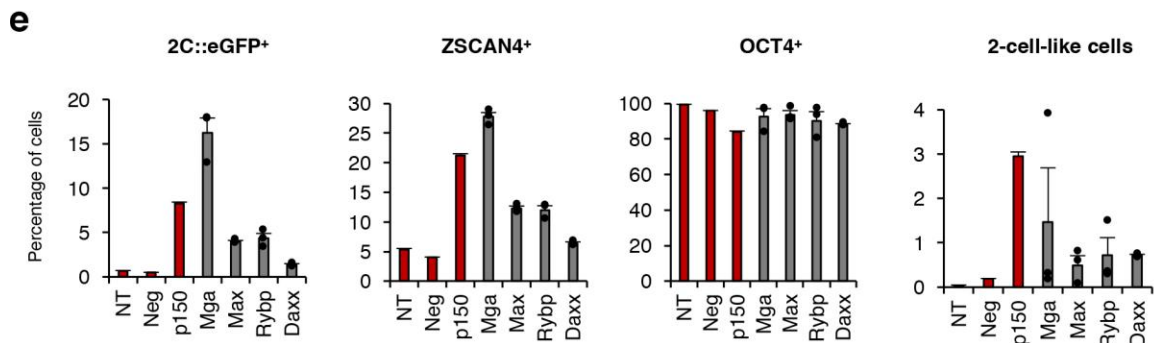
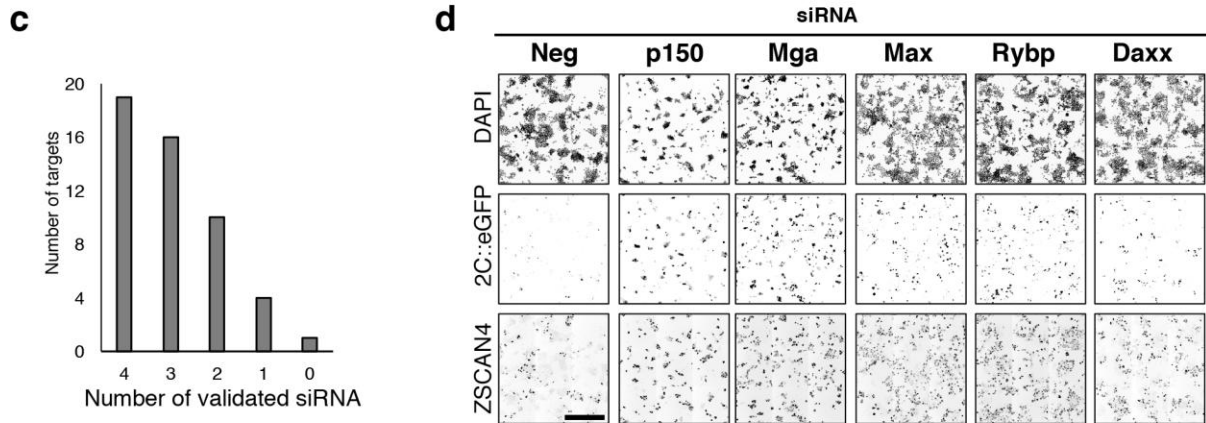
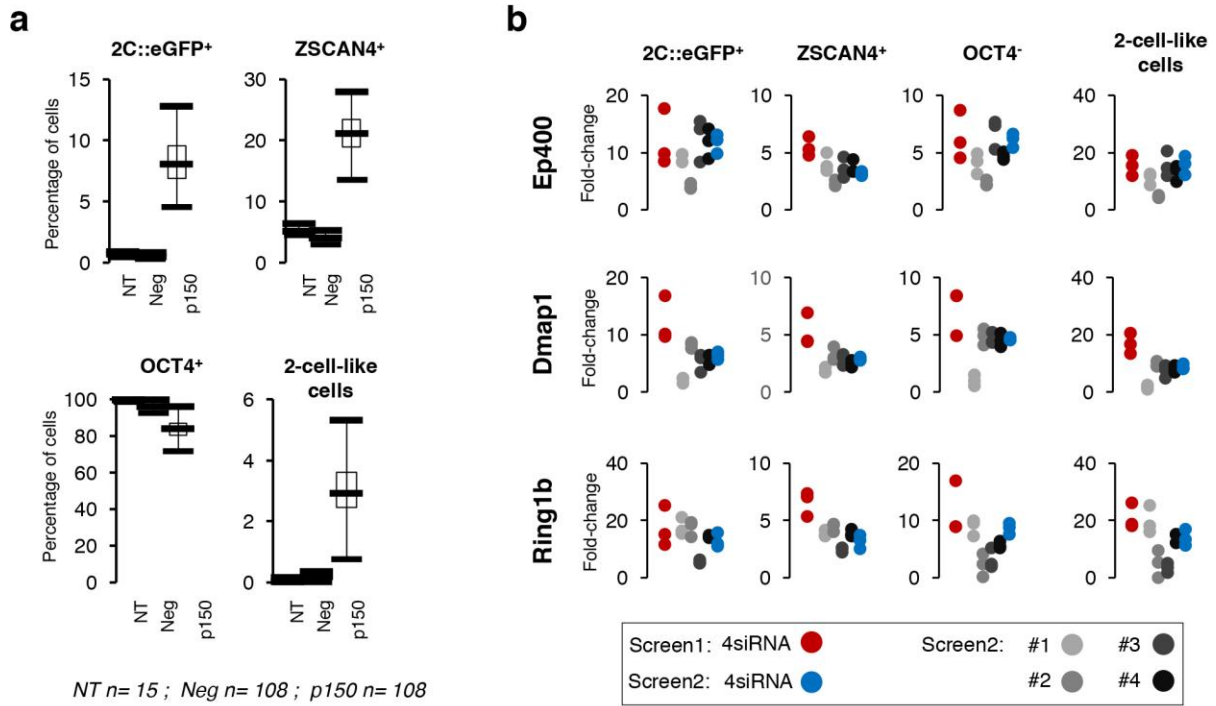


Supplementary Figure 6

Pipeline and controls for analysis of the primary siRNA screen

a, Screening was based on nuclear segmentation, following DAPI staining, for which a representative image is shown. Nuclei were segmented based on DAPI intensity, and only nuclei that met the quality control were used for further analysis (blue outlines). Scale bars, 100 μm (left) and 5 μm (right). **b**, Box-and-whisker plots for the negative (non-transfected cells (NT; $n = 45$ wells) and negative-control-siRNA-transfected cells (neg; $n = 270$ wells)) and positive (p150-siRNA-transfected cells (p150; $n = 270$ wells)) controls from the primary screen. The percentage of EGFP⁺, ZSCAN4⁺ and OCT4⁺ cells was determined for each cell culture well. Two-cell-like cells were defined as cells fitting all three criteria, namely: positive for EGFP and ZSCAN4 but negative for OCT4. On the graphs, boxes indicate 25% and 75% quartiles, and the whiskers extend to 1.5 times the interquartile range. Outlier wells are not shown. **c**, Complete results from the primary screen depicting the z scores of the 1,167 targets (mean z score of triplicate wells for each target) relative to the negative controls. The positive control p150 is depicted in red. **d**, Analysis of cell toxicity, as inferred by cell number, elicited upon treatment with siRNA for the top 50 hits. The heat map displays the top 50 hits ranked by their ability to induce 2-cell-like cells (left) and the cell number per well upon siRNA transfection (right). Note that, because all siRNAs were transfected using the same number of cells, changes in cell number indicate cell death and/or cell growth defects resulting from RNAi.

Supplementary Figure 7

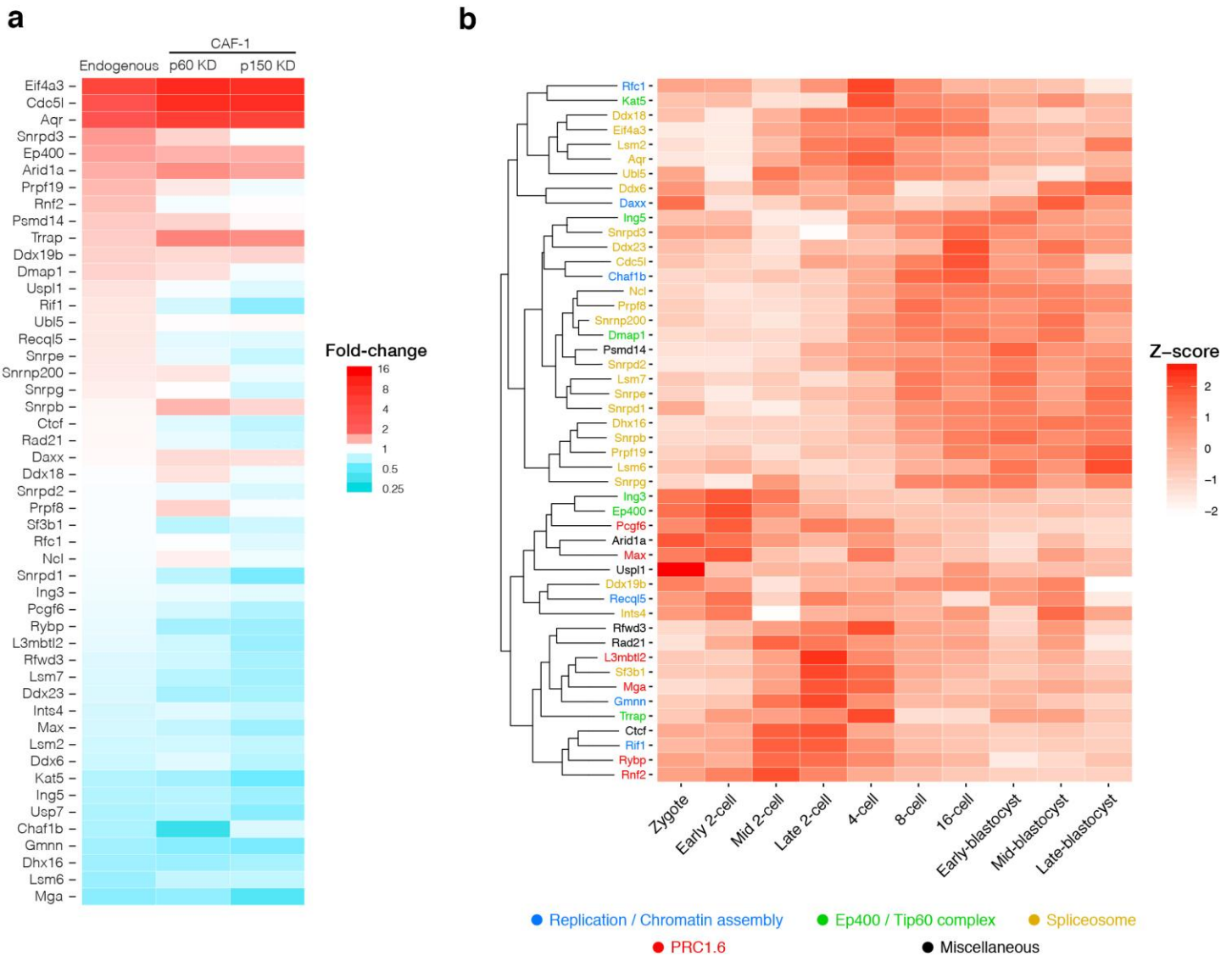


Supplementary Figure 7

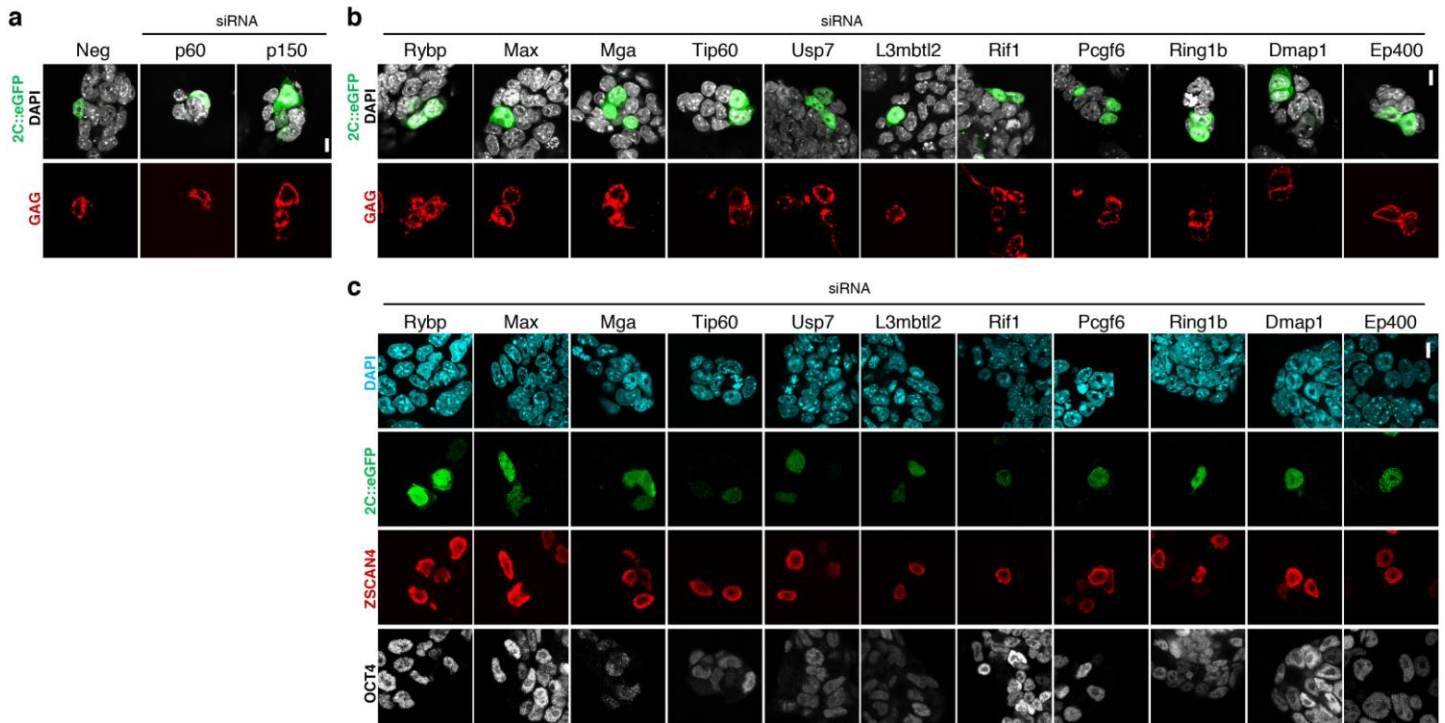
Validation of the hits obtained in the primary screen by a secondary screen and identification of new hits

a, Box-and-whisker plots representing the results of the secondary screen for the non-transfected cells (NT), scrambled-siRNA-transfected cells (Neg) and cells transfected with siRNA for p150 (positive control). *n* indicates the number of cell culture wells analyzed. Two-cell-like cells are defined as cells positive for EGFP and ZSCAN4 but negative for OCT4. Boxes indicate 25% and 75% quartiles, and the whiskers extend to 1.5 times the interquartile range. Outlier wells are not shown. The mean \pm s.d. of two technical replicates is shown. **b**, Comparison of the primary and secondary screen results for three selected hits (*Ep400*, *Dmap1* and *Ring1b*). Fold changes relative to the negative control are indicated. **c**, Validation of individual siRNAs from the siRNA pool for the top 50 hits. The top 50 hits from the primary screen were selected for validation in the secondary screen by transfecting four different individual siRNAs, and the effect of each individual siRNA on 2-cell-like cell emergence was assessed. The number of validated hits (*z* score > 2 as compared to the negative control) by 4, 3, 2 or 1 siRNA is depicted. Only one hit (*Dnmt3b*) from the primary screen was not validated by any of the four individual siRNAs. **d**, Representative random, inverted dynamics merged fields of view from the secondary screen for the indicated siRNAs as compared to the negative and positive controls. Scale bar, 500 μ m. **e**, Percentage of EGFP⁺, OCT4⁺ and ZSCAN4⁺ cells and of 2-cell-like cells obtained in the secondary screen for *Mga*, *Max*, *Rybp* and *Daxx* as compared to the negative (NT and Neg) and positive (p150) controls. Mean values \pm s.d. derived from triplicate cell cultures are shown.

Supplementary figure 8



Supplementary Figure 9



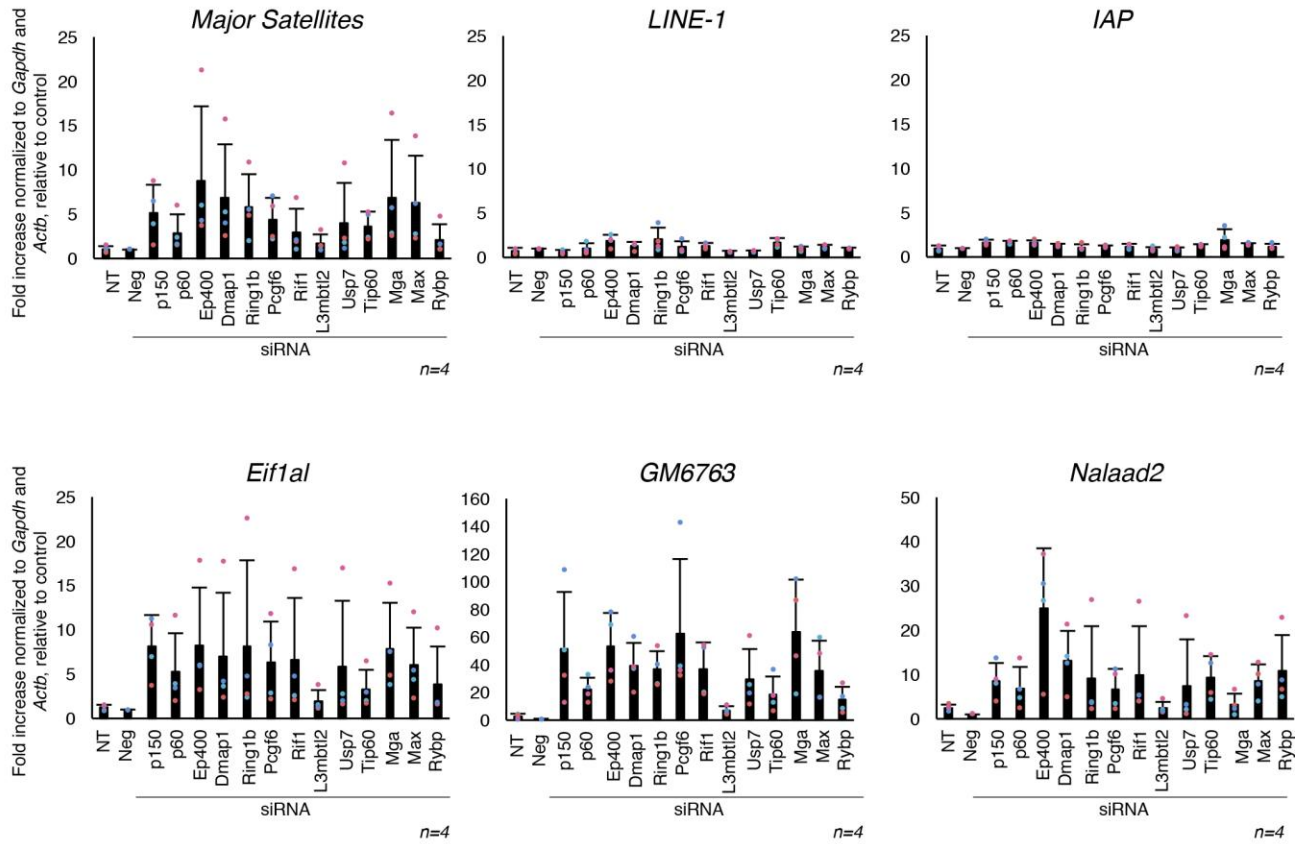
Supplementary Figure 9

Characterization of 2-cell-like cell protein markers in 2-cell-like cells induced upon siRNA of the hits identified in the siRNA screen

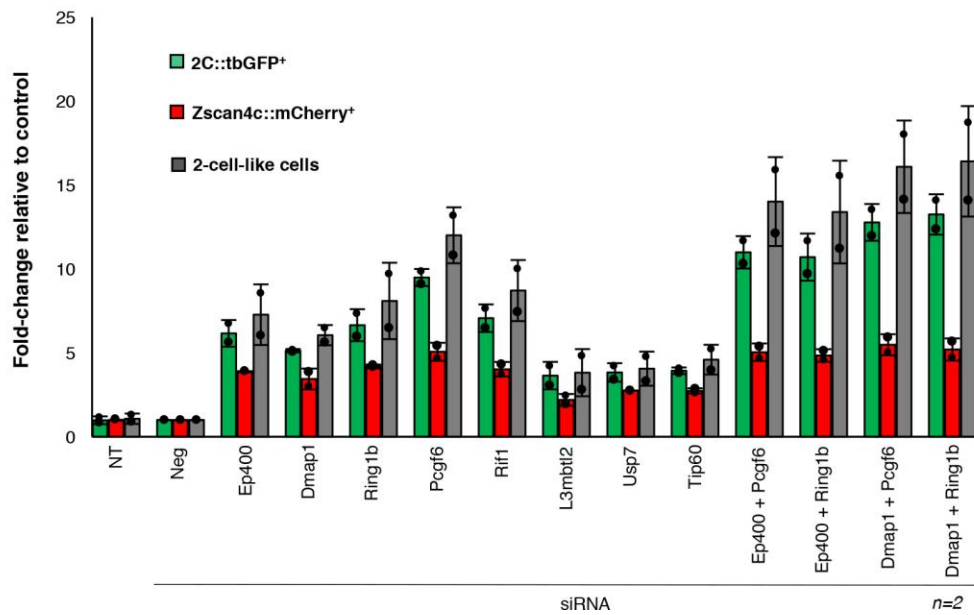
Characterization of 2-cell-like cell markers in 2-cell-like cells induced upon siRNA targeting of the hits identified in the siRNA screens. **a,b**, Immunostaining with an antibody against the protein from MERVL reveals expression of endogenous MERVL loci in cells expressing the 2C::EGFP reporter in controls (**a**) as well as in 2-cell-like cells induced upon siRNA targeting of the indicated chromatin modifiers (**b**). **c**, Immunostaining for ZSCAN4 and EGFP in the 2C::EGFP reporter ES cell line. Representative images from at least three independent cell cultures performed on different days are shown. Scale bars, 10 μ m.

Supplementary Figure 10

a



b

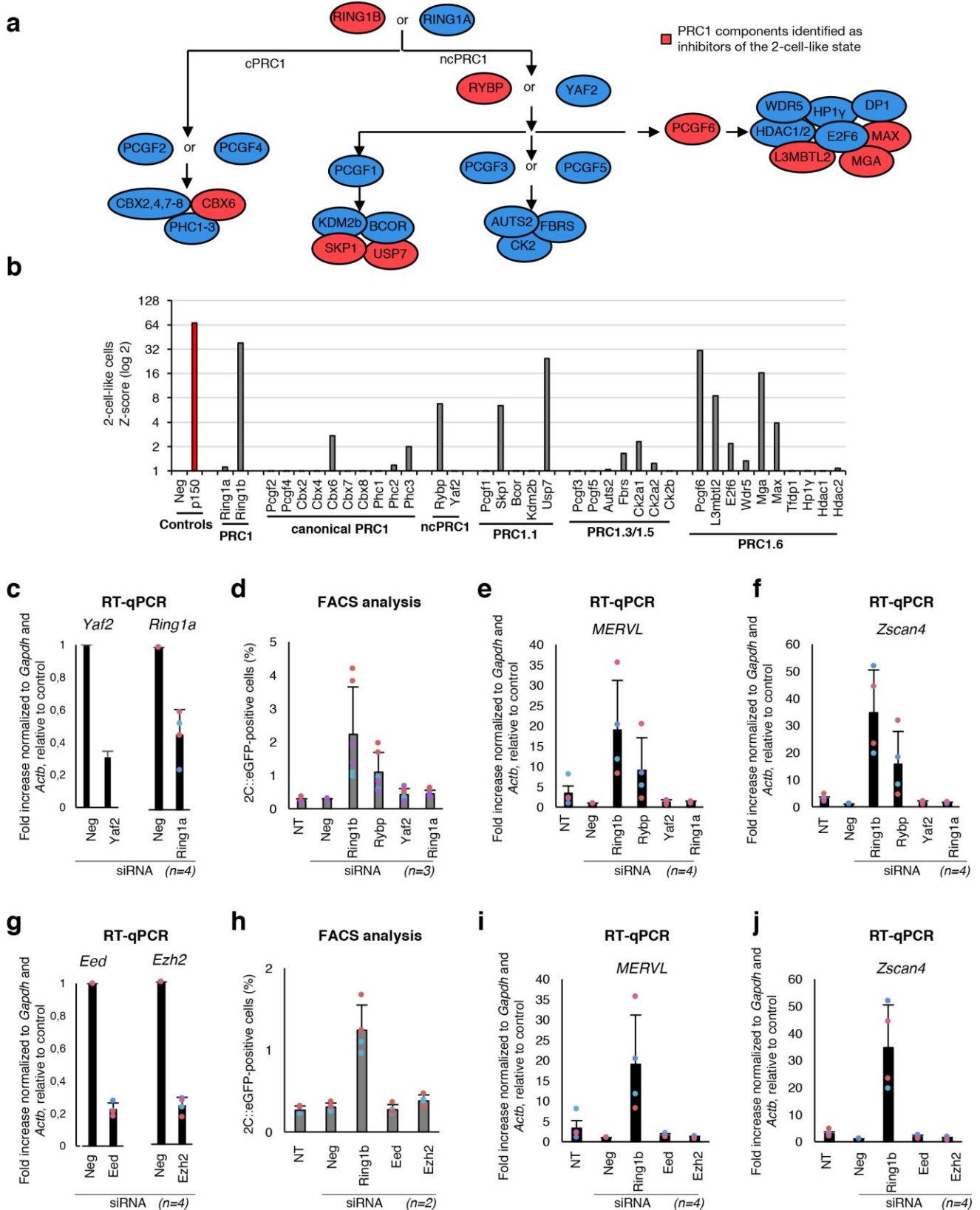


Supplementary Figure 10

Characterization of 2-cell-like cell transcriptional markers in 2-cell-like cells induced upon siRNA of the hits identified in the siRNA screen

a, Expression of 2-cell-like genes upon siRNA targeting of the identified hits. RT-qPCR analysis was performed for repetitive elements (top) and chimeric LTR transcripts (bottom) upon transfection with the indicated siRNAs. Shown are the mean values \pm s.d. from four independent cell cultures performed on two different days. **b**, FACS analysis of the 2C::turboGFP and *Zscan4*::mCherry cell line after transfection with the indicated siRNAs individually or in pairs. Fold changes in turboGFP⁺, mCherry⁺ and double-positive (2-cell-like) cells are shown. The mean \pm s.d. of the indicated number of cell cultures is shown.

Supplementary Figure 11

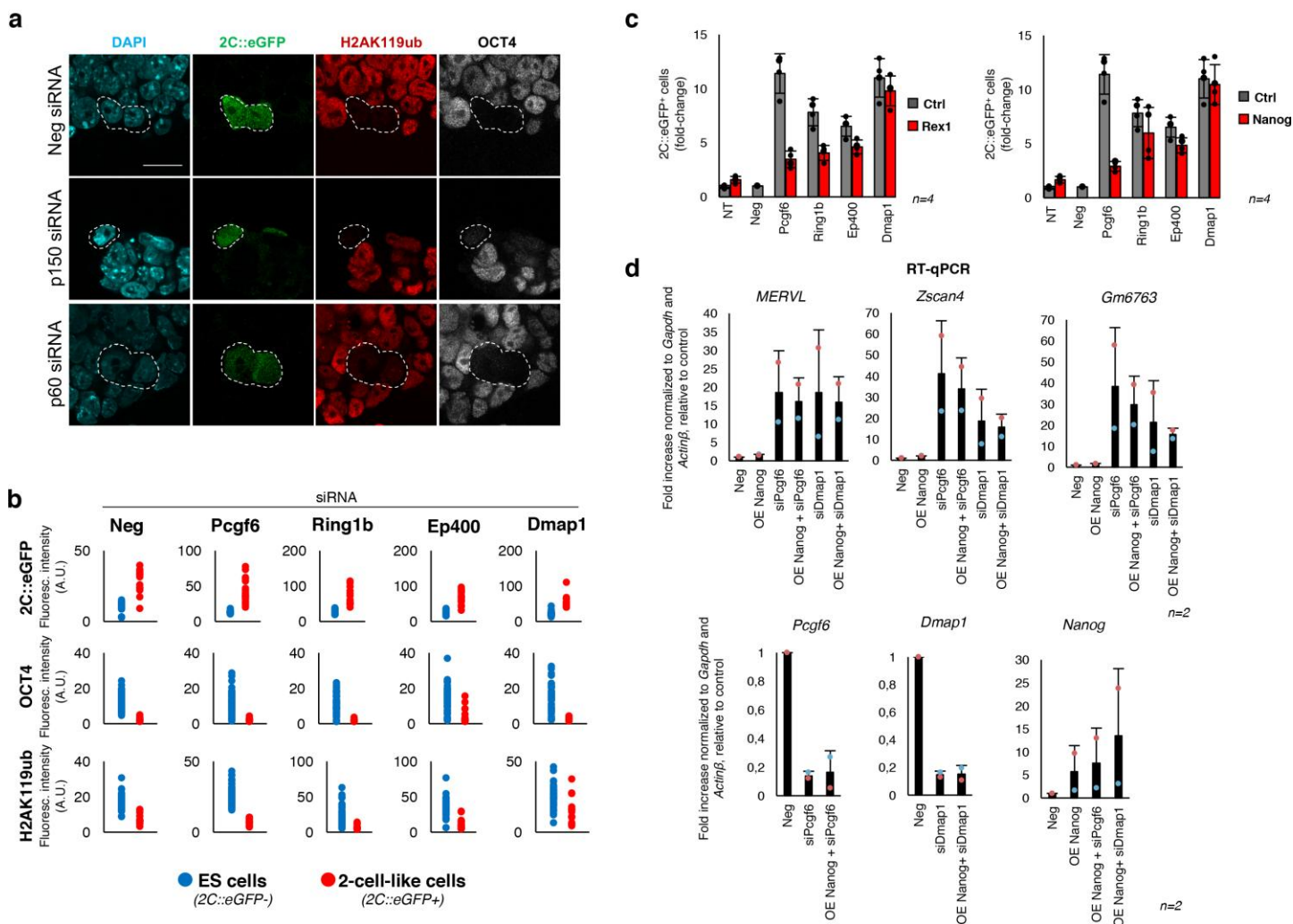


Supplementary Figure 11

PRC1.6 subunits negatively regulate the 2-cell-like state

a. Schematic of PRC1 complexes identified in mammals. PRC1 complexes are divided into cPRC1 (canonical PRC1) (left) and ncPRC1 (non-canonical PRC1) (right). RING1a and RING1b interact with distinct PCGF proteins. PCGF2 and PCGF4 are present only in canonical PRC1 complexes (PRC1.2 and PRC1.4, respectively). PCGF1, PCGF3, PCGF5 and PCGF6 proteins associate with RYBP or YAF2 to form the non-canonical PRC1 complexes (PRC1.1, PRC1.3, PRC1.5 and PRC1.6, respectively). **b.** Two-cell-like cell induction after transfection with siRNAs for all PRC1 components. Results for the siRNA pools identified in the primary or secondary screen are shown as a z score. **c,g,** RT-qPCR analysis was performed to measure siRNA efficiency for *Yaf2* and *Ring1a* (**c**) or *Eed* and *Ezh2* (**g**) after transfection with the corresponding siRNAs as compared to scrambled siRNA in the 2C::EGFP reporter cell line. Mean values \pm s.d. from four independent cell cultures performed on two different days are shown. **d,h,** Quantification of EGFP⁺ cells (%) by FACS after transfection with the indicated siRNAs. Shown are the means \pm s.d. of the indicated number of cell cultures. **e,f,i,j,** Expression of 2-cell-like genes upon treatment with the indicated siRNAs. RT-qPCR analysis was performed of MERVL (**e,i**) and *Zscan4* (**f,j**) expression in the 2C::EGFP reporter cell line after transfection with the indicated siRNA. The mean values \pm s.d. from four independent cell cultures performed on two different days are shown.

Supplementary Figure 12



Supplementary Figure 12

Two-cell-like cells are characterized by low levels of H2AK119Ub

a, Immunostaining for OCT4, EGFP and H2AK119Ub in the 2C::EGFP reporter ES cell line depicting endogenous (Neg siRNA) as well as p60- and p150-knockdown-induced 2-cell-like cells. Representative single-section confocal images of at least three independent cell cultures are shown. Dashed white lines demarcate EGFP⁺ cells. Scale bar, 20 μ m. **b**, H2AK119Ub levels in endogenous 2-cell-like cells and in 2-cell-like cells induced upon transfection with the indicated siRNAs. EGFP (top), OCT4 (middle) and H2AK119Ub (bottom) fluorescence was quantified in ES cells (blue, EGFP negative) and in 2-cell-like cells (red, EGFP positive). Each dot represents a single cell. Shown are raw values obtained in one representative experiment of three independent biological replicates performed on different days. **c**, Quantification of EGFP⁺ cells (fold change as compared to negative control Neg) by FACS after transfection with the indicated siRNAs in combination with *Rex1* (left) or *Nanog* (right) siRNA. The mean \pm s.d. of the indicated number of cell cultures is shown. **d**, RT-qPCR analysis of *MERVL*, *Zscan4* and *Gm6763* in the 2C::EGFP reporter cell line after transfection with the indicated siRNAs and/or overexpression of *Nanog* (OE *Nanog*). Expression of *Pcgf6* (lower left), *Dmap1* (lower middle) and *Nanog* (lower right) is shown as controls for siRNA and overexpression efficiency. The mean \pm s.d. of the indicated number of cell cultures performed on different days is shown.

Transgene 1	Transgene 2	Transgene 3	Experiments
2C::3XturboGFP-NLS-PEST	n.a.	n.a.	-RNA-seq data in Fig. 1 -ZSCAN4 immunofluorescence in Fig. 2 -SOX2 and REX1 immunofluorescence in Fig. 4
2C::3XturboGFP-NLS-PEST	CAG::NLS-tdTomato	n.a.	-Single-cell expression profiling in Fig. 1, Supplementary Figs. 1 and 4
2C::3XturboGFP-NLS-PEST	Zscan4c::tdTomato-PEST	n.a.	-Single cell expression profiling in Fig. 2 -PRDM14 and TFAP2C immunofluorescence in Fig. 3
2C::3XturboGFP-NLS-PEST	Zscan4c::mCherry-NLS-PEST	CAG::H2B-tdiRFP	- Time-lapse experiments in Fig. 2 - FACS analysis experiments in Fig. 7a, 7c and Supplementary Fig. 7.
2C::EGFP	n.a.	n.a.	-OCT4 immunofluorescence in Fig. 3 -Screening, Figures 5 to 7
Rex1::EGFP-PEST (knock in, ref. ⁴⁰)	Zscan4c::tdTomato-PEST	n.a.	-All experiments in Fig. 4

Supplementary Table 10. Summary of the reporter cell lines used in this study

Results

Part 2 – The molecular basis of mammalian totipotency *in vivo*

A systematic comparison of mammalian pre-implantation development transcriptomes

Diego Rodriguez-Terrones¹, Camille Noll¹, Irene Aksoy², Kilian Simmet³, Mayuko Kurome³, Valeri Zakhartchenko³, Nathalie Beaujean², Eckhard Wolf³ and Maria-Elena Torres-Padilla^{1,4}

¹ *Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München D-81377 München, Germany*

² *Stem Cell and Brain Research Institute, F- 69500 Lyon, France*

³ *Genzentrum, Ludwig-Maximilians-Universität München D-81377 München, Germany*

⁴ *Faculty of Biology, Ludwig-Maximilians-Universität München D-81377 München, Germany*

Abstract

The transient capacity to autonomously form and organize all of the embryonic and extra-embryonic tissues involved in the development of a complete organism is termed totipotency. However, our understanding of the regulatory mechanisms responsible for the establishment and termination of such a highly plastic regulatory state is greatly limited. Transposable elements constitute some of the earliest sequences to be transcribed in the early mammalian embryo, and their activity at this time has been ascribed regulatory roles in both mouse and humans. Using a low-input hybrid full-length + 5' RNA-sequencing approach, we have generated transcriptional profiles at seven different developmental timepoints across five different mammalian species. Using this data, we have determined the transcriptional dynamics of transposable elements during this developmental period. Our results pave the way to understand the contribution of transposable elements to mammalian pre-implantation development.

Introduction

How does a single cell give rise to an entire multicellular organism is one of the most fascinating questions in biology. This transient capacity to autonomously form and organize all of the embryonic and extra-embryonic tissues involved in the development of a new organism is termed totipotency. In mammals, totipotency is a feature restricted to the earliest cells of the pre-implantation embryo, and its developmental timing differs depending on the species in question. Over the course of mouse development, only the zygote and each of its two daughter cells are strictly totipotent since only they can produce an entire organism on their own (Tarkowski, 1959). For example, if both of the 2-cell stage blastomeres from the mouse embryo are physically separated and allowed to continue to develop, genetically identical mouse offspring can be

obtained. This capacity is lost by the time the next cell cycle occurs, and no mouse pups have been reported to arise from single 4-cell or 8-cell stage blastomeres, owing to the fact that the resulting blastocysts tend to lack an Inner Cell Mass (Rossant, 1976; Tarkowski, 1959). In a series of classical experiment conducted over the course of the last six decades, the totipotent windows for a series of mammalian species have been defined. While both cow and rhesus embryos have been reported to retain totipotency until up to the 4-cell stage (Chan, 2000; Johnson et al., 1995), single blastomeres of the 8-cell stage in pigs, sheep and rabbit have all been reported to be able to give rise to live offspring (Moore et al., 1968; Saito and Niemann, 1991; Willadsen, 1981).

The molecular determinants underlying this singular capacity to initiate development of an entire organism have proven elusive, and it remains an open question whether a gene regulatory network akin to the one underlying pluripotency exists in totipotent cells, or whether some other factor might underlie it. In mouse, totipotency is accompanied by a set of highly-intermingled biological processes such as global chromatin remodeling (Burton and Torres-Padilla, 2014; Eckersley-Maslin et al., 2018), an unusual set of metabolic characteristics (Kaneko, 2016; Nagaraj et al., 2017) and the de-repression of the vast majority of transposable elements (Fadloun et al., 2013b; Jachowicz et al., 2017; Peaston et al., 2004; Rodriguez-Terrones and Torres-Padilla, 2018). In particular, while transposable elements (TE) such as ERV-L have been shown to provide promoter sequences for a wide variety of genes activated at the time of Zygotic Genome Activation (ZGA) (Peaaston et al., 2004), other transposable elements such as L1 have been proposed to directly regulate the early developmental process through their RNA or even through the act of transcription itself (Jachowicz et al., 2017; Percharde et al., 2018).

In spite of their emerging roles as essential regulators of pre-implantation development, their study during early development has been hampered by the low cell numbers present at these stages, the lack of RNA-sequencing methods capable of distinguishing read-through transcription from transcription initiation roles, and from the computational challenges inherent to highly repetitive sequences (Goerner-Potvin and Bourque, 2018). In addition, transcriptional dynamics of transposable elements have yet to be characterized beyond mouse (Fadloun et al., 2013b) and human (Göke et al., 2015), and it remains an open question how can the highly divergent repetitive landscapes seen in mammalian genomes mediate such a conserved developmental process. In this work, we present a low-input hybrid full-length + 5' RNA-sequencing approach capable of simultaneously distinguishing read-through transcription from transcription initiation sites and employ it to characterize the transcriptional landscapes of pre-implantation development across five different mammalian species. While still in progress, our

work aims to ultimately assess the contribution of transposable element transcription to early development and to identify a common transcriptional program underlying the totipotent state.

Materials and methods

Embryo collection

Because of the higher cDNA yields obtained compared to the Qiagen TCL lysis buffer (Cat. #1031576), the Clontech 10X lysis buffer (Cat. #635013) was used for all embryo collections. A 1:581,000 dilution of ERCC RNA spike-ins was included at the moment of lysis buffer preparation. Lysis buffers were prepared from a single working solution and aliquoted in low-binding PCR strips (Sarstedt, 72.991.002) carrying 5.8 μ L of lysis buffer in each tube. For the single-cell RNA-seq method comparisons, pools of 5x mid-2-cell stage C57BL6 mouse embryos were collected at ~46 hours post-fertilization from natural matings, washed in PBS and deposited in the lysis buffer with the zona pellucida intact. Lysates were then flash frozen in liquid nitrogen and no pipetting of the lysate was performed. For the systematic comparison of mouse, pig, cow and rabbit pre-implantation transcriptomes, embryos were collected at the timings shown in Supplementary Table 1, washed in PBS and a single embryo was deposited in an individual tube of a PCR strip as described above. PBS washes consisted of 2x 2-min washes and 1x 5-min wash in sterile PBS.

Preparation of Smart-seq+5 samples

Samples were prepared as per the Smart-seq2 protocol (Picelli et al., 2014) with a few modifications (Supplementary Table 2). Following unfreezing of the embryo lysates at room temperature, RNA was purified using 15 μ L of AMPure RNA magnetic beads (Beckman Coulter, A63987), washed once in 100 μ L of 80% ethanol and resuspended in 3 μ L of Annealing mix containing 1 μ L dNTP mix (ThermoFisher, R0192), 1 μ L oligo-dT30 (Sigma, 5'-AAGCAGTGGTATCAACGCAGAGTACT₃₀V-3') at 10 μ M and 1 μ L nuclease free water with 5% RNase inhibitor (Clontech 2313A). Following a 3 min incubation at 72° C, samples were held at 4° C or in ice. Afterwards, 7 μ L of reverse transcription mix were added containing: 2 μ L Superscript II RT buffer (ThermoFisher, 18064014), 2 μ L 5M Betaine solution (Sigma, B0300-1VL), 0.5 μ L DTT, 0.5 μ L Superscript II RT, 0.25 μ L RNase inhibitor (Clontech 2313A), 0.1 μ L rGrG+G TSO (TIB MolBiol, 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3') at 100 μ M, 0.06 μ L 1M MgCl₂ (Sigma, M1028) and 1.6 μ L of 40% PEG-8000 solution (Sigma, P1458). Samples were then incubated for 90 min at 42° C and the enzyme was afterwards inactivated at

70° C for 15 min. PCR was performed using the KAPA HiFi ReadyMix (KM2605) for 14 cycles and further purified as described in the Smart-seq2 protocol, although using 12.5 µL of AMPure beads instead. Finally, tagmentation was carried out with the Nextera XT kit (Illumina, 15032354) as described in the Smart-seq2 protocol, with a few modifications. Namely, 2.5 µL of 120 pg/µL cDNA were used per reaction (for a total of 300 pg) and mixed with 2.5 µL of Amplicon Tagment Mix and 5 µL of Tagment DNA buffer. After incubation for 5 min at 55° C, 2.5 µL of NT buffer were pipetted into each sample and another incubation for 5 minutes at room temperature was performed. Finally, 5 µL of a custom sequencing adaptor mix (IDT) was added. This mix contained two standard i5 and i7 Nextera Unique Double Indexes and an additional tailed i7 index with an overhang overlapping the Smart-seq2 adaptors (Index and overhang shown in bold: 5'-CAAGCAGAAGACGGCATAACGAGATNNNNNNNNGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGA**AAGCAGTGGTATCAACGC***A*G-3'), all at 2.5 µM concentration. Afterwards, 7.5 µL of NPM mix was added and PCR was carried out for 12 cycles. Libraries were sequenced in a HiSeq4000 using 2 x 150 bp reads. All oligo sequences are described in Supplementary Table 3.

Preparation of mcSCRB-seq and Hybrid protocols

Samples were prepared as described above, except for alternative reverse transcription mixes. In the case of the mcSCRB-seq based protocol (Bagnoli et al., 2018), the reverse transcriptase mix consisted of 7 µL containing: 2 µL Maxima H- RT buffer (ThermoFisher, EP0752), 0.5 µL Maxima H- RT, 0.25 µL RNase inhibitor (Clontech 2313A), 0.1 µL rGrGrG TSO (Sigma, 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrGrG-3'), 2.55 µL nuclease free water and 1.6 µL of 40% PEG-8000 solution (Sigma, P1458). In the case of the Hybrid protocol, the reverse transcriptase mix consisted of 7 µL containing: 2 µL Maxima H- RT buffer (ThermoFisher, EP0752), 0.5 µL Maxima H- RT, 0.25 µL RNase inhibitor (Clontech 2313A), 0.1 µL rGrG+G TSO (TIB MolBiol, 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3'), 2 µL 5M Betaine solution (Sigma, B0300-1VL), 0.01 µL 1M MgCl₂ (Sigma, M1028), 0.54 µL nuclease free water and 1.6 µL of 40% PEG-8000 solution (Sigma, P1458). Smart-seq2 adaptor sequences – with the possible exception of the LNA base in the TSO – were used in all protocols tested to ensure an even comparison.

Preparation of Smart-seq2+PA libraries

Pools of 5x 2-cell stage mouse embryos were collected in 10 µL Qiagen TCL lysis buffer (Cat. #1031576) and flash frozen. Upon thawing, RNA was purified using AMPure RNA magnetic

beads (Beckman Coulter, A63987), washed once in 100 μ L of 80% ethanol and resuspended in 10 μ L of Polyadenylation mix containing 7 μ L water, 0.5 μ L E. coli Poly(A) polymerase (NEB, #M0276), 1 μ L 10 Polymerase reaction buffer, 1 μ L ATP and 0.5 μ L RNase inhibitor (Clontech 2313A). Following a 10 min incubation at 37° C, RNA was purified using magnetic beads once again and resuspended in annealing mix as per the standard Smart-seq2 protocol (Picelli et al., 2014).

Bioinformatic analyses used for protocol comparisons

Following sample demultiplexing, Nextera adaptors were trimmed using Trimmomatic's palindrome mode and reads shorter than 20 bases were discarded. In order to sort reads into internal, 5' or 3' fragments, a custom python script was used that operated as following. Using the biopython library, the first 25 bases in read #2 (in the case of the Hiseq 4000, since the i7 adaptor carried the overhang in our setup) were aligned to the 25 bases in common between both the 5' and 3' adaptors (5'-AAGCAGTGGTATCAACGCAGAGTAC-3'). If less than 20 identities were found between the adaptor and the beginning of the read, the read was considered an internal fragment and sorted to its respective fastq file. If 20 or more bases overlapped between the adaptor and the beginning of the read, the first 30 bases of the read were taken and aligned to the 30 bases of the 5' adaptor (5'-AAGCAGTGGTATCAACGCAGAGTACATGGG-3') and the first 30 bases of the 3' adaptor (5'-AAGCAGTGGTATCAACGCAGAGTACTTTTT-3'). The read was defined as a 5' or 3' fragment depending on which of these two adaptors provided the best score. Subsequently, reads from each fragment type were aligned to the mouse reference genome using Star. For the metagene analyses in Fig. 1B, sequencing depth normalized pileup tracks were generated using bedtools genomecov and plotting was performed using deeptools. Pileup tracks for the 5' fragments were generated based solely on the coordinates of the 5' end of read #2. For the visualization of the representative examples shown in Fig. 1C, the Gviz R package was employed using the aforementioned tracks. All differential expression analyses were performed using the DESeq2 R package.

For the 5' specificity measurements over ERCC RNA spike-ins shown in Figs. 2F and 2G, reads were considered at the true +1 base if the 5' coordinate of the 5' fragment was exactly the first base of the ERCC spike-in. For the 5' specificity measurements over annotated transcripts shown in Figs. 2H and 2I, rRNA mapping reads were removed based on the RepeatMasker annotation and the proportion of 5' fragments aligning within a 500 bp window on either side of

annotated TSSs in the GENCODE M18 release was calculated. For the precision and sensitivity analysis in Fig. 2J, 5 million reads from each of the three 5' methods tested were sampled, and the 5' fragments extracted and processed as described (resulting in a variable number of 5' tags per method, depending on the proportions shown in Fig. 2E). Paraclu was employed to identify 5' tag peaks using a minimum size parameter of 1. Subsequently, all integer combinations of d and s parameters between 0 and 50 were tested, and the paraclu peaks were filtered using these cutoffs. The p parameter was not used for filtering. For each of these parameter combinations, the precision, sensitivity and F1 statistic was calculated, and ultimately the best parameter combination was chosen based on the F1 statistic. Using these method- and replicate- specific optimal d and s parameters, paraclu peaks were filtered and used to calculate the precision, sensitivity and F1 statistic shown in Fig. 2J. These statistics were calculated as described elsewhere from the number of True positives (TP), False positives (FP) and False Negatives (FN), which were calculated as follows. The number of TPs were calculated as the number of peaks overlapping a 500 bp window centered on the 5' ends of transcripts in the GENCODE M18 release (i.e. from here on out 'promoter regions'). The number of FPs were calculated as the number of peaks overlapping a transcript but not any promoter region. The number of FNs was calculated as $FN = (FN_rate / (1 - FN_rate)) * TP$. To calculate the FN rate, we first selected a set of high confidence genes. A high confidence gene was considered as such if it exhibited more than 100 counts (across both 5', 3' and internal fragments) in all of the methods tested. The FN rate was calculated as the number of promoter regions belonging to high confidence genes that don't overlap any peak. Finally, for the determination of TSSs from vanilla Smart-seq2 data shown in black in Figs 2J and 2K, 5 million sampled reads were used to assemble transcripts using Stringtie without a reference GTF. Calculation of the variables defined above was performed in the same manner but employing the 5' ends of the assembled transcripts rather than paraclu peaks. In the case of Fig. 2K, TP and FP numbers were estimated at multiple sequencing depths.

Results

Smart-seq+5 allows for the identification of TSSs from full-length scRNA-seq data

We first set out to establish a transcriptional profiling method capable of distinguishing read-through transcription from transcription initiation sites in low RNA input amounts. While recent advances in single-cell RNA-sequencing technologies have permitted the routine interrogation of single-cell transcriptomes in a highly sensitive manner, these approaches are typically tailored to

either the internal or the terminal ends of a transcript and so far, do not allow for the joint interrogation of both of them.

Cap Analysis of Gene Expression (CAGE) is a cap-trapping-based method that enables the identification of transcription start sites (TSSs) throughout the genome (Carninci et al., 2006; Faulkner et al., 2009). By selecting for the transcript sequence proximal to the RNA cap, it identifies TSSs with base-pair resolution and permits the elucidation of promoter architectures. However, even in its most advanced versions (Cvetesic et al., 2018), CAGE protocols still require nanogram-amounts of RNA. Although the technological development of both CAGE and single-cell RNA-sequencing (scRNA-seq) approaches has occurred in parallel in recent years, recent installments of both methodologies are both based on the template switching activity of the reverse transcriptase, and therefore share most of their molecular intermediates. Thus, because of the extensive similarities between both sets of protocols, we conjectured that we could achieve full-length transcript coverage and TSS identification from a protocol combining different aspects of the two approaches.

Because of its high sensitivity, full-length transcript coverage and adequate throughput, we chose to adapt the scRNA-seq Smart-seq2 protocol to also provide TSS information (Ziegenhain et al., 2017). Smart-seq2 generates full-length cDNAs of the target transcripts by incorporating primer sequences on the 3' end through poly-A priming, and on the 5' end through a template-switching reaction (Picelli et al., 2013, 2014). Thus, addition of 'tailed' sequencing adaptors harboring an overhang overlapping these adaptors should allow for sequencing of the terminal ends in addition to the internal fragments generated during the standard fragmentation reaction (Fig. 1A). Indeed, this approach is used routinely in scRNA-seq protocols to sequence the 3' of molecules in molecular barcode approaches (Soumillon et al., 2014), and when applied to the Smart-seq2 protocol it indeed permitted us to capture the 5' ends of transcripts. In addition, recent reports indicate that the addition of molecular crowding agents such as polyethylene glycol (PEG) allows for a considerable increase in sensitivity when applied to the SCR-seq scRNA-seq protocol (Bagnoli et al., 2018). When applied to the Smart-seq2 protocol, we indeed observed a dramatic increase in the efficiency of the reverse transcription reaction which resulted in 2 PCR cycles less being required to obtain equivalent yields of full-length cDNAs (data not shown). Throughout this work, we refer to this molecular crowding variant of the Smart-seq2 protocol, and its subsequent fragmentation using a mix of standard and tailed indexes – 'hybrid' fragmentation – as **Smart-seq+5**.

In order to assess 5' specificity and identify any changes in sensitivity in the modified Smart-seq+5 protocol, we generated RNA-seq profiles in pre-implantation mouse embryos using both the original and the modified protocols. For this comparison, we employed pools of 2-cell stage mouse embryos in order to minimize biological variability in this technical comparison and included ERCC RNA spike-ins as a positive control because of their synthetically-defined 5' coordinates. In order to ensure an even comparison, PEG was employed at an equal concentration in both protocols. Following preparation and sequencing of these libraries, we bioinformatically sorted 5' fragments from internal fragments based on the presence of the 5' adaptor in the first positions of the i7 read (Fig. 1A) and aligned them to the mouse genome. This strategy successfully permitted us to identify the +1 base of both ERCC spike-ins and annotated protein coding transcripts, without impacting the distribution of reads corresponding to the internal regions of transcripts (Fig. 1B-C). Indeed, while the synthetically-defined ERCC spike-ins exhibited a sharp peak at their true +1 base but not at their +2 base (Fig. 1B), protein coding transcripts exhibited a broader distribution centered on the annotated start of the transcript, reflecting the existence of various promoter architectures (Carninci et al., 2006).

This approach had no impact on the sensitivity of the protocol, as a differential expression analysis performed between the internal reads of the samples prepared using the Smart-seq2 and the Smart-seq+5 variant showed no changes in gene expression (Fig. 1D). Intriguingly, when both internal reads and 5' fragments were included in the analysis (of course, for the Smart-seq+5 samples only), 33 genes were shown to be enriched in the Smart-seq+5 samples. Upon further examination, these genes were revealed to possess significantly shorter exonic gene lengths compared to all other detected genes (Fig. 1E), suggesting an increased sensitivity of the Smart-seq+5 variant to short transcripts compared to the Smart-seq2 protocol. Because the Smart-seq+5 tagmentation approach requires only a single transposition reaction to occur along the length of a cDNA molecule for it to be captured, this suggest that this protocol provides a better representation of short transcripts compared to the Smart-seq2 tagmentation approach, which requires one additional transposition reaction to occur along a cDNA molecule in order for it to be captured (Fig. 1F).

Smart-seq+5 provides better sensitivity and specificity for the 5' end of transcripts

Having determined that the tagmentation strategy employed in the Smart-seq+5 protocol efficiently identified the 5' ends of transcripts without an impact on sensitivity, we assessed its

specificity for 5' ends. Since reverse transcription conditions are known to affect the efficiency of the template switching reaction and to favor or hinder the appearance of strand invasion artifacts by the template switching oligo – which would affect specificity for the 5' end – we decided to compare three contrasting reverse transcription conditions (Fig. 2A). While the Smart-seq2 scRNA-seq protocol relies on the addition of additives such as betaine, additional magnesium and an LNA TSO for optimal yields (Picelli et al., 2013), the mcSCRB-seq protocol (Bagnoli et al., 2018) relies instead on a highly processive reverse transcriptase and doesn't require any of those additives because of the use of an RNA TSO. We explored whether a higher cDNA yield could be obtained by systematically combining different concentrations from all these components. We obtained maximum cDNA yields when we employed the Maxima H- RT from the mcSCRB-seq protocol in the Smart-seq2 protocol, in the presence of diminished $MgCl_2$ concentrations (data not shown and Fig. 2A-B). Finally, we applied these three protocols to pools of 2-cell stage mouse embryos as described above.

Overall, the Smart-seq+5 protocol offered the best performance, as it exhibited the highest sensitivity (Fig. 2C) and the lowest proportion of rRNA mapping reads (Fig 2D). While the hybrid protocol had exhibited the highest cDNA yields (Fig. 2B), it actually exhibited the lowest sensitivity and the highest proportion of rRNA mapping reads (Fig. 2C-D), suggesting that the increases in cDNA yield obtained from the intermixed reverse transcription reaction conditions resulted from strand invasion artifacts of the TSO. Thus, cDNA yield is a poor proxy for the sensitivity of scRNA-seq protocols and should be avoided when optimizing them. Intriguingly, the 3' ends of transcripts were essentially uncaptured (constituting less than 1% of the reads, Fig. 2E) by the hybrid tagmentation approach used across all protocol variants, despite this end of the transcript being used in both the SCRIB-seq (Soumillon et al., 2014) and Drop-seq protocols (Macosko et al., 2015). Given that those protocols specifically and selectively amplify the 3' ends of transcripts, we speculate that when both ends are amplified simultaneously, the AT-richness of the 3' end poly-A tail places these sequences on an amplification disadvantage relative to the 5' end.

Next, we compared the 5' end specificity of the different protocols over the length of the ERCC RNA spike-ins, because of their synthetically-defined defined 5' ends. While all three variants performed similarly (Fig. 2F), the mcSCRB-seq+5 protocol performed best both when assessed across all ERCC mapping reads (Fig. 2F) and when its performance was evaluated in each ERCC transcript individually (Fig. 2G). In contrast, when determined across protein coding transcripts (Fig. 2H-I and 2K), the Smart-seq+5 protocol came out on top, providing the best sensitivity, precision and F1 statistic (Fig. 2J). Thus, because of its overall better performance across

endogenous transcripts, we selected the Smart-seq+5 protocol for the subsequent characterization of pre-implantation transcriptomes across mammals. Importantly, it's worth noting that because of the alternative lysis conditions, the use of different adaptor sequences and the alternative tagmentation approach, these results should not be interpreted as indicative of the existence of a superior protocol among the three, but rather of the identification of – marginally – better performing reverse transcription conditions for our 5' tagging approach.

***In vitro* RNA poly-adenylation permits the detection of non-poly-A transcripts in single-cell protocols**

In many RNA-sequencing protocols, oligo-dT priming serves an important role in reducing the amount of rRNA sequenced and greatly improving their cost effectiveness. By priming the reverse transcription reaction at the transcript's poly-A tail, this approach permits for the selection of mRNA molecules while excluding rRNA, tRNAs and numerous other coding and non-coding transcripts that lack a poly-A tail (Fig. 3A). However, because of the high levels of pervasive transcription seen in pre-implantation embryos and the possible existence of abundant non-polyadenylated transcripts during these stages, we explored the effectiveness of polyA-independent priming mechanisms. We reasoned that perhaps the easiest way to extend a polyA-dependent protocol such as Smart-seq2 into capturing non-poly-adenylated transcripts would be to *in vitro* poly-adenylate all transcripts. Therefore, we decided to introduce an additional step after purification of sample RNA where we would poly-adenylate all transcripts using the *E. coli* Poly(A) polymerase (Fig. 3A), before proceeding with the standard Smart-seq2 protocol (throughout this text, this polyadenylation protocol is referred to as Smart-seq2+PA).

While both the Smart-seq2 and the poly-adenylation-dependent Smart-seq2+PA variant exhibited similar levels of sensitivity, the standard Smart-seq2 detected slightly more genes at each sampling level (Fig. 3B), probably due to the higher rRNA capture seen in the poly-adenylation variant (Fig. 3C). Notably, the Smart-seq2+PA succeeded in capturing non-polyadenylated transcripts such as histone genes, although at the expense of the less efficient capture of other protein-coding transcripts (Fig. 3D). Thus, because of the overall lower – although extended – sensitivity in the polyadenylation variant, we decided to proceed with the non-polyadenylation protocol for the generation of pre-implantation transcriptomes across mammals.

A comparison of pre-implantation development transcriptomes across mammals

Because of their known totipotent windows, the availability of a reference genome and the attainability of pre-implantation embryos, we decided to characterize the mouse, cow, pig, rabbit and rhesus macaque pre-implantation transcriptomes. Since the rhesus macaque embryo collection and preparation is still ongoing at the time of this submission, in this thesis I will only present the four other species. We have generated transcriptional profiles using the Smart-seq+5 protocol in single embryos for 7 pre-implantation development stages – ranging from oocytes to morula – across each of these different mammalian species, amounting to more than 330 samples at the time of writing. We aim to investigate the conservation of the oocyte and zygotic genome activation transcriptomes, as well as the transcriptional dynamics of transposable elements during this developmental window. ***However, because this work is still ongoing, in this thesis I will only briefly present an initial assessment of the conservation of transposable element expression during early development.***

Ancient transposable element families exhibit conserved expression patterns

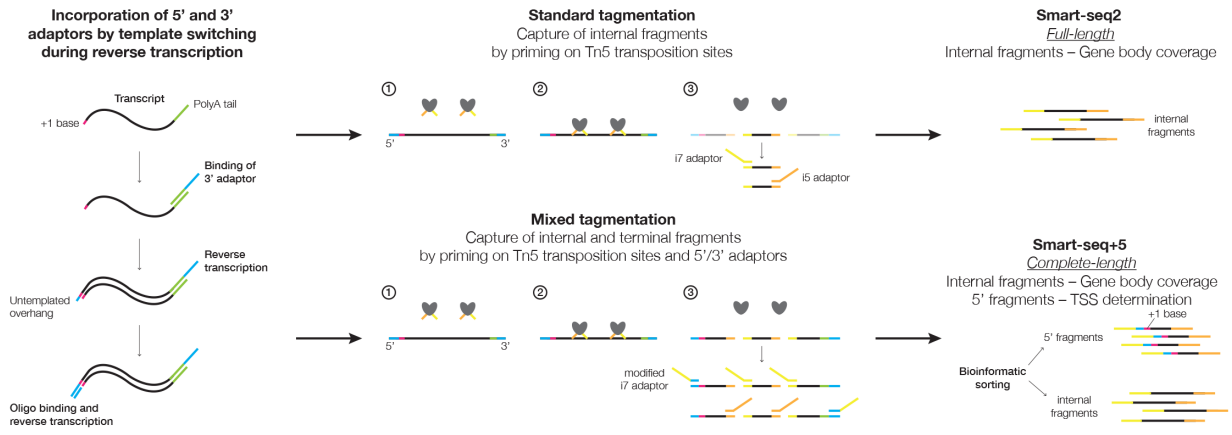
Because transposable element landscapes quickly diverge in-between species, assessing the degree of conservation of their expression dynamics during pre-implantation development must necessarily focus on older families preceding the speciation events. The L2 and L3 clades of LINE elements jointly constitute ~4% of the human genome and are amongst the most ancient transposable element groups found in mammals, having amplified approximately 200-300 mya (Pavlicek and Jurka, 2006), at the time of the radiation between reptiles, birds and mammals. Interestingly, despite their extraordinary age, we were able to detect transcription initiating within these elements using our 5' transcript profiling approach, suggesting that the regulatory sequences housed within have remained functional throughout this time. Indeed, with the possible exception of L2 in rabbit, both L2 and L3 appear to have retained a conserved expression pattern across the four species analyzed (Fig. 4A and 4B), exhibiting elevated expression levels in-between the oocyte and the species-specific timing of major zygotic genome activation. Similarly, the MIR family of SINEs – which is of a comparable age, and when active relied on L2 for its transposition – shows a similar expression pattern across these four species, being enriched in the oocyte, dropping at the zygote and further dropping off around the morula stage (Fig. 4C). L1 elements in mammalian genomes appear to evolve in a series of single non-branching lineages of successively dominant L1 families, among which the 80-150 million year old L1M4 family stands out for being one of the oldest (Lander et al., 2001; Smit et al., 1995),

and is present across the four species analyzed. Indeed, similarly to the other families described above, L1M4 also exhibited a remarkably consistent expression pattern that increases following the major zygotic genome activation event (Fig. 4D). Thus, ancient non-LTR elements appear to have retained a consistent expression pattern across the mammalian species assayed.

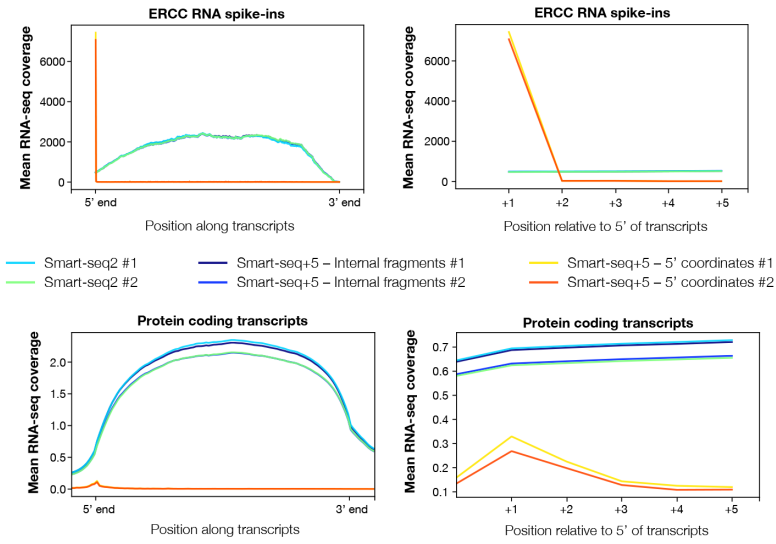
Stage-specific expression of DNA transposons

While numerous retrotransposon families have been ascribed stage-specific expression patterns and even functional roles in both mouse and human, the transcriptional dynamics of DNA transposons during pre-implantation development have remained so far unexplored. As before, we explored whether DNA transposons are expressed in pre-implantation embryos by analyzing the distribution of 5' reads generated through our approach. We identified stage-specific expression of various DNA transposons families across the four species, including the four representative examples shown in Fig. 4E. Further analyses are currently underway to identify orthologous DNA transposon families, assess the conservation of their expression and determine whether they have been co-opted into alternative promoters. Overall, our 5' transcriptional profiling approach permits us to examine the transcriptional dynamics of transposable elements during pre-implantation development, without the confounding influence of read-through transcription typically experienced in standard RNA-seq approaches.

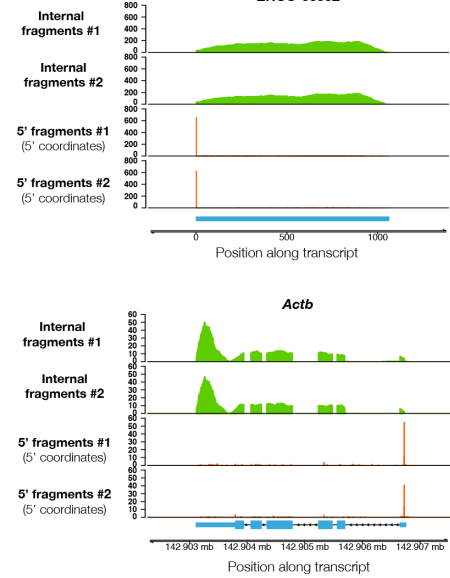
A



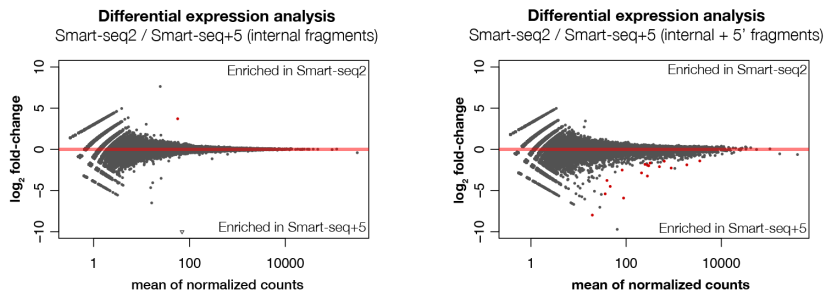
B



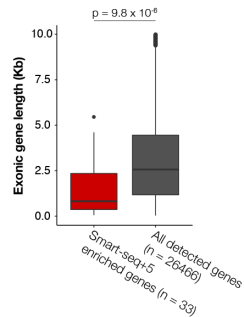
C



D



E



F

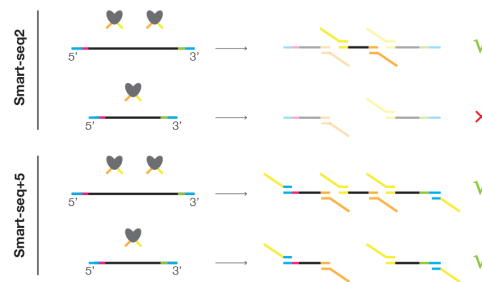
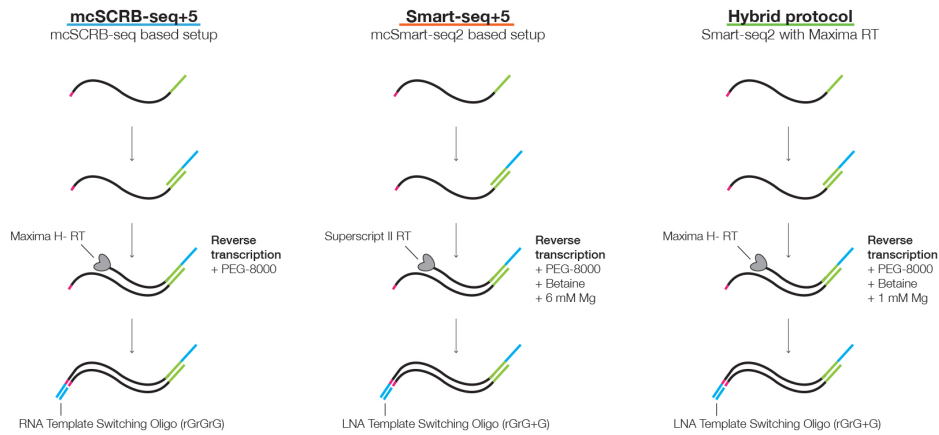
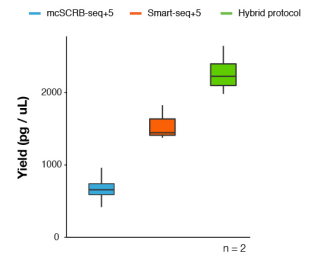


Figure 1. Identification of Transcription Start Sites from single-cell samples.

A) Schematic representation of the Smart-seq2 single-cell RNA-seq protocol and its molecular intermediates. Upon binding of an oligo-dT 3' adaptor to the transcript's poly-A tail, the target mRNA is reverse transcribed and incorporation of the 5' adaptor is performed through template switching of the reverse transcriptase. Following PCR amplification of full-length cDNAs, the sequencing library is prepared by incubating these cDNAs with adaptor-loaded Tn5 transposase. In the classical Smart-seq2 protocol, sequencing adaptors are incorporated by performing PCR using primers complementary to the adaptors inserted by the Tn5 transposase, thus omitting the terminal ends of the cDNA molecules, where there the transposase cannot bind. In the alternative approach presented here (Smart-seq+5), PCR is performed using primers complimentary to both the adaptors introduced by the transposase, and to the 3' and 5' adaptors used for reverse transcription and pre-amplification. It is thus possible to bioinformatically assign reads to the terminal or internal regions of transcripts based on the presence or absence of these adaptor sequences at the start of the read. B) Comparison of metagene profiles over protein coding transcripts or ERCC spike-ins using either the classical Smart-seq2 protocol or the internal or 5' fragments produced with the Smart-seq+5 protocol. C) Representative examples over one ERCC spike-in transcript and one endogenous gene. D) Differential expression analysis between the classical Smart-seq2 tagmentation protocol and the Smart-seq+5 tagmentation protocol, using either internal fragments only, or both 5' and internal fragments. E) Exonic gene length comparison between Smart-seq+5 genes and all detected genes. F) Smart-seq+5 might result in additional sensitivity by permitting the capture of short transcripts that are typically cleaved by a single transposase.

A**B****All protocols**

- Magnetic-bead purification of RNA
- Same adaptors (Smart-seq2 sequences)
- Molecular crowding (mc)
- Mixed tagmentation

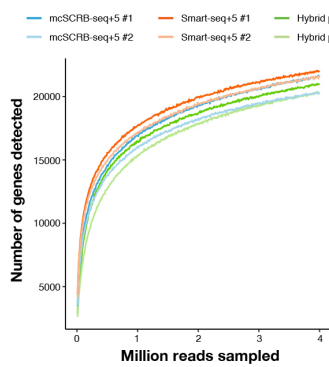
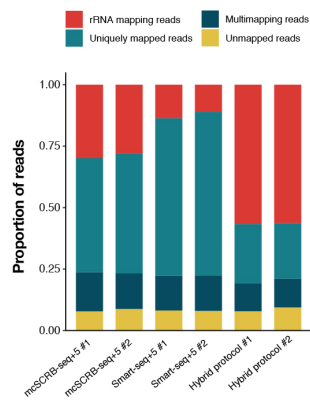
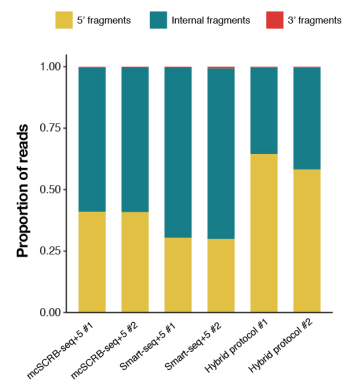
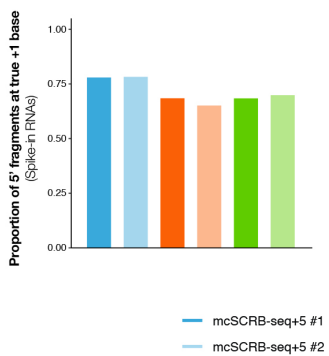
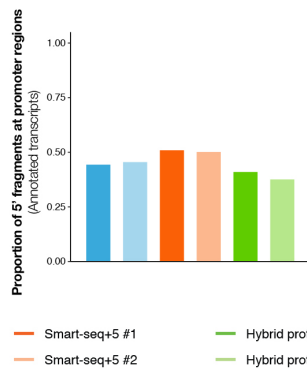
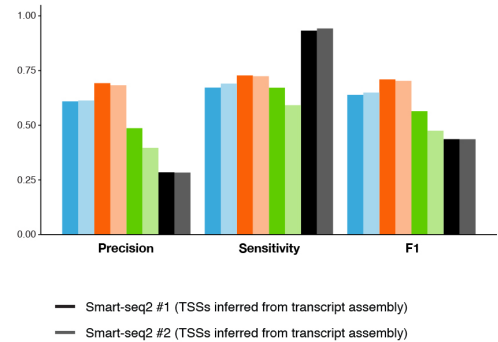
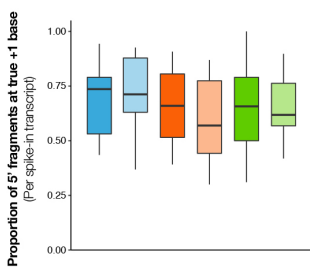
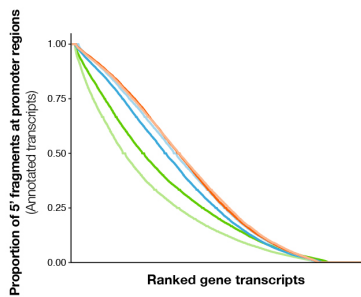
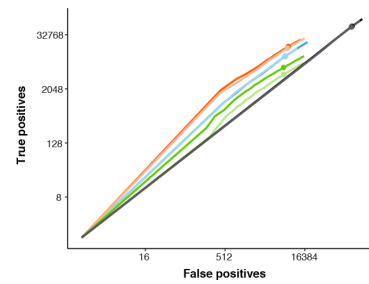
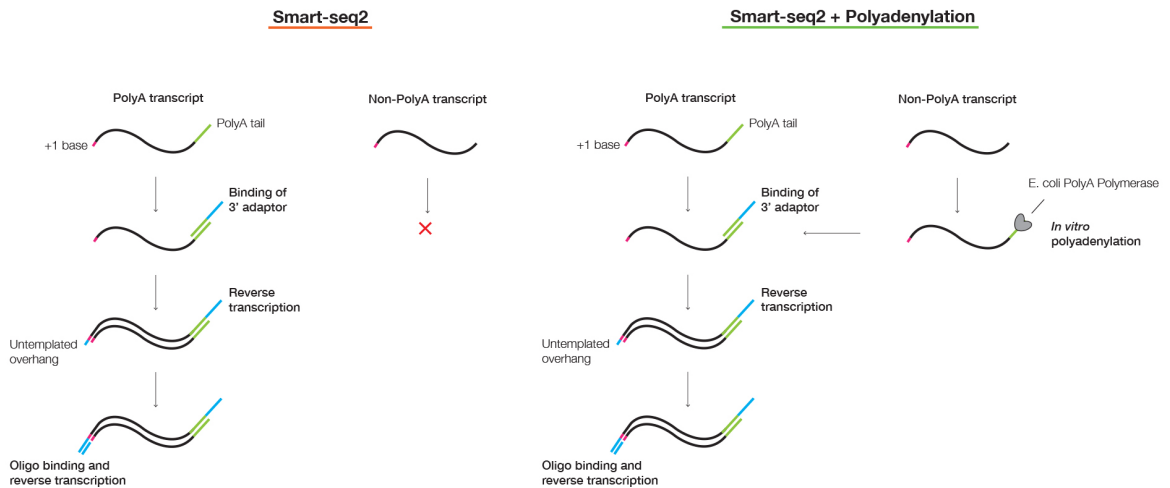
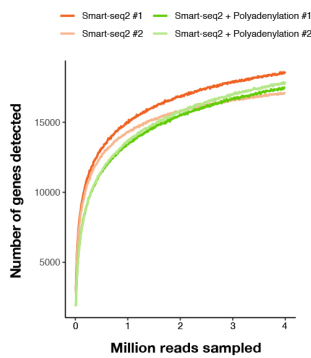
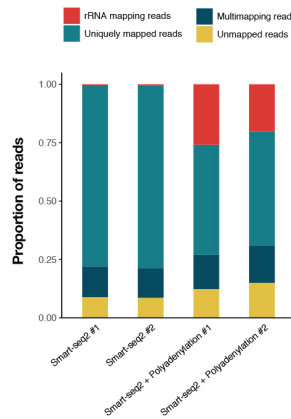
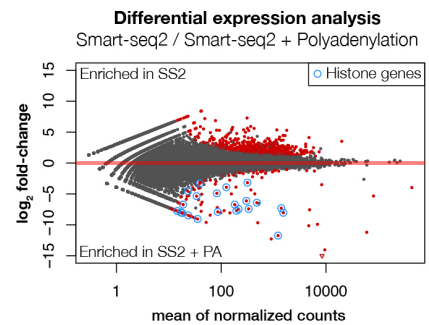
C**D****E****F****H****J****G****I****K**

Figure 2. Comparison between different single-cell RNA-seq protocols for sensitivity and 5' specificity.

A) Schematic representation of the different conditions employed for the reverse transcription reaction in each of the different protocols. B) cDNA yield obtained after pre-amplification in each of the different protocols. C) Number of genes detected at various estimated sequencing depths. D) Mappability profiles of each protocol, including proportion of rRNA reads. E) Proportion of reads in each protocol corresponding to terminal or internal fragments. F and G) Proportion of 5' fragments mapped to true 5' end of ERCC spike-in RNAs. H and I) Proportion of 5' fragments mapped to promoter regions of annotated transcripts. J) Precision, sensitivity and F1 statistic of TSS identification using the three different protocols compared to TSS identification using transcript assembly of reads derived using the vanilla Smart-seq2 protocol. K) Comparison of true positive versus false positive rates at different sequencing depths. Shown are the three protocol variants in addition to TSS identification using transcript assembly of reads derived using the vanilla Smart-seq2 protocol. Dot corresponds to performance at 5 million reads.

A**B****C****D****Figure 3.** Comparison between Smart-seq2 and Smart-seq2+PA

A) Schematic representation of the effect of the RNA polyadenylation reaction used in the Smart-seq2+PA variant. B) Number of genes detected at various estimated sequencing depths. C) Mappability profiles of each protocol, including proportion of rRNA reads. D) Differential expression analysis between the vanilla Smart-seq2 protocol and the Smart-seq2+PA variant. Significantly differentially expressed genes are shown in red, and non-polyadenylated histone genes are shown in blue.

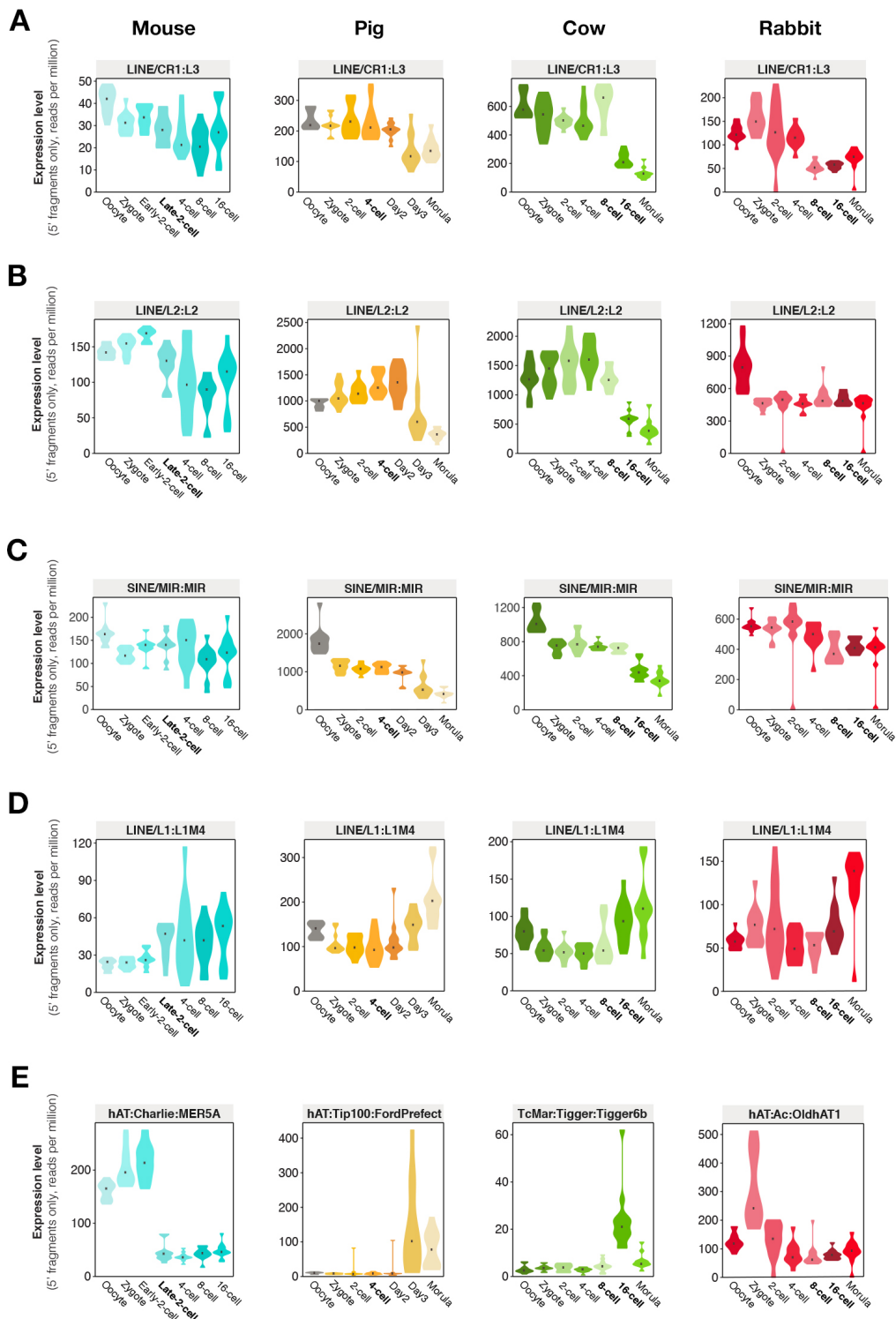


Fig 4. Comparison of transposable element expression dynamics during pre-implantation development across four mammalian species.

Violin plots representing the expression dynamics of individual transposable element families annotated by RepeatMasker. Only 5' fragments were used for this analysis. Multimapping reads were also considered if all reported alignments corresponded to different integrations of the same transposable element family. Bold font indicates the timing of major zygotic genome activation.

SUV4-20 activity in the preimplantation mouse embryo controls timely replication

Statement of contribution

I hereby state that my contribution to the publication:

Eid, A.*, **Rodriguez-Terrones, D.**, Burton, A., and Torres-Padilla, M.E. (2016).
SUV4-20 activity in the preimplantation mouse embryo controls timely replication.
Genes Dev. 30, 2513–2526.

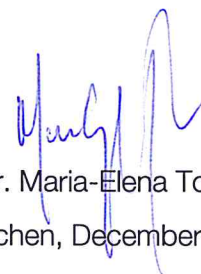
consisted in the computational analysis of gene expression data in preimplantation embryos.



Diego Rodriguez-Terrones
München, December 12th, 2018

Confirmation of contribution

I hereby confirm that the statement of contribution reproduced above is both truthful and accurate.



Prof. Dr. Maria-Elena Torres-Padilla
München, December 12th, 2018

SUV4-20 activity in the preimplantation mouse embryo controls timely replication

André Eid,¹ Diego Rodriguez-Terrones,² Adam Burton,^{1,2} and Maria-Elena Torres-Padilla^{1,2}

¹Institut de Génétique et de Biologie Moléculaire et Cellulaire, U964, Centre National de la Recherche Scientifique (CNRS)/ Institut National de la Santé et de la Recherche Médicale (INSERM), F-67404 Illkirch, CU de Strasbourg, France; ²Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, D-81377 München, Germany

Extensive chromatin remodeling after fertilization is thought to take place to allow a new developmental program to start. This includes dynamic changes in histone methylation and, in particular, the remodeling of constitutive heterochromatic marks such as histone H4 Lys20 trimethylation (H4K20me3). While the essential function of H4K20me1 in preimplantation mouse embryos is well established, the role of the additional H4K20 methylation states through the action of the SUV4-20 methyltransferases has not been addressed. Here we show that *Suv4-20h1/h2* are mostly absent in mouse embryos before implantation, underscoring a rapid decrease of H4K20me3 from the two-cell stage onward. We addressed the functional significance of this remodeling by introducing *Suv4-20h1* and *Suv4-20h2* in early embryos. Ectopic expression of *Suv4-20h2* leads to sustained levels of H4K20me3, developmental arrest, and defects in S-phase progression. The developmental phenotype can be partially overcome through inhibition of the ATR pathway, suggesting that the main function for the remodeling of H4K20me3 after fertilization is to allow the timely and coordinated progression of replication. This is in contrast to the replication program in somatic cells, where H4K20me3 has been shown to promote replication origin licensing, and anticipates a different regulation of replication during this early developmental time window.

[*Keywords:* mouse embryo; Suv4-20; heterochromatin; replication stress]

Supplemental material is available for this article.

Received August 11, 2016; revised version accepted November 7, 2016.

Fertilization of the oocyte by the sperm results in the formation of a totipotent zygote that has the ability to generate all extraembryonic and embryonic tissues necessary for development. The earliest stages of development prior to implantation are of critical importance for setting up the first embryonic lineages to generate an embryo competent for implantation. Therefore, a central question in biology is the defining role of the organization of the chromatin and its architecture during the first cell divisions and how they enable changes in cellular plasticity and fate.

In mice, preimplantation development is characterized by a distinctive atypical state of chromatin signatures, with many histone post-translational modifications (PTMs) reduced or absent after fertilization. In addition, the paternal and maternal chromatin remain physically segregated in two separate pronuclei that maintain distinctive chromatin marks, with the maternal chromatin containing most constitutive heterochromatin PTMs, while the paternal chromatin is enriched with PTMs of facultative heterochromatin, thought to substitute for the absence of constitutive heterochromatin therein

(for review, see Burton and Torres-Padilla 2014). The histone H4 Lys20 dimethylation and trimethylation (H4K20me2/3) are hallmarks of constitutive heterochromatin in somatic cells. H4K20me3 localizes primarily at centromeres, pericentromeres, and telomeres that are enriched in repetitive sequences and are gene-poor (Schotta et al. 2008). This is in contrast to H4K20 monomethylation (H4K20me1), which is one of the most abundant modifications on H4 and localizes to a wide variety of genomic regions in a cell cycle-dependent manner across somatic cells and in the early embryo (Barski et al. 2007; Jorgensen et al. 2007; Houston et al. 2008; Oda et al. 2010). While, in yeast, one single enzyme catalyzes the three H4K20 methylation states, in mammals, H4K20me1 is catalyzed by PR-Set7, and H4K20me2 and H4K20me3 are both catalyzed by the histone methyltransferases SUV4-20H1 and SUV4-20H2 (*Kmt5b* and *Kmt5c*, respectively) (Nishioka et al. 2002; Rice et al. 2002; Schotta et al. 2004, 2008). In the zygote, immediately after fertilization, H4K20me3 is detected exclusively on the maternal pronucleus, where

Corresponding author: torres-padilla@helmholtz-muenchen.de

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.288969.116>.

© 2016 Eid et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

it appears distributed mostly around ring-like structures formed by the nucleoli precursors (nucleolar-like bodies [NLBs]), which harbor the pericentromeric and centromeric regions (Probst et al. 2007; Wongtawan et al. 2011). Importantly, H4K20me3 is undetectable from the two-cell stage onward and remains so until the peri-implantation period (Wongtawan et al. 2011).

This transient loss of H4K20me3 is perplexing and raises two important questions. The only other cell types displaying the absence of H4K20me3 seem to be cancer cells with increased pluripotent capacity and proliferation activity, resulting in poor prognosis for patients (Fraga et al. 2005; Van Den Broeck et al. 2008; Schneider et al. 2011; Yokoyama et al. 2014). It is thus essential to understand how fluctuations in H4K20me3 levels impact cell proliferation and cellular potency. In addition, the lack of conventional constitutive heterochromatin in zygotes and two-cell stage embryos has been linked to their characteristic nuclear organization and high chromatin dynamics, believed to support a higher developmental plasticity. However, whether changes in this atypical heterochromatin configuration play a functional role in developmental plasticity beyond a mere correlation has not been addressed.

Because the absence of H4K20me3 correlates with the highest developmental potency, we hypothesize that lack of H4K20me3 is required for zygotic reprogramming to take place. We thus set out to address whether enforcing the maintenance of H4K20me3 during preimplantation restricts developmental capacity and investigate how H4K20me3 relates to cellular proliferation *in vivo*. For this, we first determined that *Suv4-20h1* and *Suv4-20h2* are only weakly expressed after fertilization. Accordingly, in order to achieve sustained maintenance of H4K20me3 throughout preimplantation development, we ectopically expressed *Suv4-20h1* and *Suv4-20h2*. Our results show that ectopic expression of *Suv4-20h2* is sufficient to enable global levels of H4K20me3. *Suv4-20h2* displayed a markedly higher ability to restore H4K20me3 than *Suv4-20h1*, independently of changes in H3K9me3. Embryos expressing *Suv4-20h2*—but not *Suv4-20h1*—do not efficiently develop beyond the two-cell stage, indicating that the remodeling of H4K20me3 is required for developmental progression. *Suv4-20h2* expression led to a proliferation defect accompanied by replication abnormalities. Importantly, the developmental phenotype was partially rescued by inhibition of the ATR pathway, suggesting that H4K20me3 induces replication stress and S-phase arrest. Our results shed light on the functional role of the absence of H4K20me3 during preimplantation development and suggest that, in contrast to somatic cells, H4K20me3 is incompatible with the timely progression of DNA replication of embryonic chromatin.

Results

Expression of H4K20 modifiers during preimplantation development

SUV4-20H1 and SUV4-20H2 are the two mammalian homologs of *Drosophila* Set8. In mammals, SUV4-20H2

has a slight preference for H4K20me3, but the combined knockout of *Suv4-20h1/h2* completely abolishes H4K20me3 (Schotta et al. 2004, 2008), indicating that they are the major H4K20me3 methyltransferases in mammalian cells. We thus analyzed the expression of both genes by RT-qPCR in all stages of preimplantation development. The pattern of expression of both *Suv4-20h1* and *Suv4-20h2* resembles that of maternally inherited transcripts, with higher levels in the zygote and a reduction at or after the two-cell stage (Supplemental Fig. S1). However, both enzymes are expressed at very low levels compared with the control housekeeping gene (*Actinb*), with *Suv4-20h1* exhibiting lower levels of expression than *Suv4-20h2* (Supplemental Fig. S1). A third enzyme, *Smyd5*, was reported to methylate H4K20 *in vitro* (Stender et al. 2012), although the contribution of SMYD5 to global H4K20me3 levels *in vivo* is unclear. We found that, in contrast to the two *Suv4-20* enzymes, the expression of *Smyd5* is strongly induced from the two-cell stage onward and is expressed continuously thereafter (Supplemental Fig. S1). Given the strong expression of SMYD5 during these developmental time periods, when H4K20me3 is undetectable on embryonic chromatin (Wongtawan et al. 2011), it is unlikely that SMYD5 contributes to the global remodeling of H4K20me3 after fertilization. Note that there are no specific antibodies available for SUV4-20H1, SUV4-20H2, or SMYD5 (our unpublished observations), and therefore our analysis for these three enzymes focuses on mRNA exclusively. To date, only one demethylase has been shown to be able to act on H4K20me3 *in vitro*: PHF2, which can also demethylate H3K9me1 (Wen et al. 2010; Stender et al. 2012). RT-qPCR showed that the mRNA for *Phf2* is abundant in the zygote, in comparison with later stages, as *Phf2* is practically absent from the eight-cell stage onward (Supplemental Fig. S1), suggesting that *Phf2* mRNA is inherited maternally but degraded after fertilization. Immunostaining revealed that PHF2 is present throughout all stages of preimplantation development concomitantly with the absence of H4K20me3 (data not shown). While PHF2 may contribute toward keeping H4K20me3 practically absent from the embryonic chromatin, the results above suggest that low H4K20me3 levels throughout the cleavage stages is in part due to low expression of SUV4-20 methyltransferases.

Expression of Suv4-20h2 results in accumulation of H4K20me3

Given the above results, in order to maintain sustained H4K20me3 during preimplantation development, we chose to ectopically express *Suv4-20h2* in zygotes, in particular because manipulating PHF2 levels may also directly affect H3K9me1. Zygotes were microinjected with mRNA for HA-tagged *Suv4-20h2* in combination with mRNA for *GFP* as an injection control (Fig. 1A). Control groups included embryos injected with mRNA for *GFP* alone as well as noninjected embryos. Embryos were cultured until the late zygote stage and analyzed by immunofluorescence using an HA-antibody, which

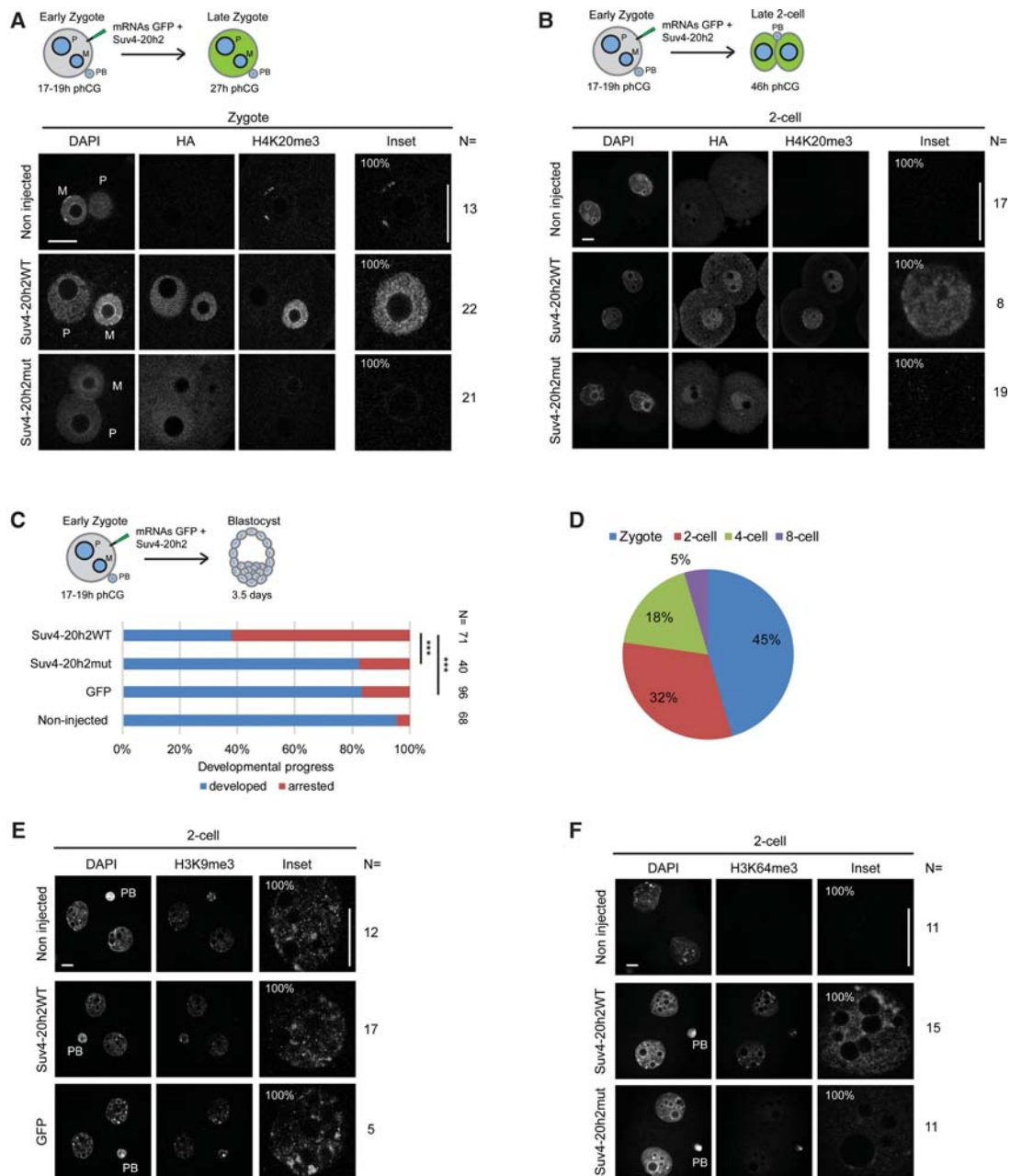


Figure 1. Maintenance of H4K20me3 through *Suv4-20h2* ectopic expression blocks embryonic development prior to the two-cell stage. (A) A schematic representation of the experimental design is shown at the top. Zygotes between 17 and 19 h post-human chorionic gonadotropin (phCG) were microinjected with mRNA *Suv4-20h2WT* or *Suv4-20h2mut* in addition to *GFP*, cultured in KSOM (K-modified simplex optimized medium), and then fixed in 4% formaldehyde (as indicated in the Material and Methods) at 27 h phCG. Representative images showing single Z-projections of confocal sections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected zygotes stained with DAPI, HA, and H4K20me3 antibodies. An inset of the maternal pronucleus is shown in the right panels. N numbers are indicated at the right. (M) Maternal, (P) paternal. Bars: inset, 5 μ m; all others, 10 μ m. (B) Microinjections were performed as in A, except that embryos were fixed at 46 h phCG. Representative images showing single Z-projections of confocal sections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected two-cell stage embryos stained with DAPI, HA, and H4K20me3 antibodies. An inset of one of the two nuclei is shown in the right panels. N numbers are indicated at the right. Bars: inset, 5 μ m; all others, 10 μ m. (C) Zygotes were microinjected as in A and cultured until the blastocyst stage. The number of embryos reaching the morula stage (developed) was quantified for noninjected, *GFP* alone-injected, *Suv4-20h2mut*-injected, and *Suv4-20h2WT*-injected embryos. Total numbers of embryos are indicated next to the plot. Statistical testing was performed using the N-1 two-proportion test for comparing independent proportions. (***) $P < 0.0001$. (D) Pie chart showing the distribution of arrested *Suv4-20h2WT*-injected embryos by stage. (E,F) Zygotes were microinjected as in A and analyzed with an H3K9me3 (E) or H3K64me3 (F) antibody at the two-cell stage. Representative images showing single Z-projections of confocal sections of noninjected, *Suv4-20h2WT*-injected, and *GFP*-injected or *Suv4-20h2mut*-injected embryos. An inset of one of the two nuclei is shown in the right panels. N numbers are indicated. Bars: inset, 5 μ m; all others, 10 μ m. (PB) Polar body.

revealed that SUV4-20H2 was efficiently translated and localized to both maternal and paternal pronuclei (Fig. 1A). In nonmanipulated embryos, H4K20me3 is detected only around the NLBs and in the nuclear periphery at 4'-6-diamidino-2-phenylindole (DAPI)-rich regions in the maternal pronucleus and is undetectable in the paternal chromatin (Fig. 1A, noninjected) (Kourmouli et al. 2004; Wongtawan et al. 2011). Expression of *Suv4-20h2* resulted in a clear increase in H4K20me3 levels in the maternal pronucleus but not in the paternal pronucleus (Fig. 1A). This observation was surprising considering that SUV4-20H2 was distributed equally between both pronuclei and suggests that SUV4-20H2 is unable to modify the levels of H4K20me3 on the paternal chromatin in zygotes, perhaps due to the absence of H3K9me3 (Lange et al. 2013). At the two-cell stage, SUV4-20H2 as well as H4K20me3 were readily detected in the nuclei of both blastomeres at levels comparable with those in zygotes (Fig. 1B; Supplemental Fig. S1D). This was in contrast to noninjected embryos, where there were no detectable levels of H4K20me3, in agreement with earlier findings (Fig. 1B). The distribution of H4K20me3 throughout two-cell stage nucleus, as opposed to only half of the nucleus, indicates that SUV4-20H2 can methylate both paternal and maternal chromosomes at this stage, suggesting that methylation on the paternal chromatin is delayed in comparison with the maternal one, which takes place in the zygote.

To test for the specific effects of the methyltransferase activity of SUV4-20H2, we generated a mutant in the SET domain. The SET domain is shared across several other histone methyltransferases, including SUV3-9H1, in which a mutation in the amino acid sequence (NHDC) abrogates its catalytic activity (Rea et al. 2000; Lachner et al. 2001). We therefore generated a SUV4-20H2 construct in which the corresponding NHDC motif was replaced by AAAG. The resulting mutant protein is referred to here as SUV4-20H2MUT, while the wild type is referred to here as SUV4-20H2WT. We microinjected early zygotes as above with mRNA for *Suv4-20h2mut* and *GFP* and analyzed embryos at the late zygote stage. The AAAG mutation did not affect the localization of SUV4-20H2MUT (which remained evenly distributed in both pronuclei) but efficiently abolished the methyltransferase activity of SUV4-20H2, since expression of SUV4-20H2MUT did not lead to an increase in H4K20me3 levels in either the maternal or paternal pronucleus (Fig. 1A). Likewise, levels of H4K20me3 remained low in two-cell stage embryos expressing SUV4-20H2MUT, similar to control embryos (Fig. 1B).

Our results show that SUV4-20H2WT efficiently increases H4K20me3 levels *in vivo*, allowing us to sustain H4K20me3 levels during preimplantation development.

H4K20me3 is incompatible with preimplantation development

We next addressed whether embryos displaying sustained H4K20me3 can develop normally. Embryos were microinjected with mRNA for *Suv4-20h2WT* as above and cultured for 3 d until the blastocyst stage. As controls,

we used noninjected embryos, embryos microinjected with *GFP* mRNA alone, and embryos expressing *Suv4-20h2MUT* in combination with *GFP*. Control embryos showed robust development, with 95.5%, 83.3%, and 82.5% developing to the morula stage for the noninjected embryos ($n = 68$) or embryos expressing *GFP* ($n = 96$) and *SUV4-20H2mut* ($n = 40$), respectively (Fig. 1C). These percentages reflect typical developmental rates obtained in these assays (Santenard et al. 2010; Jachowicz et al. 2013). In contrast, embryos expressing SUV4-20H2WT displayed a strikingly lower developmental rate (38%; $n = 71$) (Fig. 1C). Since the SUV4-20H2MUT did not exhibit a developmental delay, this indicates that the developmental phenotype of the SUV4-20H2WT embryos is due to its histone methyltransferase activity. Thus, we conclude that the embryonic arrest observed for SUV4-20H2WT embryos is most likely due to the presence of H4K20me3, suggesting that the removal of this heterochromatic mark is a requisite for preimplantation development.

Next, we asked whether the developmental phenotype upon expression of SUV4-20H2WT at the zygote stage is specific to that stage exclusively. For this, we asked whether expression of SUV4-20H2WT at a different stage results in a similar cellular arrest. We introduced mRNA for *Suv4-20h2WT* in combination with *GFP*—or with mRNA for *GFP* alone as a negative control—in a single two-cell stage blastomere (Supplemental Fig. S2A). Expression of SUV4-20H2WT led to an increase in H4K20me3 levels in the injected cell in two-cell embryos (Supplemental Fig. S2B), which developed until the blastocyst stage. However, a detailed analysis of the number of cells in these embryos revealed a more limited cell progeny in SUV4-20H2WT-expressing blastomeres compared with the negative control (*GFP*-only) (Supplemental Fig. S2C), indicating that expression of SUV4-20H2WT in two-cell embryos leads to a cellular proliferation defect. Immunostaining of these embryos showed that cell arrest was often accompanied by nuclear fragmentation (Supplemental Fig. S2D). In conclusion, SUV4-20H2-mediated H4K20me3 leads to cell proliferation arrest in preimplantation embryos independently of the developmental stage.

Sustained H4K20me3 results in developmental failure prior to the two-cell stage

To understand the mechanism behind the developmental arrest in *Suv4-20h2WT*-expressing embryos, we dissected the developmental stages at which embryos arrest. Most embryos arrested at the zygote and two-cell stages (45% and 32%, respectively). The distribution of arrested embryos across preimplantation development suggests that H4K20me3 affects the earlier stages of development during which epigenetic reprogramming takes place. Thus, we first addressed whether maintenance of H4K20me3 perturbs other heterochromatic marks in zygotes and two-cell stage embryos. Analysis of H3K9me3 revealed no global differences between noninjected, *GFP*-expressing, or *Suv4-20h2WT*-expressing embryos (Fig. 1E; Supplemental Fig. S3A), in line with the suggested model for

heterochromatin establishment in which SUV39H1/H2 acts upstream of SUV4-20H1/H2 (Schotta et al. 2004). Thus, increased H4K20me3 occurred without global changes in H3K9me3, allowing us to distinguish between phenotypic effects of the typical “H3K9me3-directed” heterochromatin and those effects specific to changes in H4K20me3.

The distribution of H3K64me3, another PTM of constitutive heterochromatin, strongly resembles that of H4K20me3 (Daujat et al. 2009). H3K64me3 is present in the maternal pronucleus but is undetectable from the two-cell stage onward. In agreement with this, H3K64me3 was undetectable in noninjected embryos at the two-cell stage (Fig. 1F). In contrast, SUV4-20H2WT embryos showed a marked increase in H3K64me3 in two-cell stage nuclei, in contrast to SUV4-20H2MUT embryos, where H3K64me3 was not detected (Fig. 1F; Supplemental Fig. S3B). This observation is surprising considering that double *Suv4-20h1/Suv4-20h2* knockout mouse embryonic fibroblasts (MEFs) retain H3K64me3, which had led to the suggestion that H3K64me3 occurs independently of SUV4-20 activity (Lange et al. 2013). Thus, the interplay between H4K20me3 and H3K64me3 in the embryo may obey different regulatory mechanisms than in somatic cells, or, alternatively, de novo acquisition of H4K20me3 may affect H3K64me3.

The developmental failure elicited by H4K20me3 is mediated by SUV4-20H2

Suv4-20h1 is the second mammalian homolog of Set8 and is expressed only weakly in the early embryo (Supplemental Fig. S1). To address whether the embryonic arrest observed following H4K20me3 maintenance is specific to the histone methyltransferase activity of SUV4-20H2, we undertook the same experimental approach as above with SUV4-20H1. When ectopically expressed, SUV4-20H1 displayed a nuclear localization similar to that of SUV4-20H2 at the zygote stage. However, SUV4-20H1 did not detectably increase levels of H4K20me3 at the zygote stage (Fig. 2A), and H4K20me3 levels increased only weakly at the two-cell stage (Fig. 2B), indicating that, in vivo, in the embryo, the catalytic activity toward H4K20me3 is higher for SUV4-20H2 than for SUV4-20H1. This is in agreement with previous suggestions from crystallography work (Southall et al. 2014). The low histone methyltransferase activity of SUV4-20H1 was lost upon mutating the NHDC sequence of its SET domain into AAAG, similar to SUV4-20H2 (Fig. 2B, SUV4-20H1MUT). Next, we monitored development of embryos expressing SUV4-20H1, similar to that of SUV4-20H2 (Fig. 2C). We microinjected early zygotes with either *Suv4-20h1WT* or *Suv4-20h1MUT* mRNA in combination with mRNA for *GFP* or with mRNA for *GFP* alone. Similarly, embryos expressing *Suv4-20h1WT* did not show a significant change in developmental progression ($n = 78$) (Fig. 2C). Indeed, their development rate to the blastocyst stage (66%) was not significantly different from that of embryos expressing *GFP* alone (77%; $P = 0.16$). These observations indicate that the increase in

H4K20me3 from the zygote to the two-cell stage is a primary cause of the embryonic arrest and is mediated mainly by SUV4-20H2 histone methyltransferase activity.

The effect of SUV4-20H2 on gene expression during embryonic genome activation (EGA)

Deposition of H4K20me3 through SUV4-20H2 can cause RNA polymerase II (Pol II) pausing. H4K20me3 has also been shown to repress gene expression in transformed cell lines by reducing H4K16 acetylation (H4K16ac) levels (Kapoor-Vazirani et al. 2011). In mice, the first wave of EGA takes place in the zygote, and the second wave, with a higher transcriptional activity, takes place at the two-cell stage (Flach et al. 1982; Bensaude et al. 1983; Aoki et al. 1997). Given that most SUV4-20H2 embryos arrested at the zygote and two-cell stages, we thus investigated whether the increase in H4K20me3 resulted in suppression of EGA in these embryos, which could potentially explain the developmental arrest. To evaluate global levels of gene expression, we pulsed embryos with EU (5-ethynyl uridine) at the late two-cell stage at a time point corresponding to the late phase of transcription during EGA (Fig. 3A). Noninjected and SUV4-20H2MUT embryos showed a similar distribution of EU pattern, indicating that SUV4-20H2MUT does not compromise EGA, with 85% ($n = 13$) and 94% ($n = 18$) of embryos displaying active transcription, respectively (Fig. 3A,B). A significant proportion of two-cell stage embryos expressing SUV4-20H2WT (61%, $n = 18$) also displayed active transcription (Fig. 3B). Nevertheless, it seems that most embryos displayed lower transcriptional activity, since only 22% of SUV4-20H2 embryos showed high levels of EU incorporation, in comparison with >50% for the noninjected and SUV4-20H2MUT groups (Fig. 3B). This suggests that sustained H4K20me3 does not prevent EGA but has some impact on global levels of transcription at the two-cell stage, which could partly explain the developmental block. Thus, to address whether SUV4-20WT embryos display specific changes in gene expression as opposed to a global or delayed effect on EGA, we profiled gene expression in individual embryos using a microfluidics Biomark approach, a robust and quantitative approach amenable to gene expression analysis from low amounts of material (Guo et al. 2010). We examined the expression of 45 genes in individual embryos across the four experimental groups: noninjected, SUV4-20H1WT, SUV4-20H2WT, and SUV4-20H2MUT (Supplemental Fig. S5). The 45 genes analyzed were selected on the basis of (1) their known transcriptional activation at EGA, (2) their role in developmental progression and signaling during early development, and (3) their role in blastocyst formation (Supplemental Table S1). Principal component analysis (PCA) and unsupervised clustering analysis of these data revealed that expression of SUV4-20H2 did not induce significant changes in expression of the genes assessed, compared with the three control groups (Fig. 3C; Supplemental Fig. S5; Supplemental Table S2). Because most SUV4-20H2WT embryos displayed transcriptional activity, albeit at reduced rates, and the Biomark results

Eid et al.

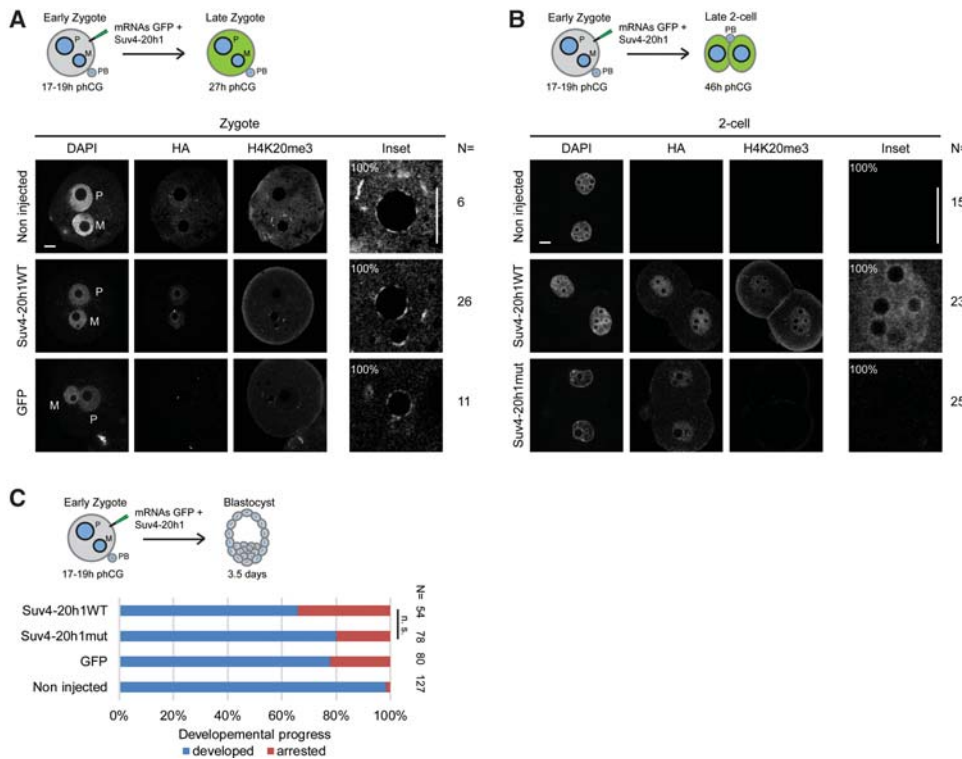


Figure 2. Ectopic expression of *Suv4-20h1* does not arrest embryonic development or increase H4K20me3 levels in zygotes. (A) Schematic representation of the experimental design. Zygotes were microinjected with mRNA for *Suv4-20h1WT* or *Suv4-20h1mut*, cultured, and fixed for immunostaining using HA or H4K20me3 antibodies at 27 h phCG. Representative images show single Z-projections of confocal sections of noninjected, *Suv4-20h1WT*-injected, and *GFP*-injected zygotes. An inset of the maternal pronucleus is shown in the right panels. *N* numbers are indicated. (M) Maternal; (P) paternal. Bars: inset, 5 μ m; all others, 10 μ m. (B) Zygotes were microinjected with mRNA for *Suv4-20h1WT* or *Suv4-20h1mut* as in A, cultured, and analyzed at the two-cell stage. Representative images show single Z-projections of confocal sections of noninjected, *Suv4-20h1WT*-injected, and *Suv4-20h1mut*-injected embryos stained with DAPI, HA, and H4K20me3 antibodies. An inset of one of the two nuclei is shown in the right panels. *N* numbers are indicated. Bars: inset, 5 μ m; all others, 10 μ m. (C) Zygotes were microinjected with mRNA for *Suv4-20h1WT* or *Suv4-20h1mut* in combination with *GFP* and cultured until the blastocyst stage. The percentage of embryos that reached the blastocyst stage (developed) is plotted for noninjected, *GFP*-injected, *Suv4-20h1mut*-injected, and *Suv4-20h1WT*-injected embryos. The total number of embryos analyzed is indicated. Statistical testing was performed using the *N*-1 two-proportion test for comparing independent proportions. (n.s.) Not significant.

indicate that changes in specific gene expression were not significant, our observations altogether could also suggest a delay in the onset of transcriptional activation. Because the timing of transcriptional activation in embryos is closely related to that of replication, it therefore remains possible that the reduced transcriptional activity in *SUV4-20H2WT* embryos reflects a delay and/or a defect in S-phase progression.

H4K20me3 perturbs developmental progression through replication in zygotes and two-cell stage embryos

Previous reports have indicated that H4K20 methylation levels may play a role in the control of replication timing and origin licensing (Tardat et al. 2007, 2010; Oda et al. 2010; Vermeulen et al. 2010). Thus, we investigated whether the developmental arrest observed upon expression of *SUV4-20H2WT* could be due to a misregulation of S phase. For this, we subjected embryos to an EdU (5-ethynyl-2'-deoxyuridine) pulse at the mid two-cell

stage, the time at which most embryos are known to be in late S phase (Bolton et al. 1984). Because it is known that replication does not proceed synchronously across embryos, we further scored replication patterns as "late" or "early" according to whether they reflect a late S phase (with low levels of EdU detected at the NLBs or the nuclear periphery) or an early-mid-S phase (where replication foci are visible and evenly distributed in the nucleus), respectively (Fig. 3D). As expected, most control embryos displayed a late replication pattern at this time point (21 out of 25) (Fig. 3D,E). Similarly, albeit with some delay presumably due to the microinjection procedure, most *SUV4-20H2MUT* embryos also displayed a late replication pattern (12 out of 20) (Fig. 3D,E). *SUV4-20H2WT* embryos instead mostly showed an early replication pattern, with 85% of embryos displaying high levels of EdU incorporation ($n = 13$) (Fig. 3D,E). This observation points toward a misregulation of S-phase progression in *SUV4-20H2WT* embryos and prompted us to further investigate the timing of S-phase initiation and completion.

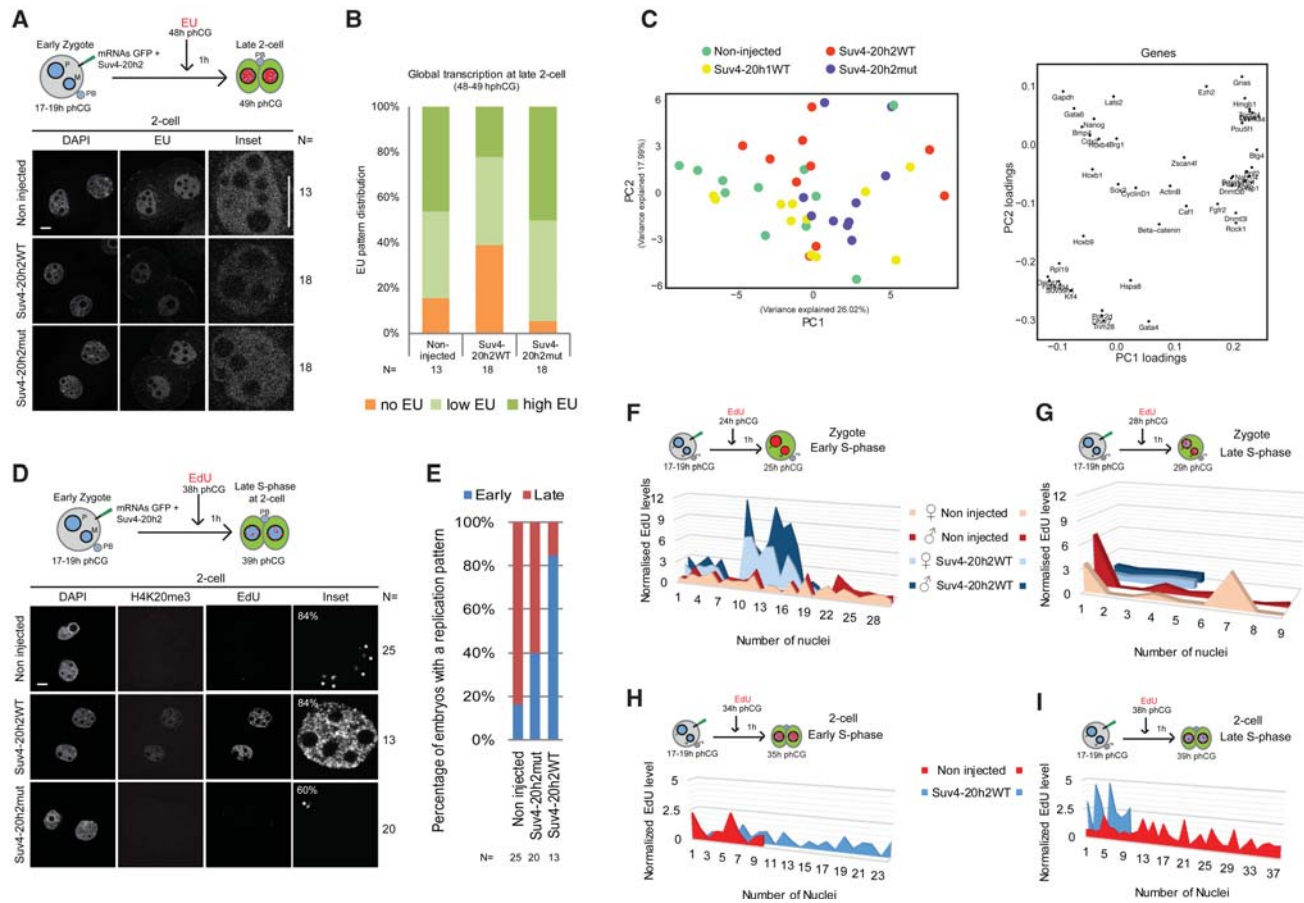


Figure 3. H4K20me3 affects replication progression. (A) Zygotes were microinjected with mRNA *Suv4-20h2WT* or *Suv4-20h2mut*, cultured until the two-cell stage, and pulsed with EU for 1 h at 48 h phCG. Representative images show single Z-projections of confocal sections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected embryos stained with DAPI and EU visualized by Click-iT reaction. An inset of one of the two nuclei is shown at the right. N numbers are indicated. Bars: inset, 5 μm; all others, 10 μm. (B) Distribution of EU patterns in late two-cell stage embryos. Embryos were divided into three groups based on their EU pattern: (1) no transcription (no EU detected), (2) low levels of transcription (as exemplified in A, *Suv4-20h2WT*), and (3) high levels of transcription (as shown in A, noninjected and *Suv4-20h2mut*). (C, right) Principal component (PC) projection of individual two-cell stage embryos based on the expression profiles of 45 genes detailed in Supplemental Table 1. Embryos are colored according to the experimental group. The first component (PC1) is shown on the X-axis, and the second component (PC2) is shown on the Y-axis. (Left) Principal component projection of gene loadings showing the weight of each gene in the analysis. (D) Zygotes microinjected with mRNA *Suv4-20h2WT* or *Suv4-20h2mut* were cultured until the two-cell stage, pulsed with 5-ethynyl-2'-deoxyuridine (EdU) for 1 h at 38 h phCG, fixed, and analyzed for EdU incorporation. Representative images showing single Z-projections of confocal sections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected embryos. An inset of one of the two nuclei is shown at the right. White asterisks indicate EdU labeling in noninjected and *Suv4-20h2mut*-injected embryos. N numbers are indicated. Bars: inset, 5 μm; all others, 10 μm. (E) Distribution of the replication patterns based on EdU labeling at 38–39 h phCG as shown in D. (F, G) Normalized EdU levels measured in each pronucleus of noninjected and *Suv4-20h2WT*-injected zygotes at 24–25 h phCG (F) and 28–29 h phCG (G). (H, I) Normalized EdU levels measured in each nucleus of noninjected and *Suv4-20h2WT*-injected two-cell stage embryos at 34–35 h phCG (H) and 38–39 h phCG (I).

We performed EdU labeling at four different time points that correspond to the onset and completion of S phase in zygotes and two-cell stage embryos (early and late, respectively). Because of the limited number of embryos available per experiment, in these experiments, we used only noninjected embryos as negative controls. At the onset of replication in zygotes, all control embryos had undertaken replication and showed stable levels of EdU incorporation (Fig. 3F), with 80% of embryos having finished replication by 29 h post-human chorionic gonadotropin (phCG) injection (Fig. 3G). In contrast, while most

SUV4-20H2WT embryos undertook replication in a timely fashion (90%; $n = 20$), the majority of them (68%) showed significantly higher levels of EdU incorporation than noninjected embryos in both pronuclei (Fig. 3F), and all embryos maintained robust levels of EdU incorporation at 29 h phCG (Fig. 3G). EdU levels in embryos expressing GFP alone at the late zygote stage were not affected, compared with noninjected embryos (Supplemental Fig. S3C). These experiments suggest that while *SUV4-20H2WT* embryos enter S phase at a time similar to that of the controls, S-phase progression is delayed.

Likewise, at the two-cell stage, control and SUV4-20H2WT embryos displayed similar levels of replication foci at the start of S phase (35 h phCG) (Fig. 3H). However, SUV4-20H2WT embryos showed persisting high levels of EdU incorporation at 39 h phCG, while noninjected embryos had mostly completed S phase at this time (Fig. 3I). These observations indicate a strong effect on S-phase progression in embryos with sustained H4K20me3 levels at the zygote and two-cell stages. Importantly, this phenotype correlates well with the timing and distribution of the embryonic arrest elicited upon ectopic expression of SUV4-20H2 (Fig. 1D).

Previous reports have shown that changes in the levels of H4K20me1 and the expression of PR-Set7 lead to an intra-S-phase checkpoint activation and cell cycle arrest (Tardat et al. 2007; Houston et al. 2008; Oda et al. 2009). Since H4K20 methylation is processive (Sims et al. 2006; Congdon et al. 2010), it is possible that the increased H4K20me3 levels upon ectopic expression of SUV4-20H2WT have repercussions on the levels of H4K20me1, and thus the developmental phenotype and the misregulation of S phase may be due to changes in H4K20me1 rather than an increase in H4K20me3. To address this, we analyzed levels of H4K20me1 in control embryos as well as embryos expressing SUV4-20H2WT or SUV4-20H2MUT at the two-cell stage during G2 phase, when H4K20me1 levels are highest (Fig. 4A). Expression of SUV4-20H2WT—but not of SUV4-20H2MUT—led to a reduction in H4K20me1 levels, suggesting that H4K20me1 is indeed used as a substrate for the higher methylation state (Fig. 4A; Supplemental Fig. S4A). To distinguish whether changes in H4K20me1 levels are related to the developmental phenotype elicited upon SUV4-20H2 expression, we analyzed H4K20me1 in embryos expressing SUV4-20H1, which do not display a developmental phenotype. Surprisingly, expression of SUV4-20H1WT showed a strong reduction of H4K20me1 (Fig. 4B) even though it did not lead to an increase in H4K20me3 (Fig. 2B). This suggests that SUV4-20H1WT converts H4K20me1 to H4K20me2. Indeed, the reduction in H4K20me1 upon expression of SUV4-20H1WT was dependent on its methyltransferase activity (Fig. 4B). However, in spite of multiple attempts, we were unable to identify a specific H4K20me2 antibody and therefore were unable to perform immunostaining for H4K20me2 (data not shown). Importantly, even though SUV4-20H1WT embryos show a reduction in H4K20me1 levels, they do not exhibit embryonic lethality or cell cycle arrest (Fig. 2C), in contrast to SUV4-20H2WT embryos, which show both reduction of H4K20me1 and developmental arrest (Fig. 1C). This observation suggests that reduction of H4K20me1 levels may not explain the developmental phenotype observed upon expression of SUV4-20H2WT, leading us to conclude that the developmental arrest in SUV4-20H2WT embryos is mostly independent of changes in H4K20me1 levels. Indeed, while PR-Set7 loss leads to a G2/M arrest (Oda et al. 2009; Tardat et al. 2010), we did not detect changes in H3S10p in SUV4-20H2WT embryos compared with noninjected or SUV4-20H2MUT controls (Fig. 4C; Supplemental Fig. S4B). We aimed to address whether over-

expression of PHF2 would rescue the developmental phenotype elicited by SUV4-20H2WT expression by reducing H4K20me3. However, under the experimental conditions tested, we were unable to detect a change in global H4K20me3 levels in spite of robust PHF2 expression (Supplemental Fig. S3D–F).

SUV4-20H2-mediated embryonic arrest is partially rescued by inhibiting ATR

All of our observations together suggest that the phenotypic arrest of embryos expressing SUV4-20H2WT is a result of a misregulation of DNA replication. H4K20 methylation can be a marker of DNA damage, and increased levels of H4K20me2/3 have been linked to the activation of the ATR pathway in cancer cells (Botuyan et al. 2006; Hajdu et al. 2011; Pei et al. 2011). We therefore hypothesized that sustained H4K20me3 could trigger DNA damage checkpoint activation during S phase in embryos. We thus investigated the levels of γ H2A.X, an indicator of DNA damage and replication stress, in addition to CHK1 and phosphorylated CHK2 (CHK2p), downstream effectors of the ATR/ATM pathway and S-phase checkpoint activation (Mechali et al. 2013). Immunostaining of two-cell stage embryos at late S phase using a γ H2A.X antibody revealed diffuse nuclear accumulation of γ H2A.X with a few foci in control noninjected embryos, in agreement with previous observations (Fig. 4D; Ziegler-Birling et al. 2009). We did not detect changes in the global levels of γ H2A.X in embryos expressing either SUV4-20H2WT or SUV4-20H2MUT (Fig. 4D; Supplemental Fig. S4C), suggesting that sustained H4K20me3 levels do not cause dsDNA damage globally. In contrast, SUV4-20H2WT embryos showed higher levels of CHK1 in comparison with noninjected and SUV4-20H2MUT embryos during S phase (Fig. 4E; Supplemental Fig. S6A–D). In addition, CHK2p was also strongly induced in both pronuclei of SUV4-20H2WT embryos compared with noninjected and SUV4-20H2MUT embryos, which showed barely any signal for CHK2p (Fig. 4F; Supplemental Fig. S6E–H). Note that we were unable to address levels of phosphorylated CHK1, since none of the commercially available CHK1p antibodies worked in immunostaining (data not shown). The increase in CHK1 and CHK2p indicates a potential signaling from DNA damage during S phase upon expression of SUV4-20H2. Since H4K20me3 constitutes a direct platform for origin replication complex (ORC) loading (Beck et al. 2012), it is possible that increased origins of replication may be used or available in SUV4-20H2WT embryos. Indeed, this could explain the increase in EdU levels in these embryos at the beginning of S phase (Fig. 3F) as well as the CHK1 and CHK2p increase through the generation of more ssDNA intermediates. The observation that CHK1 and CHK2p levels were higher in SUV4-20H2WT embryos suggested a checkpoint activation during S phase. We reasoned that if embryos with sustained H4K20me3 levels upon expression of SUV4-20H2 arrest because of an S-phase checkpoint activation, we should be able to release the developmental arrest at least partially through inhibition of the ATR

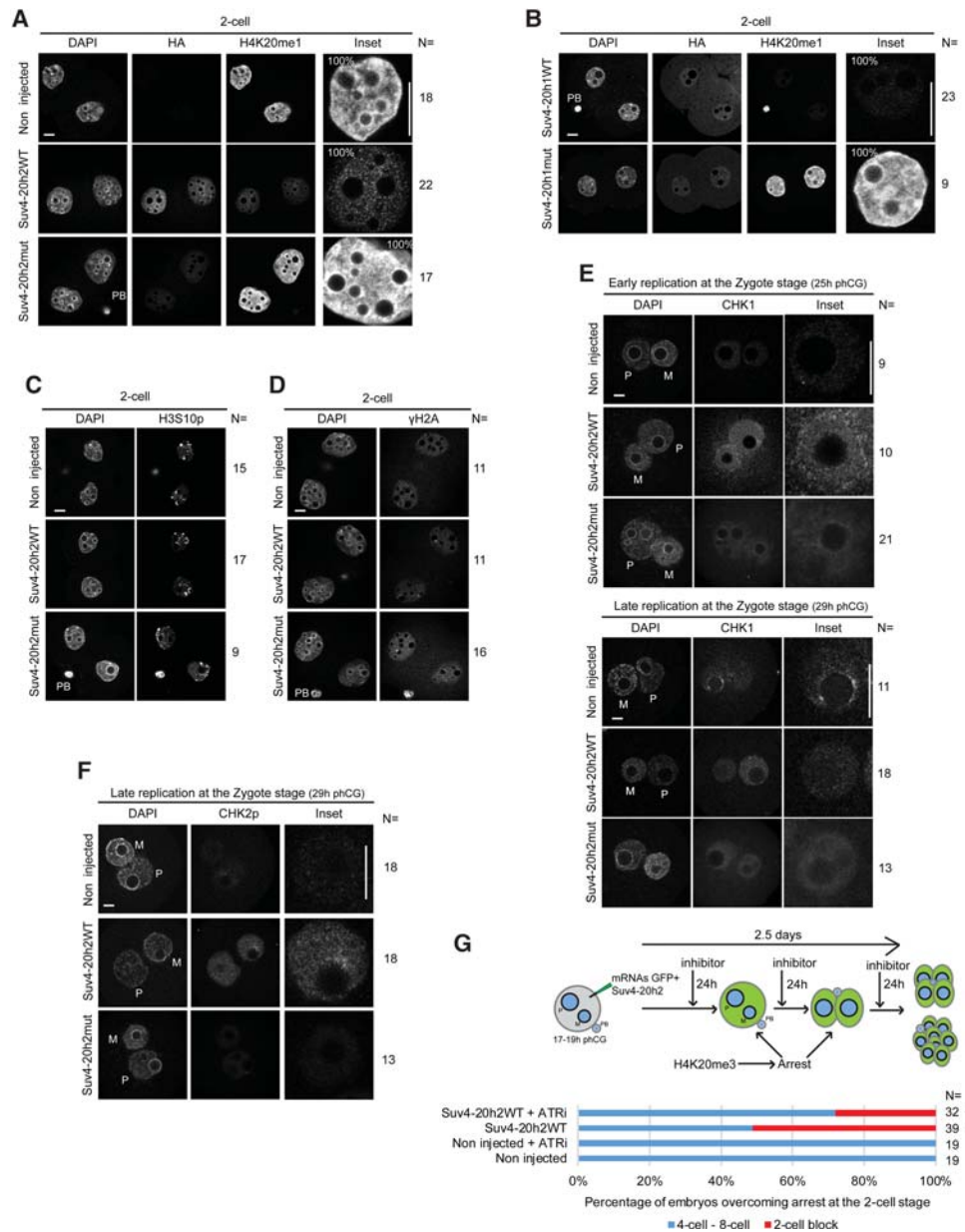


Figure 4. Embryonic arrest is partially rescued by inhibition of ATR. (A) Representative images of two-cell stage embryos analyzed at 46 h phCG with DAPI, HA, and H4K20me1 antibodies. Single Z-projections of confocal sections are shown. An inset of one of the two nuclei is shown at the *right*. N numbers are indicated. Bars: inset, 5 μ m; all others, 10 μ m. (PB) Polar body. (B) Representative single Z-projections of confocal sections of *Suv4-20h1WT*-injected and *Suv4-20h1mut*-injected two-cell stage embryos stained at 46 h phCG with DAPI, HA, and H4K20me1 antibodies. An inset of one of the two nuclei is shown in the *right* panels. N numbers are indicated. Bars: inset, 5 μ m; all others, 10 μ m. (PB) Polar body. (C) Representative two-cell stage embryos acquired at 46 h phCG stained with DAPI and H3S10p antibodies, showing confocal single Z-projections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected embryos. N numbers are indicated. Bar, 10 μ m. (PB) Polar body. (D) Representative confocal single Z-projections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected two-cell stage embryos stained at 46 h phCG with DAPI and γ H2A.X antibodies. N numbers are indicated. Bar, 10 μ m. (PB) Polar body. (E) Representative zygotes stained at 25 h phCG (early S phase) and 29 h phCG (late S phase) with DAPI and CHK1 antibodies, showing single confocal Z-projections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected embryos. Bars: inset, 5 μ m; all others, 10 μ m. (M) Maternal; (P) paternal. (F) Confocal single Z-projections of noninjected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected zygotes acquired at 29 h phCG with DAPI and CHK2p antibodies. An inset of the maternal pronucleus is shown in the *right* panels. N numbers are indicated. Bars: inset, 5 μ m; all others, 10 μ m. (M) Maternal; (P) paternal. For E and F, *SUV4-20H2MUT* and *SUV4-20H2WT* embryos were processed separately but in parallel with noninjected controls. (G) A schematic representation of the experimental design is shown at the *top*. Zygotes between 17 and 19 h phCG were microinjected with mRNA for *Suv4-20h2WT* and *GFP* and cultured in the presence or absence of an ATR inhibitor (ATRi) until the eight-cell stage. A bar chart of the developmental progression until the four-cell stage is shown at the *bottom*. Statistical testing was performed using the $N-1$ two-proportion test for comparing independent proportions. (*) $P < 0.05$.

pathway. To address this, embryos were injected as above with mRNA for *Suv4-20h2WT* and cultured in the presence of an ATR inhibitor (ATRi) from the late zygote stage (Fig. 4G). Since longer inhibition of ATR is known to block developmental progression (Brown and Baltimore 2000; Nakatani et al. 2015), we focused specifically on assessing the developmental block beyond the two-cell stage, which accounts for the majority of the phenotype (77% of SUV4-20H2 embryos arrest prior to the two-cell stage), by scoring embryos that reached the four-to-eight-cell stage transition. As shown in Figure 4G, all non-injected embryos cultured in the presence of the ATRi reached the four-cell stage at a rate similar to that of noninjected embryos cultured without inhibitor. SUV4-20H2WT embryos treated with ATRi developed at significantly higher rates than SUV4-20H2WT embryos cultured without the inhibitor ($P < 0.05$; $n = 32$) (Fig. 4G). Although not all embryos overcame the two-cell stage block upon inhibition of ATR, our results suggest that the developmental defects elicited from sustaining H4K20me3 levels are partially alleviated by inhibiting S-phase checkpoint activation. We conclude that the primary function of H4K20me3 remodeling after fertilization is to allow the timely progression of DNA replication through S phase. This is in contrast to the replication program in somatic cells, where replication occurs normally in the presence of H4K20me3 through ORC binding and anticipates a different regulation of replication during this developmental time window.

Discussion

Embryonic development requires a unique reprogramming mechanism to revert to a ground epigenetic state for a new developmental program to initiate. However, the repercussions of such remodeling at the molecular level are not understood. Zygotes and two-cell stage embryos exhibit a particular nuclear structure with distinctive and asymmetric chromatin signatures thought to be necessary for epigenetic reprogramming. Heterochromatic marks are unique identifiers of parental chromatin at fertilization, which are rapidly remodeled thereafter. To address the requirement for such a chromatin environment in vivo, we aimed to sustain the levels of one such PTM, H4K20me3, by expressing the corresponding methyltransferases (*Suv4-20h1* and *Suv4-20h2*). Indeed, expression of such chromatin modifiers can be used to interrogate the system to shed light on the regulatory mechanisms of the early embryo by studying how they respond to such perturbations. We found that expression of *Suv4-20h2* modifies H4K20 methylation by reducing H4K20me1 and increasing H4K20me3 levels and leads to embryonic arrest mostly before the two-cell stage. This embryonic arrest is dependent on the histone methyltransferase activity of SUV4-20H2. In addition, the developmental failure seems to derive from the specific activity of SUV4-20H2 and the sustained H4K20me3 levels, since expression of SUV4-20H1 does not result in developmental arrest or affect H4K20me3 levels markedly. While it is likely that

the developmental arrest observed is due mainly to the H4K20me3 increase, we cannot rule out the possibility that changes in H3K64me3 levels could also potentially contribute to the developmental phenotype observed upon *Suv4-20h2WT* expression.

It is interesting to note that, in spite of ectopic SUV4-20H2 being detected in both pronuclei, H4K20me3 is only readily detected in the maternal pronucleus prior to the two-cell stage. It is possible that SUV4-20H2 is more active on chromatin already methylated with H3K9me3, which is in line with the reinforcement loop of constitutive heterochromatin in somatic cells (Rea et al. 2000). In addition, because H4K20 methylation is processive and is among the modifications with the slowest rate of formation (Zee et al. 2010), it may take longer for SUV4-20H1 to reach full processivity in the absence of basal levels of H4K20me3 and/or H3K9me3. It is noteworthy that *Suv4-20* enzymes can be targeted to chromatin through interaction with HP1, which mediates stable SUV4-20H2 binding synergistically (Schotta et al. 2004; Hahn et al. 2013). Because HP1 β appears in the paternal chromatin only after S phase (Santos et al. 2005; Santenard et al. 2010) and because SUV4-20H1/H2 are known to function mostly in G1 (Zee et al. 2010), the lack of HP1 on the paternal pronucleus together with the processivity nature of SUV4-20 enzymes would explain why H4K20me3 is detected in the paternal chromatin only at the two-cell stage. SUV4-20H1 seems to display lesser activity toward H4K20me3 than SUV4-20H2 in the embryo. Although their SET domains are highly identical, SUV4-20H2 seems to have higher histone methyltransferase activity than SUV4-20H1 in vitro (Schotta et al. 2004). It is also possible that the longer nature of SUV4-20H1, which has 406 more amino acids than SUV4-20H2, may modulate its processivity. Strong overexpression of *Suv4-20h2* in embryonic stem cells leads to increased chromatin compaction around chromocenters and consequent mitotic segregation defects (Hahn et al. 2013). However, we did not detect segregation defects in SUV4-20H2-expressing embryos.

The expression of SUV4-20H2 also leads to a misregulation of S phase with increased replication sites, concomitant with activation of the ATR pathway (Fig. 5). We postulate that the subsequent activation of the intra-S-phase checkpoint is the cause of the cell cycle block. It is known that activation of ATR leads to a block of replication forks that exhibit ssDNA, but, in order to compensate for the delayed stalled forks undergoing repair, several dormant origins initiate replication (Gilbert 2007), which could explain the continuous EdU incorporation observed in the late S phase in zygotes and two-cell stage embryos expressing SUV4-20H2WT. Alternatively, increased loading of ORC, favored by increased accessibility of its binder target, may facilitate the activation of additional origins, resulting in more extensive ssDNA and subsequent checkpoint activation (Fig. 5). In line with the suggestion of intra-S-phase checkpoint activation, chemical inhibition of ATR partially rescues the S-phase block and developmental capacity. The persistent embryonic arrest in a proportion of embryos could result from misregulation of gene expression independently of the cell cycle/S-phase

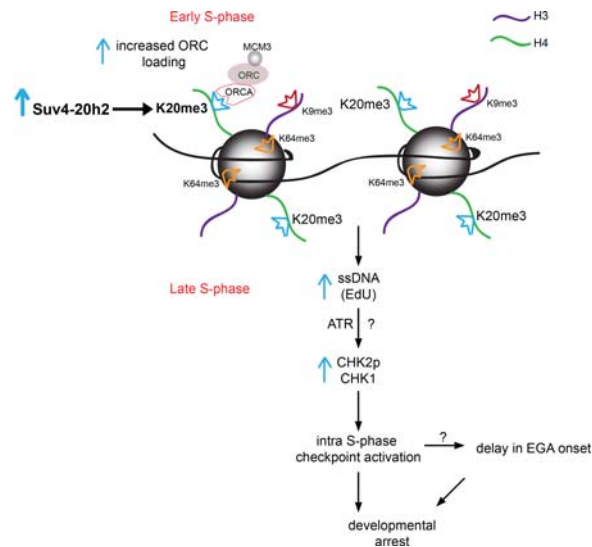


Figure 5. Working model for Suv4-20h2 and H4K20me3 effects on cell cycle progression in embryos. Expression of Suv4-20h2 establishes higher levels of H4K20me3, which is a known target of ORCA. In turn, ORCA can recruit ORC during early S phase to replication origins. The increased levels of H4K20me3 lead to increased ORC loading, which would result in higher levels of ssDNA, as measured by EdU labeling. The increase in ssDNA can activate the ATR pathway, which promotes CHK2 phosphorylation, leading to the activation of the intra-S-phase checkpoint, causing a delay in cell cycle progression and developmental arrest.

progression at the two-cell stage, which might not be overcome by ATR inhibition. However, we could not formally distinguish effects on gene expression caused by developmental arrest.

Our results underscore the necessity for a chromatin signature in zygotes and two-cell embryos that is devoid of H4K20me3 and heterochromatin domains for the correct progression of replication. Indeed, it is possible that the relatively open chromatin structure at these stages is sufficient to enable regulated access to the replication machinery (Boskovic et al. 2014; Wu et al. 2016). More broadly, this anticipates a functional difference in the organization of the replication program between the early embryos and somatic cells.

Materials and methods

Embryo collection, microinjection, and culture

Embryos were collected from 5- to 7-wk-old F1 (C57BL/6J \times CBA/H) superovulated females crossed with F1 males. Superovulation was induced by intraperitoneal injection of 5 IU of pregnant mare serum gonadotropin (PMSG, Intervet) and 7.5 IU of hCG (Intervet) 46–48 h later. Zygotes were collected between 17 and 19 h phCG injection. mRNAs were transcribed in vitro from the pRN3P plasmid using the mMESSAGE mMACHINE kit (Ambion). All cDNAs were subcloned to include identical 5' caps and untranslated regions (UTRs; including Kozak) and a poly-T 3' UTR tail to ensure equivalent expression levels after mi-

croinjection. *Suv4-20h1* and *Suv4-20h2* cDNA were obtained through a generous gift from G. Schotta (Ludwig-Maximilians-Universität München, Munich), and *Suv4-20h1mut* was obtained through a generous gift from D. Beck (New York University, New York). *Suv4-20h2mut* was prepared by site-directed mutagenesis of *Suv4-20h2* at asparagine 273 to cysteine 276 (NHDC) into AAAG (wild-type sequence CAACCATGACTG to mutated sequence CGCCGCTGCCGG) (Rea et al. 2000). Mouse *Phf2* cDNA was obtained from H. Qi (University of Iowa) and subcloned into the pRN3P plasmid. Zygotes were microinjected with 1–2 pL of the indicated mRNAs (700 ng/ μ L for *Suv4-20h1/Suv4-20h1mut/Suv4-20h2/Suv4-20h2mut/Phf2* or 250 ng/ μ L for *Gfp*) and allocated to the experimental groups at random. Embryos were cultured in KSOM (K-modified simplex optimized medium) microdrops under oil at 37°C and 5% CO₂ until they were fixed. Microinjections at the two-cell stage were performed in one of the blastomeres at random after embryo collection at 41–43 h phCG. Rescue experiments of the developmental block were performed by adding KSOM containing 10 μ M ATRi (Millipore, catalog no. 5.04972.0001) after injection and renewing the medium daily for 2 d. Embryos were monitored and scored daily to determine developmental progression. Data were derived from at least three independent biological replicates. For statistical analysis of embryonic development, the $N-1$ two-proportion test for comparing independent proportions for small and large sample sizes was used, which is based on the $N-1$ χ^2 test originally proposed by Pearson and recommended by Campbell. All experiments were performed after approval of the Ethics Committee of the Université de Strasbourg and according to French and European legislation on animal experimentation.

EU and EdU incorporation

Embryos were incubated with 50 μ M EU for 1 h at 48 h phCG treatment and then visualized by Click-iT chemistry (Life Technologies) as described in the manufacturer's instructions. Quantifications were performed as described in "Quantification of Fluorescence Intensity" (below). Embryos were incubated with 50 μ M EdU for 1 h at the times indicated in the figure legends and figure schemes and then visualized by Click-iT chemistry (Life Technologies) as described in the manufacturer's instructions.

Immunostaining and confocal microscopy

Fixation of freshly collected embryos from F1 superovulated, microinjected, and cultured embryos was performed as described (Torres-Padilla et al. 2006). Briefly, the zona pellucida was removed with Acid Tyrode solution followed by two washes in PBS and fixation in 4% paraformaldehyde, 0.04% Triton, 0.3% Tween20, and 0.2% sucrose at 37°C to ensure preservation of nuclear architecture. After permeabilization with 0.5% Triton in PBS, embryos were washed three times in PBSt (0.1% Tween20 in PBS), quenched in 2.6 mg/mL freshly prepared ammonium chloride, washed twice in PBSt, blocked for 3–4 h or overnight at 4°C in blocking solution (BS: 3% BSA in PBSt), and incubated with primary antibodies in BS. The antibodies used were as follows: anti-HA (Abcam, 16B12), anti-H3K9me3 (Millipore, 07-442), anti-H4K20me3 (Millipore, 07-463), anti-H3K9me3 (Millipore, 07-442), anti-H3K64me3 (generous gift from S. Daujat Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch), anti-H4K20me1 (Abcam, ab9051), anti-H3S10p (Abcam, ab5176), anti- γ H2A.X (Millipore, 05-636), anti-CHK1 (Cell Signaling, 2G1D5), and anti-CHK2p (Cell Signaling, T68 2661S). After overnight incubation at 4°C, embryos were washed three

Eid et al.

times in PBSt, blocked for 20 min in BS, and incubated for 3 h at room temperature in BS containing secondary antibodies labeled with Alexa fluorophores (Invitrogen). After washing twice in PBSt and once in PBS, embryos were mounted in VectaShield (Vector Laboratories) containing DAPI to visualize DNA. Confocal microscopy was performed on a 63 \times oil objective in a TCS SP5 inverted confocal microscope (Leica). Z-sections were taken every 0.5–1 μ m. Image analysis was performed using the LAS-AF (Leica) and Imaris (Bitplane) software.

Quantification of fluorescence intensity

Confocal z-series stacks were reconstructed in three dimensions using Imaris software (Bitplane), and the pronuclei (zygote) or nuclei (two-cell stage) were segmented based on the DAPI channel. The average fluorescence intensity for the channel of interest within the segmented region was calculated after uniform background subtraction with a 95% confidence. The fluorescence intensity for each embryo was normalized to the average of the noninjected control group. The data were tested for normality using the Kolmogorov-Smirnov test. If the data were found to not be normally distributed, the Mann-Whitney *U*-test, a nonparametric test, was used.

Gene expression analysis

Control noninjected, *Suv4-20h1WT*-injected, *Suv4-20h2WT*-injected, and *Suv4-20h2mut*-injected two-cell stage embryos were washed in PBS and flash-frozen in liquid nitrogen in 5 μ L of 2 \times reaction buffer (CellsDirect One-Step qRT-PCR kit, Invitrogen) at 44 h phCG. TaqMan gene expression assays (20 \times ; Applied Biosystems), previously tested using embryonic stem cell cDNA for amplification efficiency, were pooled to a final concentration of 0.2 \times for each of the 48 assays. To each of the single-cell samples in 2 \times reaction buffer, 2.5 μ L of 0.2 \times assay pool, 0.5 μ L of RT/Taq enzyme (CellsDirect qRT-PCR kit, Invitrogen), and 2.3 μ L of water were added. Cell lysis and sequence-specific reverse transcription were performed for 20 min at 50°C. The reverse transcriptase was inactivated by heating for 2 min to 95°C. Sequence-specific preamplification was performed by denaturing for 15 sec at 95°C and then annealing and amplification for 4 min at 60°C for 18 cycles. The resulting cDNA was diluted fivefold before analysis with Universal PCR master mix and TaqMan gene expression assays (Applied Biosystems) in 48:48 Dynamic Arrays on a Biomark system (Fluidigm). C_t values were calculated from the system's software (Biomark real-time PCR analysis, Fluidigm). All raw C_t values were normalized to the assumed detection C_t level of 28 following the recommendation from Fluidigm technical support as in Guo et al. (2010) and Burton et al. (2013). C_t values >28 and those with curve qualities <0.65 were deemed unreliable measurements and were substituted with values of 28. Whenever C_t values or quality scores were judged unreliable in one replicate but not the other, those of the successful replicate were kept. Additionally, all samples lacking expression of reference genes *Actin-b* and *Gapdh* were removed from further analysis. The remaining C_t values were subtracted from 28 in order to achieve a scale in which zero corresponded to the lack of expression, and an increase of 1 U indicated a doubling of the expression level (Guo et al. 2010; Burton et al. 2013). Violin plots of the resulting data set were generated using the ggplot2 R package. Statistical analysis was performed using the Mann-Whitney *U*-test. The PCA was performed using the pcaMethods R package, and variables were scaled to unit variance. All plotting was done in R using the ggplot2 package. The dendrogram clustering was generated in R using Euclidean distance.

Acknowledgments

We are grateful to Gunnar Schotta (Ludwig-Maximilians-Universität München) for providing antibodies, wild-type *Suv4-20h1* and *Suv4-20h2* constructs, and helpful discussions; David Beck (New York University) for providing the *Suv4-20h1mut* construct; and Hank H. Qi for the *Phf2* cDNA. M.-E.T.-P. acknowledges funding from EpiGeneSys NoE, ERC-Stg "Nuclear-Potency," the EMBO Young Investigator Programme, the Schlumberger Foundation for Research and Education, and the Helmholtz Gesellschaft. A.E. was a recipient of a doctoral fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche and the Fondation pour la Recherche Médicale (FDT20150532012).

References

- Aoki F, Worrall DM, Schultz RM. 1997. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* **181**: 296–307.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837.
- Beck DB, Burton A, Oda H, Ziegler-Birling C, Torres-Padilla ME, Reinberg D. 2012. The role of PR-Set7 in replication licensing depends on Suv4-20h. *Genes Dev* **26**: 2580–2589.
- Bensaude O, Babinet C, Morange M, Jacob F. 1983. Heat shock proteins, first major products of zygotic gene activity in mouse embryo. *Nature* **305**: 331–333.
- Bolton VN, Oades PJ, Johnson MH. 1984. The relationship between cleavage, DNA replication, and gene expression in the mouse 2-cell embryo. *J Embryol Exp Morphol* **79**: 139–163.
- Boskovic A, Eid A, Pontabry J, Ishiuchi T, Spiegelhalter C, Raghu Ram EV, Meshorer E, Torres-Padilla ME. 2014. Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev* **28**: 1042–1047.
- Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G. 2006. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**: 1361–1373.
- Brown EJ, Baltimore D. 2000. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev* **14**: 397–402.
- Burton A, Torres-Padilla ME. 2014. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol* **15**: 723–734.
- Burton A, Muller J, Tu S, Padilla-Longoria P, Guccione E, Torres-Padilla ME. 2013. Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. *Cell Rep* **5**: 687–701.
- Congdon LM, Houston SI, Veerappan CS, Spektor TM, Rice JC. 2010. PR-Set7-mediated monomethylation of histone H4 lysine 20 at specific genomic regions induces transcriptional repression. *J Cell Biochem* **110**: 609–619.
- Daujat S, Weiss T, Mohn F, Lange UC, Ziegler-Birling C, Zeissler U, Lappe M, Schubeler D, Torres-Padilla ME, Schneider R. 2009. H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. *Nat Struct Mol Biol* **16**: 777–781.
- Flach G, Johnson MH, Braude PR, Taylor RA, Bolton VN. 1982. The transition from maternal to embryonic control in the 2-cell mouse embryo. *EMBO J* **1**: 681–686.
- Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, et al.

2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* **37**: 391–400.
- Gilbert DM. 2007. Replication origin plasticity, Taylor-made: inhibition vs. recruitment of origins under conditions of replication stress. *Chromosoma* **116**: 341–347.
- Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P. 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18**: 675–685.
- Hahn M, Dambacher S, Dulev S, Kuznetsova AY, Eck S, Worz S, Sadic D, Schulte M, Mallm JP, Maiser A, et al. 2013. Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. *Genes Dev* **27**: 859–872.
- Hajdu I, Ciccio A, Lewis SM, Elledge SJ. 2011. Wolf-Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proc Natl Acad Sci* **108**: 13130–13134.
- Houston SI, McManus KJ, Adams MM, Sims JK, Carpenter PB, Hendzel MJ, Rice JC. 2008. Catalytic function of the PR-Set7 histone H4 lysine 20 monomethyltransferase is essential for mitotic entry and genomic stability. *J Biol Chem* **283**: 19478–19488.
- Jachowicz JW, Santenard A, Bender A, Muller J, Torres-Padilla ME. 2013. Heterochromatin establishment at pericentromeres depends on nuclear position. *Genes Dev* **27**: 2427–2432.
- Jorgensen S, Elvers I, Trelle MB, Menzel T, Eskildsen M, Jensen ON, Helleday T, Helin K, Sorensen CS. 2007. The histone methyltransferase SET8 is required for S-phase progression. *J Cell Biol* **179**: 1337–1345.
- Kapoor-Vazirani P, Kagey JD, Vertino PM. 2011. SUV420H2-mediated H4K20 trimethylation enforces RNA polymerase II promoter-proximal pausing by blocking hMOF-dependent H4K16 acetylation. *Mol Cell Biol* **31**: 1594–1609.
- Kourmouli N, Jeppesen P, Mahadevhaiah S, Burgoyne P, Wu R, Gilbert DM, Bongiorno S, Prantera G, Fanti L, Pimpinelli S, et al. 2004. Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. *J Cell Sci* **117**: 2491–2501.
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116–120.
- Lange UC, Siebert S, Wossidlo M, Weiss T, Ziegler-Birling C, Walter J, Torres-Padilla ME, Daujat S, Schneider R. 2013. Dissecting the role of H3K64me3 in mouse pericentromeric heterochromatin. *Nat Commun* **4**: 2233.
- Mechali M, Yoshida K, Coulombe P, Pasero P. 2013. Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr Opin Genet Dev* **23**: 124–131.
- Nakatani T, Yamagata K, Kimura T, Oda M, Nakashima H, Hori M, Sekita Y, Arakawa T, Nakamura T, Nakano T. 2015. Stella preserves maternal chromosome integrity by inhibiting 5hmC-induced γ H2AX accumulation. *EMBO Rep* **16**: 582–589.
- Nishioka K, Rice JC, Sarma K, Erdjument-Bromage H, Werner J, Wang Y, Chuikov S, Valenzuela P, Tempst P, Steward R, et al. 2002. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol Cell* **9**: 1201–1213.
- Oda H, Okamoto I, Murphy N, Chu J, Price SM, Shen MM, Torres-Padilla ME, Heard E, Reinberg D. 2009. Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Mol Cell Biol* **29**: 2278–2295.
- Oda H, Hubner MR, Beck DB, Vermeulen M, Hurwitz J, Spector DL, Reinberg D. 2010. Regulation of the histone H4 monomethylase PR-Set7 by CRL4(Cdt2)-mediated PCNA-dependent degradation during DNA damage. *Mol Cell* **40**: 364–376.
- Pei H, Zhang L, Luo K, Qin Y, Chesni M, Fei F, Bergsagel PL, Wang L, You Z, Lou Z. 2011. MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature* **470**: 124–128.
- Probst AV, Santos F, Reik W, Almouzni G, Dean W. 2007. Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. *Chromosoma* **116**: 403–415.
- Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593–599.
- Rice JC, Nishioka K, Sarma K, Steward R, Reinberg D, Allis CD. 2002. Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev* **16**: 2225–2230.
- Santenard A, Ziegler-Birling C, Koch M, Tora L, Bannister AJ, Torres-Padilla ME. 2010. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* **12**: 853–862.
- Santos F, Peters AH, Otte AP, Reik W, Dean W. 2005. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol* **280**: 225–236.
- Schneider AC, Heukamp LC, Rogenhofer S, Fechner G, Bastian PJ, von Ruecker A, Muller SC, Ellinger J. 2011. Global histone H4K20 trimethylation predicts cancer-specific survival in patients with muscle-invasive bladder cancer. *BJU Int* **108**: E290–E296.
- Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**: 1251–1262.
- Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callen E, Celeste A, Pagani M, Opravil S, De La Rosa-Velazquez IA, et al. 2008. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev* **22**: 2048–2061.
- Sims JK, Houston SI, Magazinnik T, Rice JC. 2006. A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. *J Biol Chem* **281**: 12760–12766.
- Southall SM, Cronin NB, Wilson JR. 2014. A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. *Nucleic Acids Res* **42**: 661–671.
- Stender JD, Pascual G, Liu W, Kaikkonen MU, Do K, Spann NJ, Boutros M, Perrimon N, Rosenfeld MG, Glass CK. 2012. Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. *Mol Cell* **48**: 28–38.
- Tardat M, Murr R, Herceg Z, Sardet C, Julien E. 2007. PR-Set7-dependent lysine methylation ensures genome replication and stability through S phase. *J Cell Biol* **179**: 1413–1426.
- Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, Sardet C, Julien E. 2010. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat Cell Biol* **12**: 1086–1093.
- Torres-Padilla ME, Bannister AJ, Hurd PJ, Kouzarides T, Zernicka-Goetz M. 2006. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol* **50**: 455–461.
- Van Den Broeck A, Brambilla E, Moro-Sibilot D, Lantuejoul S, Brambilla C, Eymin B, Khochbin S, Gazerri S. 2008. Loss of

Eid et al.

- histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. *Clin Cancer Res* **14**: 7237–7245.
- Vermeulen M, Eberl HC, Matarese F, Marks H, Denisov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, et al. 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**: 967–980.
- Wen H, Li J, Song T, Lu M, Kan PY, Lee MG, Sha B, Shi X. 2010. Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J Biol Chem* **285**: 9322–9326.
- Wongtawan T, Taylor JE, Lawson KA, Wilmot I, Pennings S. 2011. Histone H4K20me3 and HP1 α are late heterochromatin markers in development, but present in undifferentiated embryonic stem cells. *J Cell Sci* **124**: 1878–1890.
- Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, Zhang B, Liu B, Wang Q, Xia W, et al. 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**: 652–657.
- Yokoyama Y, Matsumoto A, Hieda M, Shinchi Y, Ogihara E, Hamada M, Nishioka Y, Kimura H, Yoshidome K, Tsujimoto M, et al. 2014. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast Cancer Res* **16**: R66.
- Zee BM, Levin RS, Xu B, LeRoy G, Wingreen NS, Garcia BA. 2010. In vivo residue-specific histone methylation dynamics. *J Biol Chem* **285**: 3341–3350.
- Ziegler-Birling C, Helmrich A, Tora L, Torres-Padilla ME. 2009. Distribution of p53 binding protein 1 (53BP1) and phosphorylated H2A.X during mouse preimplantation development in the absence of DNA damage. *Int J Dev Biol* **53**: 1003–1011.



SUV4-20 activity in the preimplantation mouse embryo controls timely replication

André Eid, Diego Rodriguez-Terrones, Adam Burton, et al.

Genes Dev. published online December 5, 2016

Access the most recent version at doi:[10.1101/gad.288969.116](https://doi.org/10.1101/gad.288969.116)

Supplemental Material

<http://genesdev.cshlp.org/content/suppl/2016/12/05/gad.288969.116.DC1>

Published online December 5, 2016 in advance of the full issue.

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner for EXIQON, a QIAGEN company. The left side features a blue background with a microscopic view of cells and the text "Boost NGS microRNA profiling. Read about 3 methods tested". The right side has a white background with the EXIQON logo (a grid of blue dots) and the text "EXIQON Now a QIAGEN company". The QIAGEN logo (a grid of blue dots) is also present in the bottom right corner of the banner.

Supplemental Figure Legends

Figure S1. Expression levels of H4K20me3 dimethyl- and trimethyl-transferases during pre-implantation development.

Relative expression level of *Suv4-20h1* (A), *Suv4-20h2* (A), *Smyd-5* (B) and *Phf-2* (C) at all stages of pre-implantation development normalized to *Actinb*. Shown are means \pm s.d. of two independent experiments performed in technical replicates. Blue and red dots correspond to individual biological replicates.

(D) Relative expression levels of SUV4-20H2WT in zygotes and 2-cell stage embryos upon mRNA injection as in Figure 1A. Quantification was done using HA staining, and was normalized to DAPI. Shown are means \pm s.e.m. of 13 zygotes and 21 2-cell stage embryos.

Figure S2. Expression of SUV4-20H2 at the 2-cell stage results in cell proliferation arrest.

A. Experimental design: a random blastomere of a 2-cell stage embryo (37h phCG) was micro-injected with mRNA for *Suv4-20h2WT* and *GFP*.

B. Representative confocal single Z-projections of Non injected, *Suv4-20h2WT*- and *GFP*-injected 2-cell stage embryos stained at 48h phCG with DAPI, GFP; HA and H4K20me3 antibodies. An inset of one H4K20me3 of the two nuclei is shown in the right panels. Scale bar corresponds to 10 μ m, except in the inset where it corresponds to 5 μ m. PB, Polar Body.

C. Bar chart distribution of the number of GFP negative and positive blastomeres at the blastocyst stage in *GFP*- or *Suv4-20h2WT*- injected 2-cell stage embryos. The number of blastocysts analyzed is indicated at the bottom of the chart.

D-E. Representative full Z-projections of confocal sections of blastocysts injected with *Suv4-20h2WT*- and *GFP* acquired with DAPI, GFP, and phalloidin. A merged channel of the three signals is shown at the bottom. Scale bar corresponds to 10 μ m.

Figure S3. Suv4-20h2 ectopic expression affects the levels of H3K64me3.

Quantification of H3K9me3 (**A**), H3K64me3 (**B**) in late 2-cell-stage embryos at 46h phCG. The fluorescence intensity per nucleus was normalized to the mean intensity of non injected embryos following background subtraction (detailed in the material and methods) and the number of nuclei analyzed is shown below the graphs. The mean averages of these values are shown (\pm SEM). The Mann-Whitney U test was applied and only significant differences are indicated.

C. Normalized EdU levels measured in each pronucleus of non injected and *GFP*-injected zygotes at late S-phase.

D-F. PHF2 does not lead to changes in global H4K20me3 levels upon Suv4-20h2 expression. (**D**) Experimental design: zygotes were microinjected with mRNA for *Suv4-20h2* as in Figure 1 except that RFP mRNA was used as positive control for injection. Embryos were left to develop until the mid 2-cell stage at which time one of the 2 blastomeres was injected with mRNA for HA-PHF2 together with membraneGFP in order to identify the injected blastomere. (**E**) Embryos were then fixed 24h later and analyzed with an HA antibody to reveal PHF2 and with an H4K20me3 antibody as indicated. Note that because of the delay induced by SUV4-20H2, the embryo shown is still at the 2-cell stage. In (**F**), H4K20me3 levels were quantified and normalized against the negative blastomere for HA-PHF2. Note that under the experimental conditions used, PHF2 does not appear to regulate H4K20me3 globally.

Figure S4. Suv4-20h2 ectopic expression reduces the levels of H4K20me1 but does not affect H3S10p or γ H2A.X.

Quantification of H4K20me1 (**A**), H3S10p (**B**) and γ H2A.X (**C**) in late 2-cell-stage embryos at 46h phCG. The fluorescence intensity per nucleus was normalized to the mean intensity of non injected embryos following background subtraction (detailed in the material and methods) and the number of nuclei analyzed is shown below the graphs. The mean averages of these values are shown (\pm SEM). The Mann-Whitney U test was applied and only significant differences are indicated.

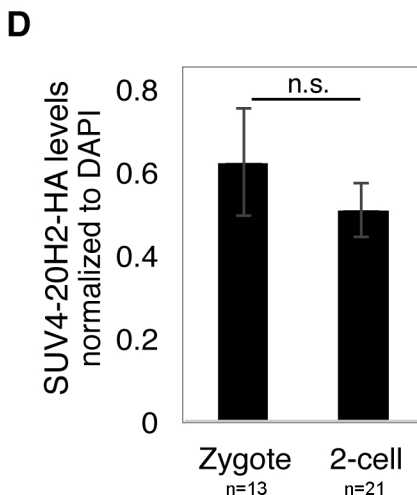
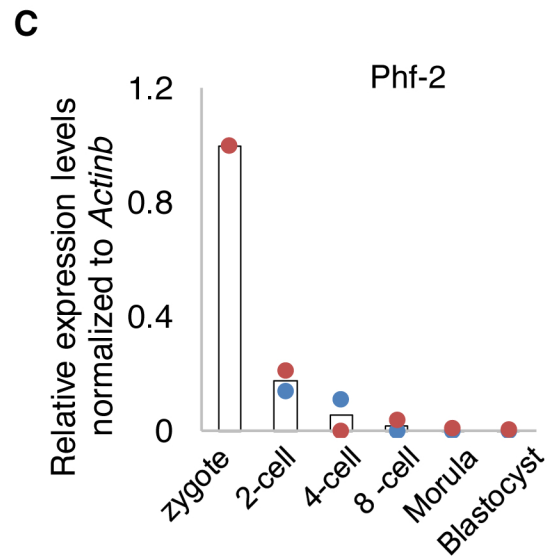
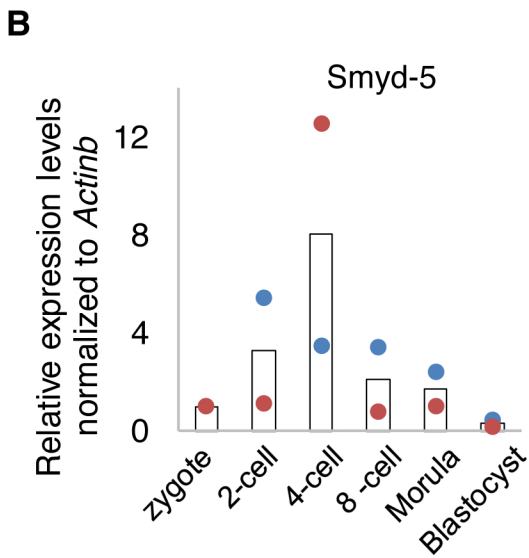
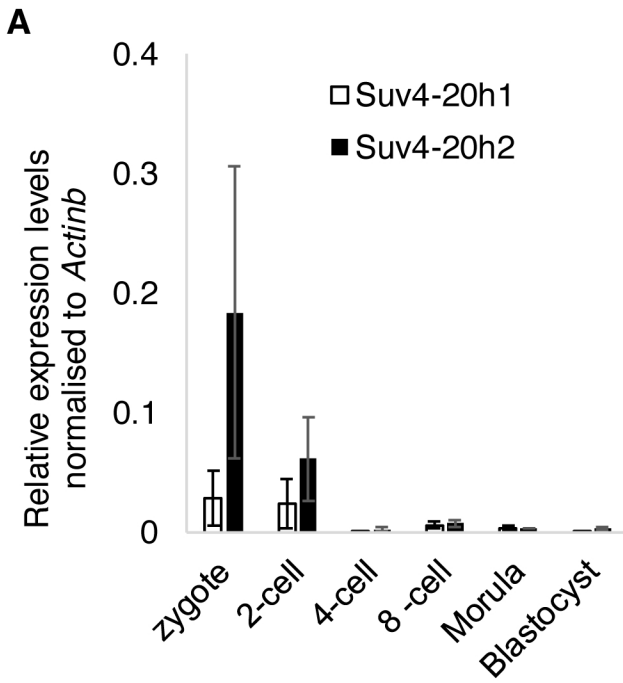
Figure S5. Gene expression analysis does not show a different expression profile in Suv4-20h2WT- expressing embryos.

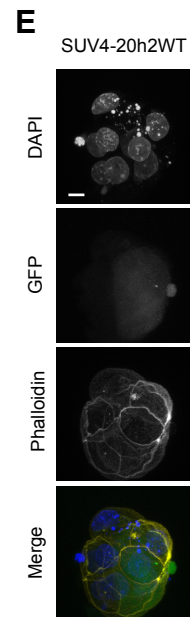
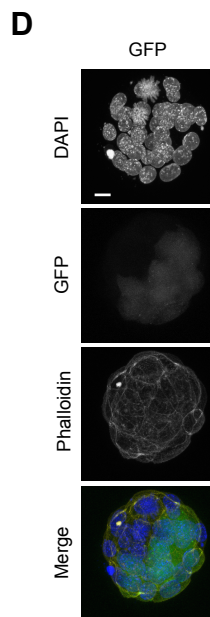
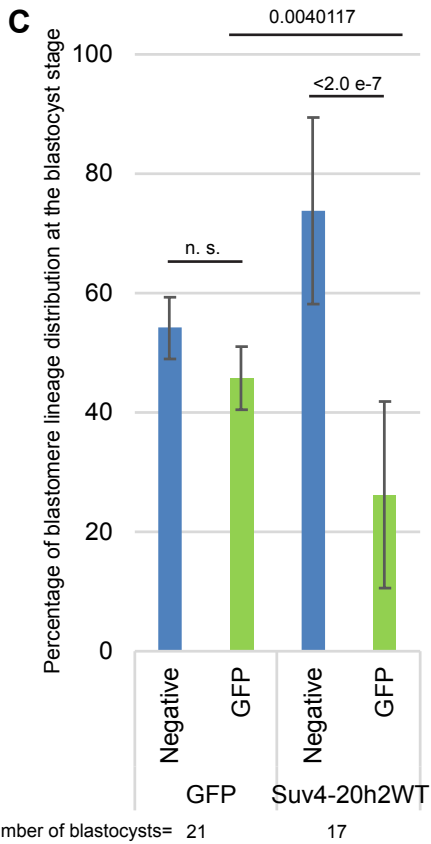
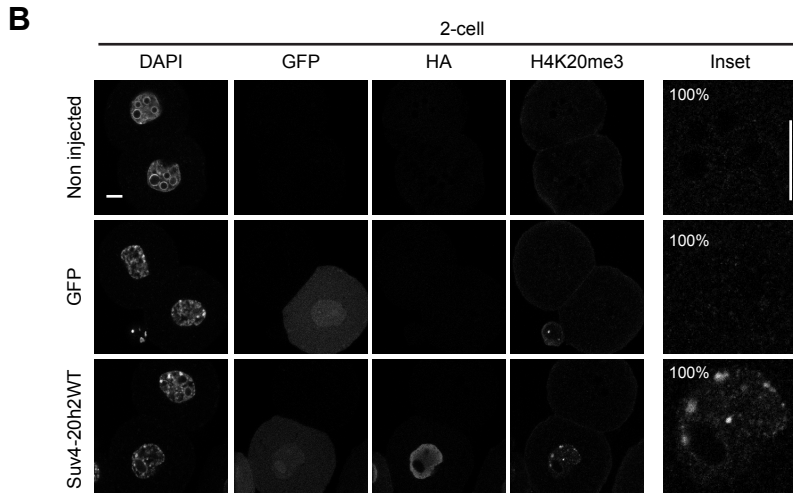
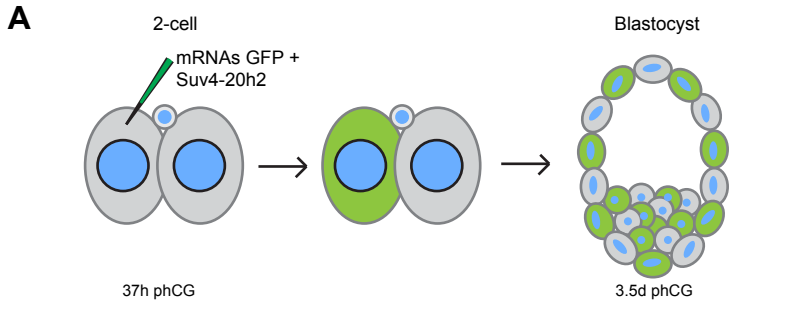
A. Cluster dendrogram of single 2-cell stage embryos at 44h phCG (during mid-EGA) based on 45 genes' expression levels. Embryos are identified by their experimental group (Non injected, Suv4-20h1, Suv4-20h2 and Suv4-20h2AAAG) followed by an identification number.

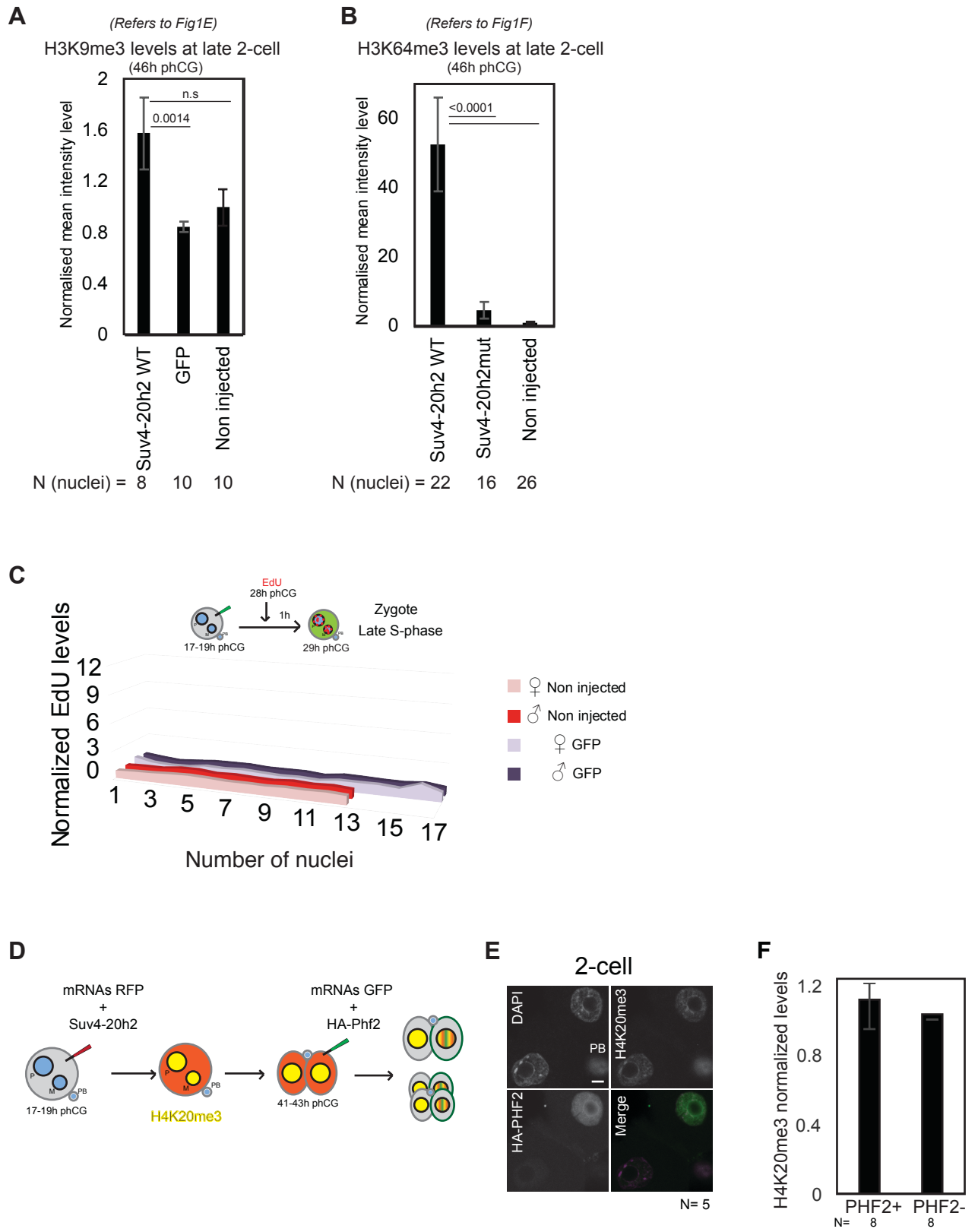
B. Violin plots of 43 genes analyzed in single 2-cell stage embryos at 44h phCG and averaged to *Actinb* and *Gapdh*. Average expression levels is indicated with a black dot.

Figure S6. Activation of the S-phase checkpoint.

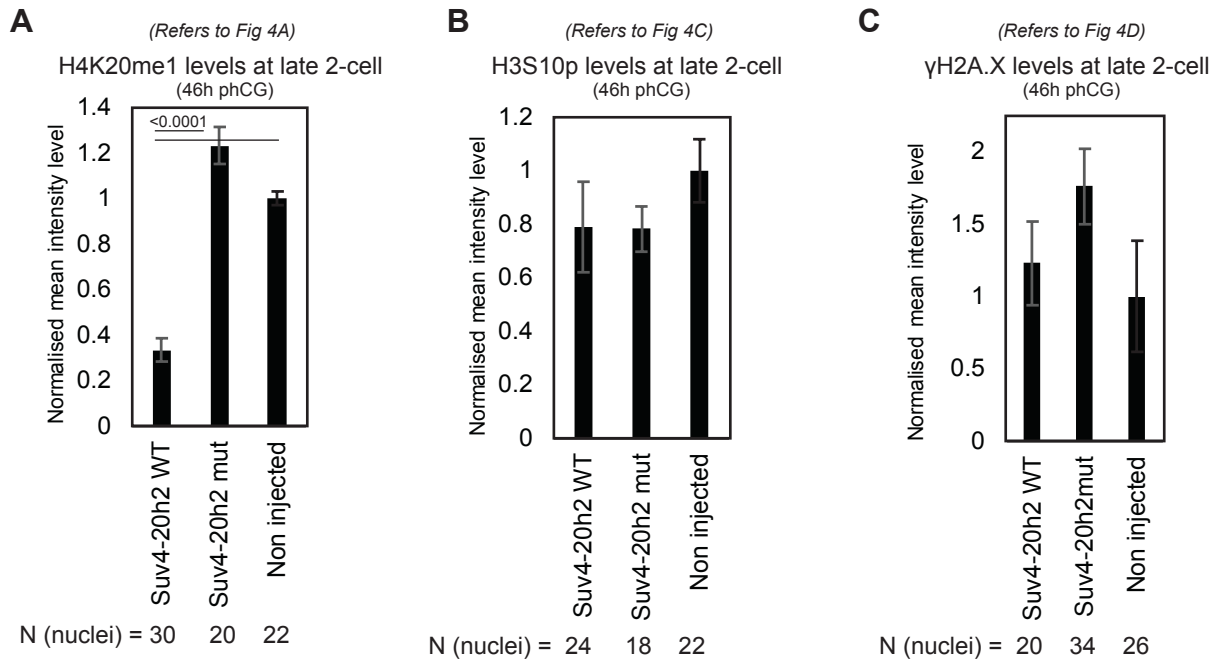
Quantification of CHK1 (**A-D**) and CHK2p (**E-F**) in early (25h phCG) and late (29h phCG) S-phase at the zygote stage in the maternal and paternal pronuclei in embryos expressing Suv4-20h2WT and Suv4-20h2MUT as indicated, compared to Non injected controls. The fluorescence intensity per pronucleus was normalized to the mean intensity in Non injected embryos following background subtraction and the number of embryos analyzed is shown to the left of the graphs. The mean averages of these values are shown (\pm SEM). The Mann-Whitney U test was applied and only significant differences are indicated.





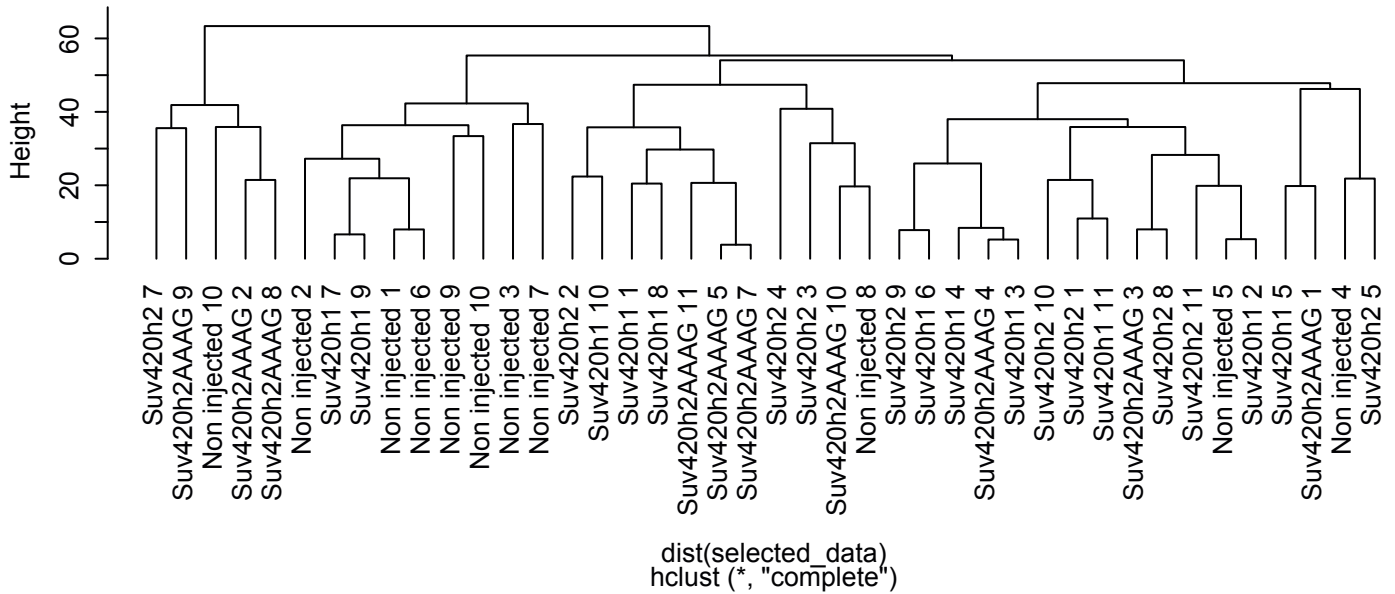


Eid et al. Supplementary Figure S4



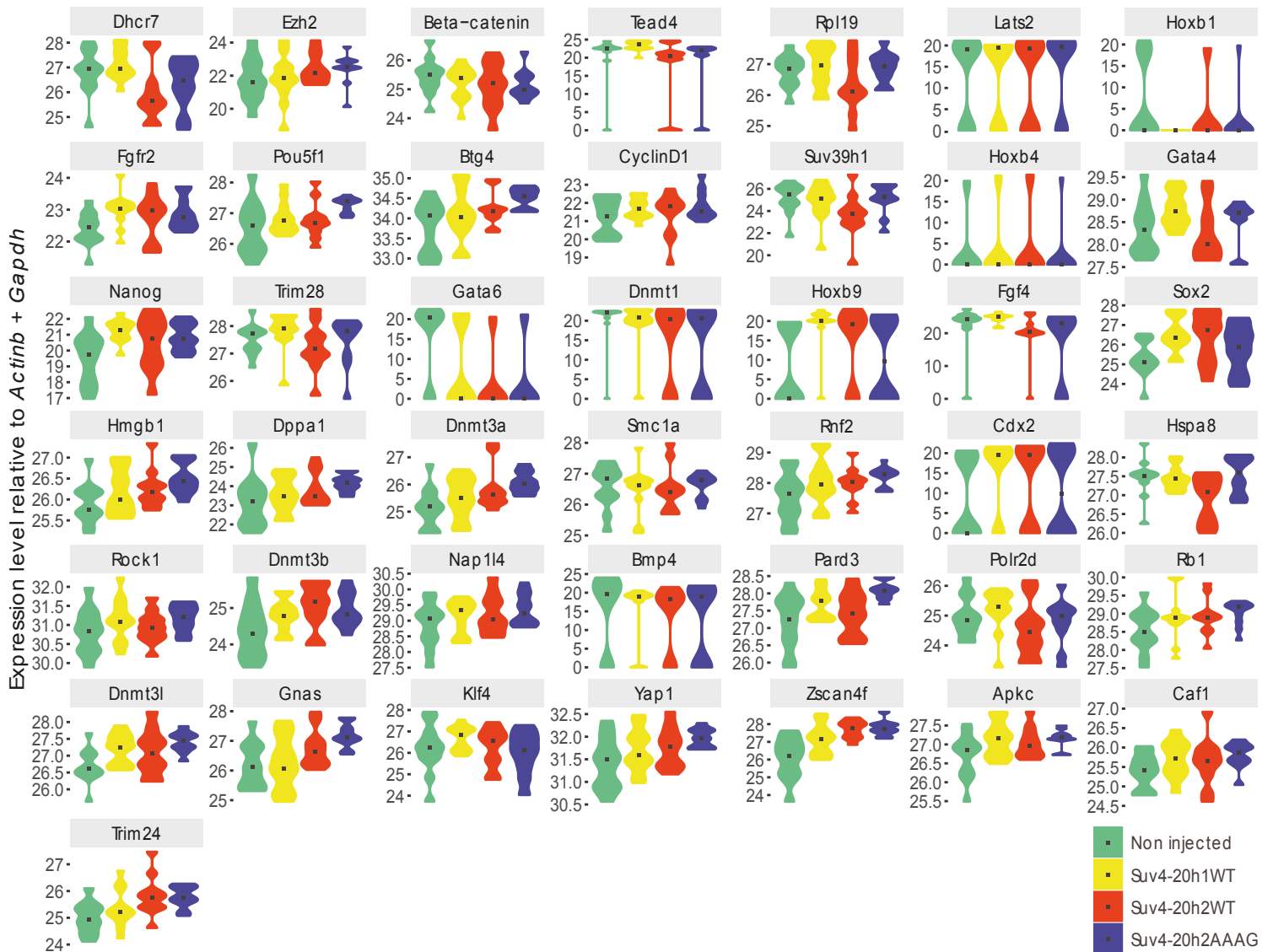
A

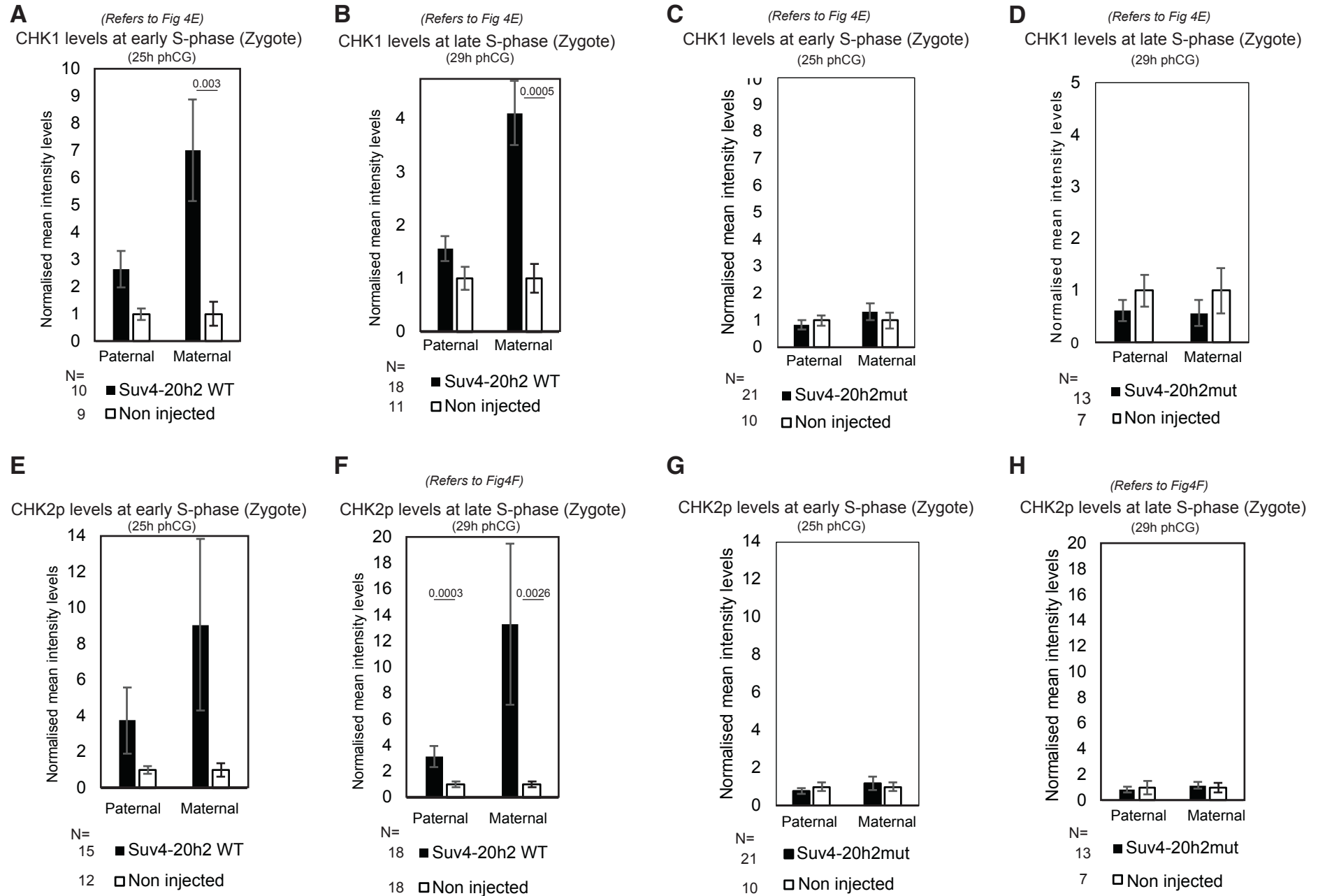
Cluster Dendrogram of 45 gene expression at EGA (2-cell, 44h phCG)



B

Gene expression of 43 genes at EGA (2-cell, 44h phCG)





Supplementary Table 1

Gene	Taqman probe	Function
<i>beta-catenin</i>	Mm00483039_m1	Signalling, polarity, axis specification, proliferation
<i>hoxb1</i>	Mm00515118_g1	Homeobox b1, transcription factor
<i>hoxb4</i>	Mm00657964_m1	Homeobox b4, transcription factor
<i>hoxb9</i>	Mm01700220_m1	Homeobox b9, transcription factor
<i>yap1</i>	Mm01143263_m1	Centromere proximal location, chr 9
<i>lats2</i>	Mm00497217_m1	Large tumor suppressor 2 kinase, cell division, Hippo pathway
<i>brg1</i>	Mm01151948_m1	ATPase subunit, remodelling complex, required for zygotic genome activation
<i>rock1</i>	Mm00485745_m1	Signalling kinase, Centromere proximal location, chr 18
<i>rb</i>	Mm00485586_m1	Cell cycle, cell proliferation
<i>cyclin D1</i>	Mm00432359_m1	Cell cycle
<i>gapdh</i>	Mm99999915_g1	Housekeeping, internal control
<i>Actin-B</i>	Mm00478923_m1	Housekeeping, internal control
<i>Smc1a</i>	Mm00490624_m1	Cohesion, can bind to Suv4-20h2
<i>Nap114</i>	Mm01131802_m1	nucleosome assembly, maternally expressed
<i>Gnas</i>	Mm01242435_m1	G-protein alpha subunit (Gs- α), imprinted
<i>Dhcr7</i>	Mm00514571_m1	production of cholesterol from NADPH, imprinted
<i>Rpl19</i>	Mm02601633_g1	60S ribosomal protein
<i>Btg4</i>	Mm00517103_m1	B-cell translocation gene 4, G1 arrest
<i>Trim28</i>	Mm00495594_m1	TF interacts with HP1a and controls gene expression
<i>Hmgb1</i>	Mm00849805_gH	Chromatin remodeler, affects transcriptional expression
<i>Hspa8</i>	Mm01731394_gH	Protein folding
<i>Polr2d</i>	Mm00770943_m1	One the RNA pol 2 subunits
<i>Zscan4f</i>	Mm01234988_g1	Exclusive expression at 2-cell stage
<i>Tif1a/Trim24</i>	Mm00446831_m1	TF binds to (AF2) region of several nuclear receptors
<i>EZH2</i>	Mm00468464_m1	Enhancer of Zest homolog 2, Polycomb Repressive Complex 2
<i>Ring1b Rnf2</i>	Mm00803321_m1	Histone H2A ubiquitylation, Polycomb Repressive Complex 1
<i>Suv4-20H1</i>	Mm00523065_m1	Di- and tri- methyl transferase
<i>Suv4-20h2</i>	Mm00525366_m1	Di- and tri- methyl transferase
<i>Suv39H1</i>	Mm00468952_m1	Centromere proximal location, chr X
<i>Dnmt1</i>	Mm00599763_m1	Maintenance DNA methyltransferase
<i>Dnmt3a</i>	Mm00432881_m1	De novo DNA methyltransferase
<i>Dnmt3b</i>	Mm01240113_m1	De novo DNA methyltransferase
<i>Dnmt3l</i>	Mm00457635_m1	De novo DNA methylation, cofactor
<i>Pou5f1</i>	Mm03053917_g1	Oct4, transcription factors
<i>Nanog</i>	Mm02019550_s1	Transcription factor, zygotically expressed, pluripotency
<i>Sox2</i>	Mm03053810_s1	Transcription factor, ICM maintenance, essential for early development
<i>Gata4</i>	Mm00484689_m1	Transcription factor, specification of primitive endoderm
<i>Fgf4</i>	Mm00438917_m1	Signalling molecule
<i>Cdx2</i>	Mm01212280_m1	Transcription factor, trophoctoderm differentiation
<i>Gata6</i>	Mm00802636_m1	Transcription factor, centromere proximal location, chr 18
<i>Bmp4</i>	Mm00432087_m1	TGF signalling, lineage specification
<i>Fgfr2</i>	Mm01269930_m1	Proliferation and differentiation, signalling
<i>Dppa1</i>	Mm00626454_m1	Developmental pluripotency associated 1, zygotically expressed
<i>Tead4</i>	Mm01189836_m1	Hippo pathway (transcription enhancer factor)
<i>Klf4</i>	Mm00516104_m1	Transcription factor, pluripotency
<i>Pard3</i>	Mm00473929_m1	Partitioning defective 3, cell polarity
<i>Apkcl</i>	Mm00435769_m1	Atypical PKC, cell polarity
<i>caf1</i>	Mm00511230_m1	Chromatin assembly

Supplementary Table 2

*Fold changes calculated based on the median of gene expression values per experimental group compared to Suv420h2MUT controls, and the corresponding P-values are indicated.

** Significant P-values between Suv420h2WT and Suv420h2MUT controls are highlighted in beige. Note that when significant changes were observed for the Suv420h2WT experimental group, values were also significant for the other two groups, suggesting that effects are not-specific to Suv420h2WT expression.

	Non injected vs Suv420h2MUT Fold-change	Suv420h1WT vs Suv420h2MUT Fold-change	Suv420h2WT vs Suv420h2MUT Fold-change	Non-Injected vs Suv420h2MUT P-value	Suv420h1WT vs Suv420h2MUT P-value	Suv420h2WT vs Suv420h2MUT P-value
<i>ActinB</i>	1.04	0.99	0.99	0.97	0.97	0.74
<i>Ezh2</i>	0.65	0.66	1.04	0.31	0.35	0.68
<i>Beta-catenin</i>	1.23	1.04	0.96	0.15	0.76	0.91
<i>Tead4</i>	7.26	49.16	0.74	0.15	0.01	0.43
<i>Rpl19</i>	0.93	1.03	0.68	0.60	0.76	0.09
<i>Gapdh</i>	0.96	1.01	1.01	0.97	0.97	0.74
<i>Hoxb1</i>	31.94	0.25	3.55	0.22	0.34	0.67
<i>Fgfr2</i>	0.71	1.07	0.90	0.08	0.51	1.00
<i>Pou5f1</i>	0.60	0.76	0.72	0.01	0.04	0.03
<i>Btg4</i>	0.60	0.74	0.82	0.002	0.08	0.12
<i>CyclinD1</i>	0.65	0.90	0.68	0.25	0.81	0.80
<i>Suv39h1</i>	1.03	0.84	0.38	0.92	0.76	0.09
<i>Hoxb4</i>	2.72	3.26	4.46	0.73	0.64	0.50
<i>Gata4</i>	0.91	1.21	0.84	0.51	0.39	0.39
<i>Nanog</i>	0.40	1.32	0.74	0.06	0.35	0.68
<i>Trim28</i>	1.02	1.27	0.86	0.60	0.39	0.39
<i>Gata6</i>	2029.69	10.07	0.86	0.04	0.45	0.87
<i>Dnmt1</i>	107.29	69.59	3.39	0.21	0.50	0.85
<i>Hoxb9</i>	0.37	414.04	17.33	0.42	0.03	0.53
<i>Fgf4</i>	48.68	256.76	8.58	0.10	0.01	0.60
<i>Sox2</i>	0.71	1.66	1.50	0.31	0.22	0.25
<i>Hmgb1</i>	0.66	0.81	0.84	0.003	0.17	0.17
<i>Dppa1</i>	0.51	0.66	0.82	0.01	0.13	0.25
<i>Dnmt3a</i>	0.52	0.67	0.87	0.00	0.08	0.12
<i>Smc1a</i>	0.96	0.91	0.94	0.60	0.47	0.31
<i>Cdx2</i>	0.32	18.82	3.91	0.42	0.61	0.84
<i>Hspa8</i>	0.95	0.95	0.65	0.81	0.76	0.06
<i>Rock1</i>	0.79	0.97	0.85	0.13	0.76	0.19
<i>Dnmt3b</i>	0.68	0.90	1.05	0.08	0.60	0.74
<i>Nap114</i>	0.71	0.85	0.91	0.17	0.47	0.68
<i>Bmp4</i>	9.88	10.67	1.91	0.27	0.86	0.46
<i>Pard3</i>	0.54	0.76	0.63	0.002	0.04	0.01
<i>Polr2d</i>	1.15	1.30	0.84	0.92	0.20	0.39
<i>Rb1</i>	0.73	0.93	0.94	0.04	0.35	0.44
<i>Dnmt3l</i>	0.61	0.87	0.81	0.00	0.28	0.25
<i>Gnas</i>	0.50	0.53	0.76	0.00	0.02	0.08
<i>Klf4</i>	1.14	1.62	1.19	0.70	0.11	0.63
<i>Yap1</i>	0.67	0.83	0.86	0.01	0.15	0.22
<i>Zscan4f</i>	0.31	0.64	0.95	0.00	0.05	0.91
<i>Apkc</i>	0.73	0.98	0.93	0.06	0.86	0.31
<i>Caf1</i>	0.76	0.94	0.89	0.03	0.70	0.31
<i>Dhcr7</i>	1.43	1.76	0.79	0.31	0.10	0.39
<i>Lats2</i>	0.33	0.33	0.21	0.37	0.30	0.42
<i>Rnf2</i>	0.64	0.90	0.86	0.05	0.35	0.31
<i>Trim24</i>	0.59	0.76	1.04	0.00	0.11	0.91

Discussion

Part 1 – The molecular basis of totipotent-like cells in culture

In the work presented in this thesis, we have addressed several open questions regarding the molecular and developmental identities of the 2-cell-like subpopulation that arises in mouse embryonic stem cell cultures. As described above, approximately 0.5% of a mouse ESC culture spontaneously transitions to a state reminiscent of the totipotent 2-cell stage embryo, and recapitulates several of its molecular features such as downregulation of OCT4 and expression of the *MERV-L* family of endogenous retroviruses (Macfarlan et al., 2012). However, before this work, it remained an open question how exactly this subpopulation arose from the rest of the pluripotent stem cell culture, and whether these cells recapitulated any other major features of the totipotent embryo such as its unique metabolism.

By employing several single cell approaches, our work has clarified the series of molecular events that occur during this transition and places into context their developmental identity. In addition to the 2-cell-like subpopulation, an altogether different heterogeneity of mouse ESCs recapitulating some features of the 2-cell stage embryo had been previously reported (Falco et al., 2007; Zalzman et al., 2010), although the relationship between these two states had never been established. The *Zscan4*⁺ fraction of mouse ESCs corresponds to ~5% of the culture at any given time, and its main defining feature consists in the expression of the 2-cell stage specific genes *Zscan4a-f*. By employing a combination of time-lapse microscopy, fluorescent reporters and a highly-precise single cell expression profiling assay on a set of differentially expressed genes between the three subpopulations, we were able to identify an ordered series of molecular events occurring during the transition from the ES cell to the 2-cell-like state (Rodriguez-Terrones et al., 2018).

The ESC to *Zscan4*⁺ cell transition

First, through an as yet unknown molecular mechanism, a small percentage of preferentially naïve pluripotent stem cells activates transcription of the *Zscan4* gene cluster. Shortly after this event, and again through an as yet undetermined molecular mechanism, key pluripotency-associated transcription factors such as OCT4 and SOX2 – but not NANOG – are downregulated at the protein level. Considering that barely any *Zscan4*⁺ cell is positive for OCT4 when assayed through immunofluorescence, this downregulation must be quick and might involve an active degradation mechanism. Once in this *Zscan4*⁺ positive state and with pluripotency-associated transcription

factors downregulated, at least two major things happen. First, *Zscan4*⁺ cells appear to reorganize their pericentromeric heterochromatin into a 'blob'-like structure within their nuclei, as described in (Akiyama et al., 2015); and second, *Zscan4*⁺ cells appear to initiate a shift in their central carbon metabolism, and exhibit significantly higher glucose consumption in spite of lower oxygen consumption (Rodriguez-Terrones et al., unpublished). Both in our targeted transcriptomic data and in the genome-wide RNA-seq data that we reanalyzed, a considerable fraction – but not the entirety of the 2-cell-like transcriptome – is upregulated once in this *Zscan4*⁺ state. Despite these transcriptional changes – including some involving 2-cell stage specific genes – no upregulation of *MERV-L* transcripts or reporters is observed, and the chromatin over their LTRs remains inaccessible.

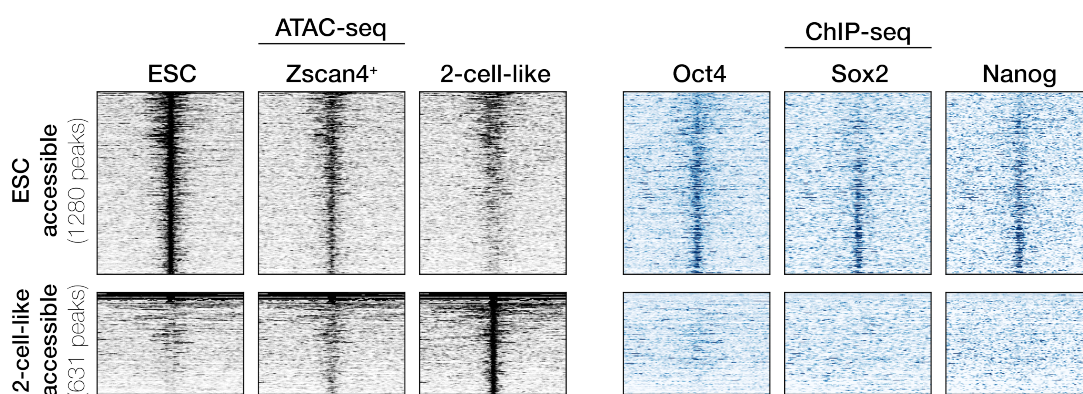


Figure 4. Heatmaps representing ATAC-seq and ChIP-seq signals over the differentially accessible regions between ESCs and 2-cell-like cells. ChIP-seq profiles correspond to the ESC state.

Interestingly, a number of genomic sites across the genome already exhibit a sharp drop in chromatin accessibility in the *Zscan4*⁺ state, mostly reflecting the loss of binding pluripotency-associated transcription factors (Figure 4, unpublished observation). Indeed, this observation is particularly surprising, given that the forced downregulation of OCT4 outside of a differentiation context results in cell death, both in our experiments (Rodriguez-Terrones et al., 2018) and in other reports (Chen et al., 2012). Given that our Rex1-reporter experiments also suggest that entry into the *Zscan4*⁺ state is independent of differentiation, this suggests that this transition involves an altogether different pathway out of pluripotency. Although the precise *in vivo* developmental analogue of the *Zscan4*⁺ subpopulation remains to be established, our Rex1-reporter work suggests that – if it exists – it must be a cellular state developmentally closer to the

pre-implantation epiblast (i.e. naïve pluripotency) rather than to the post-implantation epiblast (i.e. primed pluripotency). Indeed, in the mouse embryo, *Zscan4* expression occurs in parallel to *MERV-L* expression (Figure 5, unpublished observation), raising the questions as to whether a developmental equivalent of this *Zscan4*⁺ yet *MERV-L*⁻ state exists at all.

The *Zscan4*⁺ to 2-cell-like cell transition

Mouse embryonic stem cells have been widely reported to be able to transition in and out of the *Zscan4*⁺ state over the course of several cell cycles. Indeed, this state is essential for the self-renewal of the pluripotent culture, as it permits telomere elongation and the preservation of genomic stability. Indeed, our time-lapse microscopy experiments permitted us to confirm this observation, and in addition provided conclusive proof that a small subpopulation of *Zscan4*⁺ cells – around 1 in every 10 – would activate *MERV-L* expression rather than revert back to the ESC state. It is unclear, however, what occurs to this double positive *Zscan4*⁺ and *MERV-L*⁺ state – to which throughout this thesis I have referred to as the 2-cell-like state – and whether there are any distinct molecular states beyond *MERV-L* activation.

As previously mentioned, the LTRs of the *MERV-L* family of retrotransposons serve as promoters (Franke et al., 2017; Peaston et al., 2004) for many of the genes activated at the time of zygotic genome activation in the 2-cell stage embryo, following the opening of their chromatin at the early 2-cell stage (Wu et al., 2016). By reanalyzing previously published ATAC-seq data across these three subpopulations (Eckersley-Maslin et al., 2016), we were able to determine that opening of the chromatin at the *MERV-L* LTRs occurs in the 2-cell-like subpopulation but not in the *Zscan4*⁺ one, in agreement with the pattern of transcriptional activation. Recent reports (Hendrickson et al., 2017; De Iaco et al., 2017; Whiddon et al., 2017) have identified the transcription factor DUX as one of the main regulators of zygotic genome activation in mouse and human, acting in part through its binding to *MERV-L* LTRs. Indeed, the reanalysis and modelling of previously published

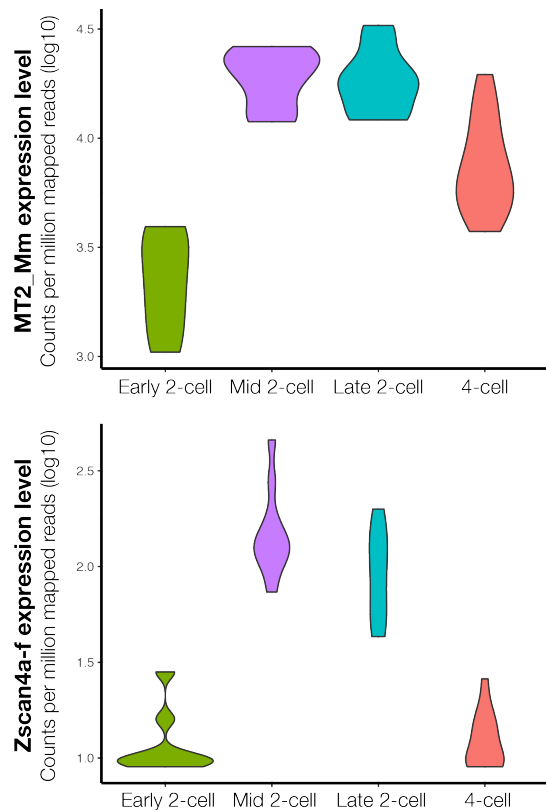


Figure 5. Violin plots representing *Zscan4* and *MERV-L* LTR expression levels over the course of early mouse development.

single-cell RNA-seq data in these three subpopulations suggests that *Dux* expression precedes *MERV-L* transcription, and that transcriptional activation of this transcription factor only occurs in a fraction of the highest-expressing *Zscan4*⁺ cells (Figure 6, unpublished observation). Thus, the joint analysis and modeling of these various datasets suggests that a small fraction of *Zscan4*⁺ cells expressing the highest *Zscan4* transcript levels activates expression of the DUX transcription factor, which subsequently binds to the *MERV-L* LTRs and leads to their transcriptional activation and chromatin opening.

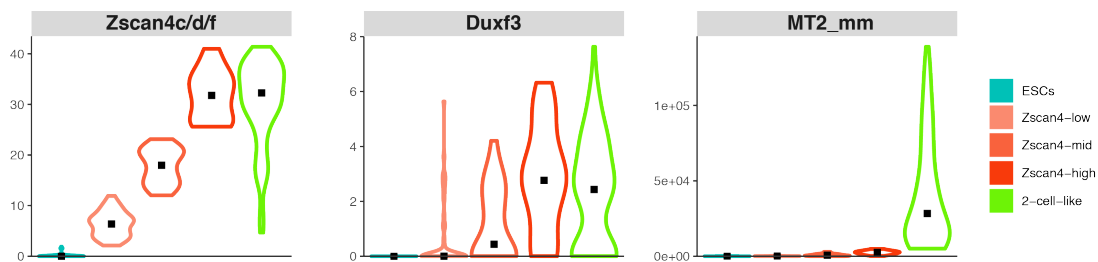


Figure 6. Violin plots representing *Zscan4*, *Dux* and *MERV-L* LTR expression levels over the course of the transition to the 2-cell-like state. Data reanalyzed from Eckersley-Maslin et al., 2016.

2-cell-like cells enable the identification of novel molecular features of totipotent cells

Indeed, as previously mentioned, the 2-cell-like subpopulation recapitulates several features of the 2-cell stage embryo, and this work permitted us to identify some novel ones. Before this work, the main defining features of this population included the downregulation of OCT4 and the transcriptional activation of *MERV-L* and other 2-cell stage specific genes and transcripts. After noticing that an overall reorganization of pericentromeric heterochromatin – i.e. major satellites – was evident in the DAPI stainings of 2-cell-like cells, Ishiuchi et al. identified another parallel to the 2-cell stage embryo and further determined that these genomic regions were being transcribed in a manner analogous to the situation seen in the zygote and the 2-cell stage embryo (Ishiuchi et al., 2015). Following up on these careful observations, it was determined that this reorganization of pericentromeric regions was reminiscent of that seen upon downregulation of replication-dependent chromatin assembly, and thus downregulation of the CAF-1 complex was identified as a powerful inducer of the 2-cell-like state. Following the publication of one of the first high-quality and comprehensive single-cell transcriptomes in the mouse early embryo (Deng et al., 2014), my contribution to the aforementioned work resided in identifying a significant enrichment in the 2-cell-like cell transcriptome of genes expressed exclusively in the 2-cell stage embryo, but not of those of any other developmental stage. As previously mentioned, the LTRs of the *MERV-L* family of retrotransposons serve as promoters for some of the genes activated at

the time of zygotic genome activation in the 2-cell stage embryo (Franke et al., 2017; Peaston et al., 2004), and these LTRs were determined in this study to be enriched in the vicinity of genes upregulated in the 2-cell-like state.

Having determined the sequence of molecular events leading to the activation of the MERV-L family of retrotransposons, we also undertook a loss-of-function screen – carried out mainly by Xavier Gaume – to identify negative regulators of this transition. At least five major molecular complexes were identified through these experiments: the spliceosome complex, the PRC1.6 complex, the CAF-1 complex and other DNA replication factors, the Ep400/Tip60 complex and the SUMO complex. While the effect of the downregulation of the spliceosome complex was not studied further, the mechanism of action of three of the four other complexes was to a certain extent determined. The Ep400/Tip60 complex proved to be particularly elusive to characterize in this context, and so far, our only functional insight is that it binds 2-cell-like upregulated genes in a mutually exclusive fashion to the PRC1.6 complex. Given the lack of gene expression changes seen in Ep400 and Tip60 catalytically inactive mutants (Acharya et al., 2017), it's likely that their mechanism of action is independent of their direct activity on chromatin, and might have a rather structural basis.

The PRC1.6 complex is a non-canonical variant of the Polycomb Repressive Complex 1, which represses target genes through the deposition of H2AK119 mono-ubiquitination. Indeed, the bioinformatic analyses that I undertook as part of that paper identified an enrichment of PRC1.6 bound genes in the set of upregulated genes in the 2-cell-like state, suggesting that this complex acts to repress at least part of the 2-cell stage transcriptional program. In addition, while we were not able to identify binding of either the PRC1.6 complexes or the Ep400/Tip60 complex to MERV-L elements, these analyses did identify the transcription factor DUX as one of its targets, providing a molecular explanation for the activation of these retrotransposons and the entry of these cells into the 2-cell-like state. Upon PRC1.6 knockdown, Dux might become de-repressed and go on to activate MERV-L transcription. Subsequently, further work from our lab and from Anne Dejean's lab identified a critical role for SUMO proteins in facilitating the recruitment of PRC1.6 to the Dux array (Cossec et al., 2018), indicating that both of these complexes act – at least partially – through the same pathway.

Finally, perhaps the most interesting of all these negative regulators of the 2-cell-like state is the CAF-1 complex, which is involved in the deposition of H3.1 and H3.2 histone variants at the time

of DNA-replication. Indeed, through its association with the DNA-replication machinery, this complex partially mediates the inheritance of the epigenetic state of the cell and thus of its precise cellular identity. The fact that the downregulation of this complex results in a considerable upregulation of the 2-cell-like transcriptome, raises the possibility that the clearance of the cell's epigenetic state might underlie the totipotent capacity seen at the beginning of development. Indeed, in this work (Ishiuchi et al., 2015), 2-cell stage embryos at an early phase of DNA-replication – that is, when gene-rich regions are replicated (Rhind and Gilbert, 2013) – were shown to be replicating in the absence of CAF-1. Thus, to a certain extent, the erasure of epigenetic memory might underlie the totipotent capacity of early-embryonic cells.

Part 2 – The molecular basis of mammalian totipotency *in vivo*

In this work, I have presented some preliminary work on a systematic comparison of pre-implantation development transcriptomes across mammals. Because of the low RNA inputs inherent to the pre-implantation embryo and the extraordinary scarcity of the embryos from some of the mammalian species we are assaying, I have adapted an existing single-cell RNA-sequencing protocol into providing both improved sensitivity and 5' information. Having verified the effectiveness of this approach, we have set out to benchmark a range of reaction conditions and identified an optimal protocol in terms of sensitivity and precision of the 5' end profiling for use in our transcriptomic comparison. The result is a highly sensitive transcriptional profiling approach that can also pinpoint transcription start sites and promoter architectures.

As a proof of concept, in this thesis I have used this transcriptional data to compare the transcriptional dynamics of a set of ancient – yet abundant – transposable element families between four mammalian species. Indeed, despite the divergent developmental timings guiding the pre-implantation development process across the mammalian species assayed, these ancient transposable elements appear to exhibit a mostly consistent expression pattern across species. In this context, equivalent developmental stages in terms of cell number are not necessarily equivalent in terms of the biological processes underlying them, and thus the conservation of the transcriptional dynamics appears to occur relative to the timing of the zygotic genome activation event and the degradation of the maternal transcriptome.

While the well-defined transcriptional profiles observed for L2, L3, MIR, L1M4 or the various DNA transposons discussed is undoubtedly not necessarily indicative of a functional role of these elements for the host organisms, it might rather be suggestive of a functional relevance of these stages and their regulatory context for the once active transposable elements. Some transposable elements such as ERV-L or HERV-H appear to achieve stage-specificity by relying on the binding of sequence specific transcription factors such as DUX (Hendrickson et al., 2017; De Iaco et al., 2017; Whiddon et al., 2017) or OCT4 (Lu et al., 2014) to activate their transcription. It remains an open question to which extent are transposable elements reliant on transcription factor binding to activate their expression, and whether they can achieve stage-specific expression during pre-implantation development through other mechanisms. Thus, the conserved transcriptional pattern observed for the analyzed elements might potentially reflect an ancestral co-option of a conserved stage-specific host factor by the once active transposable element.

Another possibility however, is that the global chromatin remodeling process that occurs during pre-implantation development is what permits transposable elements to exhibit stage-specific expression patterns. As Eid et al. show here, the absence of certain heterochromatic marks is essential for developmental progression in the early embryo, and it remains to be addressed whether any of the pathways involved in the chromatin remodeling process lead to the activation of transposable elements at this time. Instead, some evidence of the converse situation where transposable elements modulate the chromatin landscape does exist. Indeed, the transcriptional activation of L1 elements has been shown to affect chromatin accessibility, and their timely transcriptional activation has been determined to be essential for pre-implantation development (Jachowicz et al., 2017). In general, it remains to be addressed how does chromatin mediate the transcriptional activation of transposable elements during pre-implantation development.

Concluding remarks

Overall, the work carried out during my PhD has aimed to characterize the molecular features of mammalian totipotency in terms of its transcriptional profiles, its chromatin state and its metabolic properties. Further work to systematically determine which properties of totipotent cells are reproduced in totipotent-like cells *in vitro* will be needed before totipotency can be ultimately induced and stabilized in culture.

References

- Acharya, D., Hainer, S.J., Yoon, Y., Wang, F., Bach, I., Rivera-Pérez, J.A., and Fazio, T.G. (2017). KAT-Independent Gene Regulation by Tip60 Promotes ESC Self-Renewal but Not Pluripotency. *Cell Rep.* 19, 671–679.
- Akiyama, T., Xin, L., Oda, M., Sharov, A.A., Amano, M., Piao, Y., Cadet, J.S., Dudekula, D.B., Qian, Y., Wang, W., et al. (2015). Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. *DNA Res.* 22, 307–318.
- Arrell, V.L., Day, B.N., and Prather, R.S. (1991). The Transition from Maternal to Zygotic Control of Development Occurs during the 4-Cell Stage in the Domestic Pig, *Sus Scrofa*: Quantitative and Qualitative Aspects of Protein Synthesis¹. *Biol. Reprod.* 44, 62–68.
- Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCR-seq. *Nat. Commun.* 9, 2937.
- Bošković, A., Eid, A., Pontabry, J., Ishiuchi, T., Spiegelhalter, C., Raghu Ram, E.V.S., Meshorer, E., and Torres-Padilla, M.E. (2014). Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* 28, 1042–1047.
- Brinster, R.L. (1965). Studies in the development of mouse embryos in vitro. II. The effect of the energy source. *J. Exp. Zool.* 158, 59–68.
- Burton, A., and Torres-Padilla, M.-E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 722–734.
- Camous, S., Kopečný, V., and Fléchon, J.E. (1986). Autoradiographic detection of the earliest stage of [3H]-uridine incorporation into the cow embryo. *Biol. Cell* 58, 195–200.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626–635.
- Casser, E., Israel, S., Witten, A., Schulte, K., Schlatt, S., Nordhoff, V., and Boiani, M. (2017). Totipotency segregates between the sister blastomeres of two-cell stage mouse embryos. *Sci. Rep.* 7, 8299.
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* 113, 643–655.
- Chan, A.W. (2000). Clonal Propagation of Primate Offspring by Embryo Splitting. *Science* (80-.).
- Chatot, C.L., Lewis-Williams, J., Torres, I., and Ziomek, C.A. (1994). One-minute exposure of 4-cell mouse embryos to glucose overcomes morula block in CZB medium. *Mol. Reprod. Dev.* 37, 407–412.
- Chen, T., Du, J., and Lu, G. (2012). Cell growth arrest and apoptosis induced by Oct4 or Nanog knockdown in mouse embryonic stem cells: a possible role of Trp53. *Mol. Biol. Rep.* 39, 1855–1861.

- Christians, E., Rao, V.H., and Renard, J.P. (1994). Sequential Acquisition of Transcriptional Control during Early Embryonic Development in the Rabbit. *Dev. Biol.* 164, 160–172.
- Condic, M.L. (2014). Totipotency: what it is and what it is not. *Stem Cells Dev.* 23, 796–812.
- Cossec, J.-C., Theurillat, I., Chica, C., Búa Aguín, S., Gaume, X., Andrieux, A., Iturbide, A., Jouvion, G., Li, H., Bossis, G., et al. (2018). SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* 23, 742–757.e8.
- Cvetesic, N., Leitch, H.G., Borkowska, M., Müller, F., Carninci, P., Hajkova, P., and Lenhard, B. (2018). SLIC-CAGE: high-resolution transcription start site mapping using nanogram-levels of total RNA. *Genome Res.* 28, 1943–1956.
- Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. *Science* (80-.). 343, 193–196.
- Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R.V., Milagre, I., et al. (2016). MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep.* 17, 179–192.
- Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* 1.
- Eid, A., Rodriguez-Terrones, D., Burton, A., and Torres-Padilla, M.E. (2016). SUV4-20 activity in the preimplantation mouse embryo controls timely replication. *Genes Dev.*
- Evsikov, A. V., Graber, J.H., Brockman, J.M., Hampl, A., Holbrook, A.E., Singh, P., Eppig, J.J., Solter, D., and Knowles, B.B. (2006). Cracking the egg: Molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes Dev.* 20, 2713–2727.
- Fadloun, A., Eid, A., and Torres-Padilla, M.-E. (2013a). Mechanisms and dynamics of heterochromatin formation during mammalian development: closed paths and open questions.
- Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.E. (2013b). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* 20, 332–338.
- Falco, G., Lee, S.L., Stanghellini, I., Basse, U.C., Hamatani, T., and Ko, M.S.H. (2007). Zscan4: A novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.* 307, 539–550.
- Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571.
- Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F., Kuzman, M., Fulka, H., Cernohorska, M., Urbanova, J., et al. (2017). Long terminal repeats power evolution

of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res.* 27, 1384–1394.

Frankenberg, S. (2018). Pre-gastrula Development of Non-eutherian Mammals. In *Current Topics in Developmental Biology*, p.

Gardner, D.K., and Leese, H.J. (1990). Concentrations of nutrients in mouse oviduct fluid and their effects on embryo development and metabolism in vitro. *Reproduction* 88.

Goerner-Potvin, P., and Bourque, G. (2018). Computational tools to unmask transposable elements. *Nat. Rev. Genet.* 19, 688–704.

Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 16, 135–141.

Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934.

Houghton, F.D., Thompson, J.G., Kennedy, C.J., and Leese, H.J. (1996). Oxygen consumption and energy metabolism of the early mouse embryo. *Mol. Reprod. Dev.* 44, 476–485.

Hughes, R.L. (1974). Morphological Studies on Implantation in Marsupials. *Reproduction* 39.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945.

Ishiuchi, T., and Torres-Padilla, M.-E. (2013). Towards an understanding of the regulatory mechanisms of totipotency. *Curr. Opin. Genet. Dev.* 23, 512–518.

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.-E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671.

Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* 49, 1502–1510.

Johnson, W.H., Loskutoff, N.M., Plante, Y., and Betteridge, K.J. (1995). Production of four identical calves by the separation of blastomeres from an in vitro derived four-cell embryo. *Vet. Rec.*

Kaneko, K.J. (2016). Metabolism of Preimplantation Embryo Development: A Bystander or an Active Participant? *Curr. Top. Dev. Biol.* 120, 259–310.

Kelly, S.J. (1975). Studies of the potency of early cleavage blastomeres of the mouse. In *The Early Development of Mammals*, pp. 97–106.

Kemp, T.S. (Thomas S. (2005). *The origin and evolution of mammals* (Oxford University Press).

- Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* 132, 1049–1061.
- Kopečný, V., Fléchon, J.E., Camous, S., and Fulka, J. (1989). Nucleologenesis and the onset of transcription in the eight-cell bovine embryo: Fine-structural autoradiographic study. *Mol. Reprod. Dev.* 1, 79–90.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* 20, 311–316.
- Lu, X., Sachs, F., Ramsay, L., Jacques, P.-É., Göke, J., Bourque, G., and Ng, H.-H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* 21, 423–425.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63.
- Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214.
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* 9, 625–635.
- Memili, E., and First, N.L. (1999). Control of gene expression at the onset of bovine embryonic development. *Biol. Reprod.* 61, 1198–1207.
- Mintz, B. (1964). Synthetic processes and early development in the Mammalian egg. *J. Exp. Zool.* 157, 85–100.
- Moore, N.W., Adams, C.E., and Rowson, L.E. (1968). Developmental potential of single blastomeres of the rabbit egg. *J. Reprod. Fertil.* 17, 527–531.
- Morgani, S.M., Canham, M.A., Nichols, J., Sharov, A.A., Migueles, R.P., Ko, M.S.H., and Brickman, J.M. (2013). Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. *Cell Rep.*
- Nagaraj, R., Sharpley, M.S., Chi, F., Braas, D., Zhou, Y., Kim, R., Clark, A.T., and Banerjee, U. (2017). Nuclear Localization of Mitochondrial TCA Cycle Enzymes as a Critical Step in Mammalian Zygotic Genome Activation. *Cell* 168, 210–223.e11.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* 95, 379–391.
- Pavlicek, A., and Jurka, J. (2006). Ancient Transposable Elements, Processed Pseudogenes, and Endogenous Retroviruses. In *Genomic Disorders*, (Totowa, NJ: Humana Press), pp. 57–72.

- Peaston, A.E., Evsikov, A. V, Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7, 597–606.
- Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* 174, 391–405.e19.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Posfai, E., Petropoulos, S., de Barros, F.R.O., Schell, J.P., Jurisica, I., Sandberg, R., Lanner, F., and Rossant, J. (2017). Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *Elife* 6.
- Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. *Cold Spring Harb. Perspect. Biol.* 5, a010132.
- Roberts, R.M., Green, J.A., and Schulz, L.C. (2016). The evolution of the placenta. *Reproduction* 152, R179–R189.
- Rodriguez-Terrones, D., and Torres-Padilla, M.E. (2018). Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet.*
- Rodriguez-Terrones, D., Gaume, X., Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* 50, 106–119.
- Rossant, J. (1976). Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *Embryol. Exp. Morph.*
- Rossant, J., and Vijn, K.M. (1980). Ability of outside cells from preimplantation mouse embryos to form inner cell mass derivatives. *Dev. Biol.* 76, 475–482.
- Saito, S., and Niemann, H. (1991). Effects of Extracellular Matrices and Growth Factors on the Development of Isolated Porcine Blastomeres1. *Biol. Reprod.* 44, 927–936.
- Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat. Cell Biol.* 12, 853–862.
- Schramm, R.D., and Bavister, B.D. (1999). Onset of Nucleolar and Extranucleolar Transcription and Expression of Fibrillarin in Macaque Embryos Developing In Vitro1. *Biol. Reprod.* 60, 721–728.
- Selwood, L., and Johnson, M.H. (2006). Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: Evolution and origins. *BioEssays.*
- Smit, A.F.A., Tóth, G., Riggs, A.D., and Jurka, J. (1995). Ancestral, Mammalian-wide Subfamilies of LINE-1 Repetitive Sequences. *J. Mol. Biol.* 246, 401–417.
- Smith, K.K. (2001). Early development of the neural plate, neural crest and facial region

of marsupials. *J. Anat.*

Soumillon, M., Cacchiarelli, D., Semrau, S., Oudenaarden, A. van, and Mikkelsen, T.S. (2014). Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *BioRxiv* 003236.

Tabansky, I., Lenarcic, A., Draft, R.W., Loulier, K., Keskin, D.B., Rosains, J., Rivera-Feliciano, J., Lichtman, J.W., Livet, J., Stern, J.N.H., et al. (2013). Developmental bias in cleavage-stage mouse blastomeres. *Curr. Biol.* 23, 21–31.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126, 663–676.

Tarkowski, A.K. (1959). Experiments on the development of isolated blastomeres of mouse eggs. *Nature*.

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

Tománek, M., Kopečný, V., and Kaňka, J. (1989). Genome reactivation in developing early pig embryos: an ultrastructural and autoradiographic analysis. *Anat. Embryol. (Berl)*. 180, 309–316.

Vaughan, T.A., Ryan, J.M., and Czaplewski, N.J. (2013). *Mammalogy* (Jones & Bartlett Learning).

Wang, H., and Dey, S.K. (2006). Roadmap to embryo implantation: clues from mouse models. *Nat. Rev. Genet.* 7, 185–199.

Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* 49, 935–940.

Willadsen, B.S.M. (1981). The developmental capacity of blastomeres from 4- and 8-cell sheep embryos. *J. Embryol. Exp. Morphol.*

Williamson, T.E., Brusatte, S.L., and Wilson, G.P. (2014). The origin and early evolution of metatherian mammals: the Cretaceous record Launched to accelerate biodiversity research. *Zookeys* 465, 1–76.

Woodland, H.R., and Graham, C.F. (1969). RNA Synthesis during Early Development of the Mouse. *Nature* 221, 327–332.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. *Biochem. Soc. Trans.*

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* 534, 652–657.

Zalzman, M., Falco, G., Sharova, L. V, Nishiyama, A., Thomas, M., Lee, S.-L., Stagg, C.A., Hoang, H.G., Yang, H.-T., Indig, F.E., et al. (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* 464, 858–863.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., et al. (2012). HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC

transition. *EMBO J.* 31, 2103–2116.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* 65, 631–643.e4.

Annex I – Review article

Review

Nimble and Ready to Mingle: Transposon Outbursts of Early Development

Diego Rodriguez-Terrones¹ and Maria-Elena Torres-Padilla^{1,*}

Transposable elements are the largest individual constituent of mammalian genomes. These elements are highly diverse, a consequence of the multiplicity of genomic habitats that they inhabit and of the complex evolutionary histories that they have developed therein. Intriguingly, a surge of transposable element transcription occurs during mammalian preimplantation development, contributing to the establishment of totipotency and pluripotency and to the activation of the embryonic genome. However, it remains an open question how such an evolutionarily divergent set can mediate such conserved developmental processes. Here, we review transposable element diversity across mammals and their evolutionary significance. We also discuss the implications that their high evolutionary divergence has for the regulation of preimplantation development across mammals.

The Diversity of Mammalian Transposons

Despite their classical rendition as selfish, parasitic genetic elements [1,2], transposable elements are major drivers of genome evolution and fundamental coordinators of regulatory function. Long suspected to harbor regulatory roles [3–5], transposable elements have recently surfaced as conspicuous actors behind the transcriptional and epigenomic remodeling processes underlying development and disease. These elements are a nimble bunch, using a compact set of molecular mechanisms in their replication yet quickly diverging in their abundance, even among closely related species (Box 1). Broadly, transposable elements can be classified into two classes (Figure 1), according to the molecular intermediate involved in their transposition [6]: class I elements or retrotransposons, which transpose through an RNA intermediate; and class II elements or DNA transposons, which directly excerpt, as DNA molecules, from a donor site in the genome before reintegrating elsewhere at a target site.

Box 1. Transposable Elements Exhibit Limited Conservation across Mammals

The transposable element content of mammalian genomes is exceedingly variable between species (see Figure 2 in the main text). Even without considering the diversity of individual families, mammalian genomes exhibit marked differences in the abundance and activity of the different classes, orders, and superfamilies of transposable elements. For instance, while DNA transposons have long been extinct in most mammals, they appear to have recently flourished in some lineages of bats. In contrast to other vertebrate genomes, LTR retrotransposons are also extinct across mammals, but have proliferated primarily as endogenous retroviruses. Given their viral origin, endogenous retrovirus content differs on a species-by-species basis and reflects the history of exposure of the lineage of the host to different exogenous retroviruses. Nevertheless, even purely vertically transmitted elements, such as the mammalian-wide ERV-L family, exhibit variability in their transposition activities, even among closely related lineages, such as mice and rats. LINE elements are also mainly vertically transmitted, yet have evolved distinct lineage-specific forms and molecular features, as can be observed when comparing human and mouse L1 elements. Despite being by far the most abundant transposable element in humans and mice, L1 appear to have died out in some lineages of rodents and bats and to be undergoing a gradual displacement by the RTE clade in ruminants. Finally, numerous SINEs have emerged *de novo* over the course of mammalian evolution and, therefore, are typically restricted to the lineage in which they appear. Thus, mammalian species exhibit limited conservation in their transposable element content and it remains an open question to what extent their functional co-options are conserved.

Highlights

Transposable elements are extensively transcribed during early development in a stage-specific manner in both mice and humans.

Stage-specific transcriptional activity of transposable elements reflects a complex interplay between their co-option by host regulatory programs and their own selective pressure to replicate within the germline.

Despite their limited conservation across mammals, transposable elements have been co-opted into regulatory functions during early development.

The ERV-L family of endogenous retroviruses can function as alternative promoters during embryonic genome activation in both mice and humans, and despite its lack of sequence conservation, an orthologous transcription factor binds their long terminal repeat (LTR) in both species.

The timely transcription of some transposable elements, such as L1 in mice, is necessary for developmental progression.

¹Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, D-81377 München, Germany

*Correspondence: torres-padilla@helmholtz-muenchen.de (M.-E. Torres-Padilla).

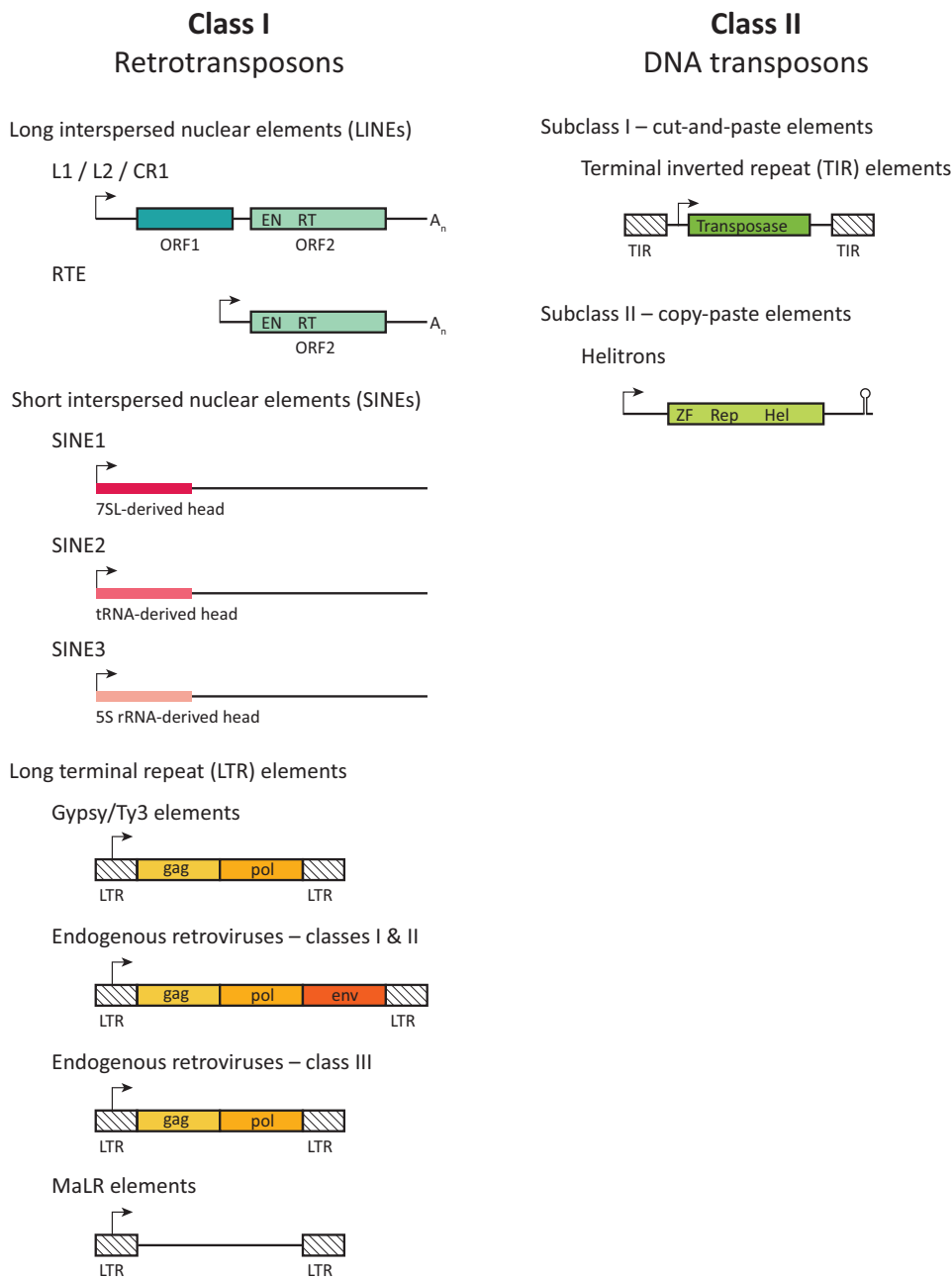
In mice and humans [7,8], transposable elements exhibit a surge in transcriptional activity during preimplantation development, which is thought to contribute to cellular plasticity [9,10] and to the activation of the embryonic genome [11–13]. However, given their limited conservation, it remains an open question how such an evolutionarily divergent set can mediate such conserved developmental processes. Here, we first summarize the diversity, abundance, and activity of the two transposon classes across mammals before proceeding to discuss their evolutionary significance and the implications that their high evolutionary divergence has for the regulation of preimplantation development across mammals.

DNA Transposons Are Ancient, Predominantly Extinct Inhabitants of Mammalian Genomes

DNA transposons can be classified into two subclasses depending on the number of DNA strands cut during transposition [14]. Whenever both strands are cleaved, elements are classified into Subclass I, which currently comprises 21 superfamilies [15]. These elements replicate via the classical ‘cut-and-paste’ mechanism, whereby a transposon-encoded protein, typically a transposase, catalyzes the excision of the transposon from the host genome and its subsequent reintegration at a different genomic site. Given their perilous ‘cut-and-paste’ nature, Subclass I elements rely on the timely action of their transposases for their proliferation. As shown in maize [16], some of these elements increase their copy number by excising from the replicated genome during DNA replication and reintegrating ahead of the replication fork. In an alternative manner, P elements in *Drosophila melanogaster* also make use of the S-phase cellular machinery to increase their copy number [17]. After excising from the genome to undertake transposition, these elements rely on homologous recombination to repair the double-stranded break left behind by their excision, consequently restoring the donor copy. Thus, Subclass I elements make prompt use of the cellular replication and repair machineries to increase their copy number [18].

At least five different superfamilies of Subclass I DNA transposons occur within the human genome, which they comprise almost 3% of [19]. Specifically, Tc1/mariner and hAT are the dominant superfamilies and account for approximately 97% of the 380 000 individual elements [20]. The last detected waves of Subclass I DNA transposon activity within the human genome appear to have occurred between ~85 and ~40 million years ago (Mya), during the early evolution of the primate lineage [20] (Figure 2). This period saw a tremendous amount of Subclass I activity, with up to one-third of the 125 DNA transposon families found in the human genome being active, but subsequently going extinct. Intriguingly, the same trend appears to apply to all other mammalian lineages studied so far [19,21,22], in which DNA transposon activity appears to have concurrently ceased ~40–50 Mya [18]. The Vespertilionidae family of bats, which ranks among the most diverse and species-rich mammalian families, is a particularly striking exception to this trend and is the only mammalian lineage wherein recent activity of Subclass I DNA transposons has been detected [23–25].

Conversely, whenever a DNA transposon requires only a single DNA strand to be cleaved for its transposition, it is termed a Subclass II element [14]. Two distinct orders of Subclass II elements exist: Mavericks and Helitrons, both of which replicate through a ‘copy-paste’ mechanism that does not involve full excision of both DNA strands from the host genome. Intriguingly, Subclass II elements do not appear to have been particularly successful in expanding throughout the mammalian lineage. Whereas Helitron-related sequences are detectable but generally found in very low abundance [26,27], Mavericks have not been detected within mammalian genomes [26,28,29]. Remarkably, the single known occurrence of Subclass II activity within mammals is once again found in the Vespertilionidae [30] (Figure 2). Approximately 6% of the genome of the



Trends in Genetics

Figure 1. Schematic Representations of the Different Transposable Element Groups Found in Mammalian Genomes. Colored boxes demarcate coding regions and the protein domains encoded within.

little brown bat (*Myotis lucifugus*) comprises Helitron elements, some of which replicated as recently as 1.8 Mya [31]. These transposition events have occurred with deep evolutionary significance, often capturing functional genomic elements, including promoters, 5'-untranslated regions (UTRs), and first exons, and introducing them at different genomic sites [31].

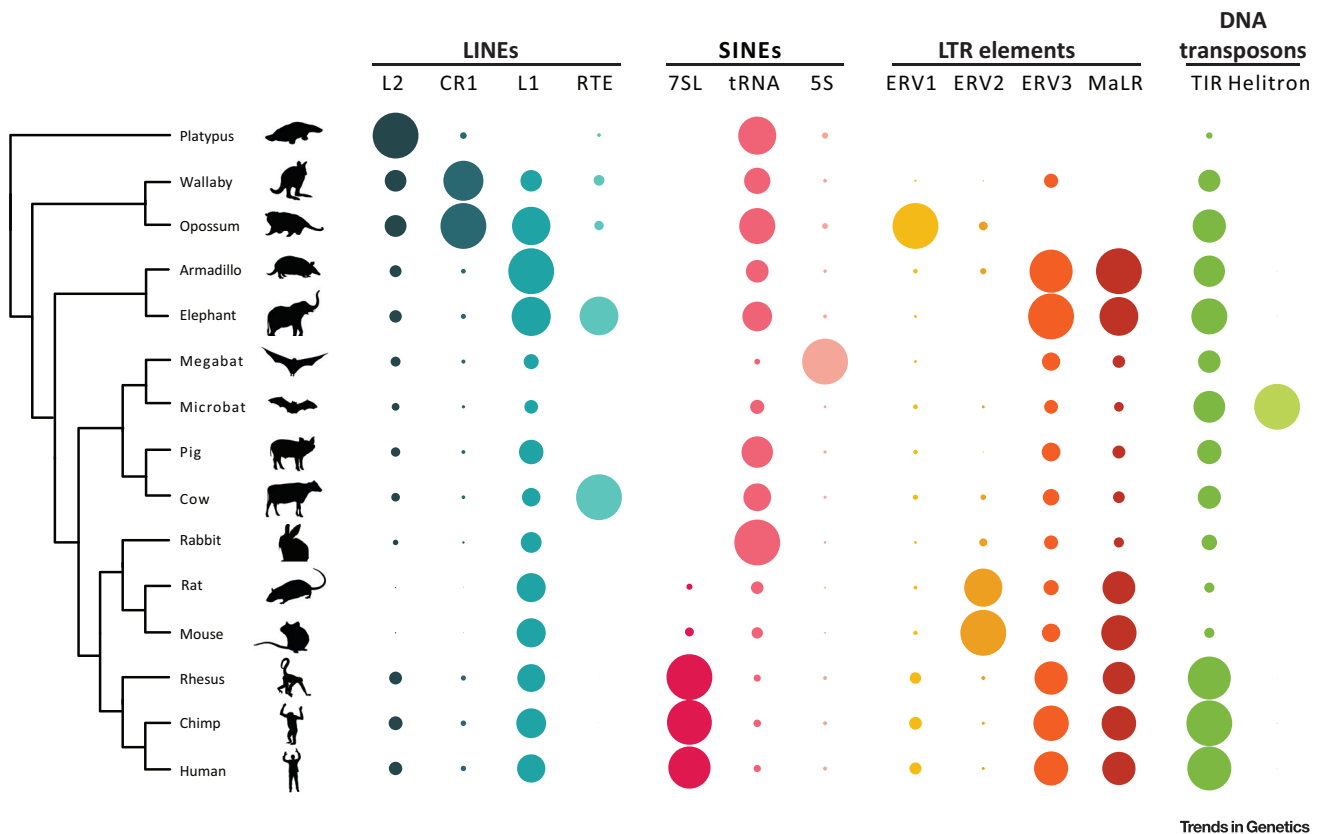


Figure 2. The Diversity of Transposable Element Landscapes in Mammals. The diameter of each dot is proportional to the absolute abundance in base pairs of the respective transposable element order or superfamily, as determined by RepeatMasker. The largest dot corresponds to the species with the highest genomic content of a particular transposable element group. The abundance of each transposon group was calculated by RepeatMasker v4.07 using Rebase library 20170127. The phylogenetic relationships shown in the dendrogram are based on [104]. Branch lengths are not representative of evolutionary distance. Abbreviations: ERV, endogenous retrovirus; LINE, long interspersed nuclear element; LTR, long terminal repeat; SINE, short interspersed nuclear element; TIR, terminal inverted repeat.

Therefore, although DNA transposons of both subclasses are generally extinct and represent a minority within the mammalian repertoire of transposable elements (Figure 1), they have contributed significantly to the evolution of mammalian genomic landscapes. Overall, DNA transposons have remained largely unexplored in the context of early mammalian development. While it is well established that specific retrotransposon families are actively transcribed during specific windows of preimplantation development [8], DNA transposons have not yet been characterized in this context.

Retrotransposons Dominate the Transposable Element Repertoire of Mammalian Genomes

In contrast to most DNA transposons, retrotransposons typically encode several protein domains harboring a wide array of biochemical activities, including the RNA-dependent DNA polymerase activity catalyzed by their idiosyncratic reverse transcriptase (RT). This RT is responsible for converting the RNA intermediate into the DNA molecule that will be integrated into the host genome. Thus, because of their dependency on transcription, elements within this class are of a strict ‘copy-paste’ nature. Retrotransposons can be classified into five orders based on an ensemble of their genetic structure, molecular mechanisms, and RT phylogeny

[14]: (i) long terminal repeat (LTR) elements; (ii) *Dictyostelium* intermediate repeat sequences (DIRSs); (iii) Penelope-like elements (PLEs); (iv) long interspersed nuclear elements (LINEs); and (v) short interspersed nuclear elements (SINEs). However, given that only ancestral activity of DIRSs and PLEs has been detected within mammalian genomes [27], these will not be discussed further here.

LTR Elements

As their name conveys, LTR elements are characterized by the presence of initially identical, equally oriented 300–1000 base-pair (bp) sequences on both ends of the retrotransposon [32]. These pairs of sequences, which are termed LTRs themselves, provide a promoter at the 5' end, and transcriptional termination and polyadenylation signals at the 3' end of the retrotransposon. Although the particularities of their sequential arrangement and their disposition into distinct open reading frames (ORFs) differ depending on the family, LTR elements typically encode a *gag* ORF with structural and nucleic acid-binding functions, and a *pol* ORF with protease, RT, RNase-H, and integrase protein domains [33]. Based on the divergence of their RT and the organization of their functional domains, at least four superfamilies exist within the LTR order: (i) Copia/Ty1; (ii) Gypsy/Ty3; (iii) Bel/Pao; and (iv) endogenous retroviruses (ERV) [14,33].

Copia/Ty1, Gypsy/Ty3, and Bel/Pao represent three distinct lineages based on their RT phylogeny [34]. Despite the wide distribution of these three superfamilies across eukaryotes, only Gypsy/Ty3 elements are detected within mammalian genomes, albeit in very low numbers [26,35]. By contrast, ERVs are among the most abundant transposable elements found in the human genome, of which they comprise ~8% [19]. Endogenous and exogenous vertebrate retroviruses share a startling resemblance to LTR retrotransposons and are presumed to have arisen from a Gypsy/Ty3 element that acquired an envelope gene [36,37]. This envelope gene, typically termed *env*, encodes the transmembrane domain that enables these retroviruses to undergo an extracellular infectious phase and determines the tropism of the virus [38]. Although this tropism is typically limited to somatic cells, an exogenous retrovirus might occasionally infect the germline and be vertically transmitted to the offspring. Over evolutionary timescales, numerous exogenous retroviruses have invaded mammalian genomes and have settled within them, establishing novel ERV families.

Vertebrate ERV families can be further classified based on their RT phylogeny into four distinct classes, commonly referred to as Class I, Class II, Class III, and the recently described reptile-specific Class IV [38,39]. Each of these classes relates to a specific group of exogenous retrovirus genera, in addition to their endogenous counterparts (Box 2). Unfortunately, while this four-class system adequately represents the phylogenetic relationships between RT sequences, it provides few insights into the phylogenetic relationships of full-length elements. A recent attempt [40] to systematically classify human ERVs led to the realization that mosaic forms between classes are common, and often form as a result of secondary integrations and recombination events between elements of distinct families and even distinct classes. Thus, while ERVs are commonly discussed in the context of this four-class system, the phylogenetic relationships between ERV families frequently involve recombination events and, therefore, their constituent sequences often have diverse origins.

Given their co-option as alternative promoters during embryonic genome activation [41] and as regulatory elements within gene regulatory networks in embryonic stem cells (ESCs) [42], LTR elements have emerged as potential leading regulators of early mammalian development, which we discuss further below.

Box 2. Endogenous Retroviruses of Mice and Humans

Mouse and human genomes represent contrasting scenarios of ERV activity (see [Figure 2](#) in the main text). While Classes I, II, and III have active members in mice, no naturally occurring ERV has been demonstrably shown to be replication competent in humans. In addition, the relative abundance of the first two classes has evolved radically differently in the two species. Thus, while Class I sequences are sevenfold more abundant in humans than in mice, Class II sequences are tenfold more abundant in mice than in humans [19,21]. In humans, the Class I repertoire is dominated by HERV-H, which by itself encompasses around one-third of all ERVs in the human genome and is the most abundant human ERV family [85]. This family integrated in several waves during primate evolution and is now extinct. By contrast, Class II elements are highly abundant in mice and all integrated following the mouse–human speciation event [86]. Notable examples of this class are HERV-K (HML2) in humans, and IAP and MusD/ETn elements in mice. HERV-K is a superfamily of elements that have invaded the human genome in various waves at different time points, with HERV-K (HML2) being the most recent and potentially still active [87]. In mice, the MusD/ETn families and the IAP superfamily of elements are both active representatives of Class II and share a common feature in that both groups lack an *env* gene. Likewise, ETn elements are an exacerbated example of this trend and, because of their complete lack of coding potential, they rely on MusD elements for their retrotransposition [88].

Class III elements comprise the ERV-L family of elements and their internally deleted nonautonomous counterparts MaLR, which rely on the machinery of the former for replication. Class III elements have been detected in all eutherian mammals [89], but not in monotremes [35] or metatherians [89–91], and were probably present in their last common ancestor [89]. Despite their ancient origins, ERV-L elements have remained active throughout eutherian evolution and have undergone several independent waves of expansion in both primates and rodents [89,92,93]. Similarly, MaLR elements have also been active throughout mammalian evolution, having reached up to 3.6% and 4.8% of the mouse and human genomes, respectively, and becoming some of the most successful retrotransposon families therein.

Lastly, in striking contrast to the situation observed in mice, humans, and all other mammalian species analyzed, ERVs are almost nonexistent in the platypus genome, where they amount to only 0.1% of its genome [35] (see [Figure 2](#) in the main text). Thus, the specific abundance of the different endogenous retrovirus families differs on a species-by-species basis and reflects the history of exposure of the host to different exogenous retrovirus families, the proficiency of these families to replicate within the germline, and the selective forces acting over these integrations.

LINEs and SINEs

LINEs comprise an additional order within retrotransposons, and greatly differ from LTR elements in structure and transposition mechanism. Full-length elements resemble an integrated mRNA and harbor an internal promoter on their 5′ end, a variable number of ORFs, and a polyA tail on their 3′ end. LINE elements can be monocistronic or polycistronic, and exhibit differing arrangements of their RT, endonuclease, and accessory domains depending on the clade in question. LINE-encoded proteins exhibit a certain degree of *cis*-preference [43] and catalyze the preferential transposition of their transcript of origin [44]. In the absence of an integrase domain, LINE elements can replicate by reverse transcribing directly into the host genome, using the DNA of the host as a primer. However, the LINE replication machinery can also act *in trans* and often catalyzes the transposition of host transcripts and additional nonautonomous elements. Despite their extraordinary diversity, most eutherian mammals display a remarkably consistent LINE landscape dominated by the L1 clade ([Box 3](#)). A typical eutherian genome harbors a L1 content of 10–30% in addition to a L2 contribution hovering in the low single digits.

SINEs are an additional order of nonautonomous retrotransposons that rely on the replication machinery of LINEs for their transposition. SINEs have composite structures and incorporate a 5′ sequence (head) of variable origin, a central sequence (body) with diverse origins, and a 3′ sequence (tail) often related to the 3′ end of a LINE. Depending on the origin of their 5′ sequence, SINEs can be classified into five superfamilies [15]: (i) SINE1 display a head derived from the 7SL RNA; (ii) SINE2 have a tRNA-derived head; (iii) SINE3 exhibit a head derived from the 5S rRNA; (iv) the recently described SINEU carry a snRNA-derived head [45]; and (v) the recently proposed SINE28 are derived from 28S rRNA [46]. However, despite originating from a common set of sequences, SINE families have repeatedly arisen independently during

Box 3. LINE Diversity across Mammals

Based on reverse transcriptase phylogenies, currently known LINE elements can be classified into 32 clades [15]. In mammals, four of these clades constitute the bulk of the LINE repertoire, with the more basal lineages harboring increased LINE diversity (see Figure 2 in the main text). In the platypus genome, to our knowledge, the only monotreme genome to be systematically analyzed, L2 is the dominant LINE clade, although small amounts of the RTE and CR1 clades are also present [35]. L2 and the L2-dependent SINE MIR have both proliferated to great lengths within this genome and each comprise around 20% of its sequence. Intriguingly, while L1 is almost nonexistent in the platypus and echidna genomes [35,94], this clade has become the dominant LINE clade in metatherians [95]. In the metatherian opossum and wallaby genomes, L1 contributes 15–20% of the genomic sequence, while the monotreme-dominant L2 is limited to a mere 5–7%. A relatively large number of CR1 and RTE elements is also present in these genomes, constituting 2–3% of their respective sequences. Together with Ruminantia and Afrotheria [96], Metatheria is one of the few mammalian lineages harboring a substantial amount of RTE LINEs and also harbors RTE-dependent SINEs [95]. Thus, metatherian genomes are exceptional among mammals because of the diversity of LINE clades that they harbor, and stand in stark contrast to monotreme genomes, which are solely populated by the L2 clade.

Despite their extraordinary diversity, most eutherian mammals display a remarkably consistent LINE landscape dominated by the L1 clade. A typical eutherian genome harbors a L1 content of 10–30% in addition to a L2 contribution hovering in the low single digits. While the monotreme-dominant L2 clade is now extinct in eutherian genomes [97], the L1 clade is generally still active and has evolved numerous lineage-specific attributes, such as a recently described novel ORF in primates [98]. Ruminant and Afrotherian genomes [96], such as those of cows and sheep and those of elephants and tenrecs, represent the only known exceptions to L1 hegemony in eutherians, because up to half of their LINE content corresponds to the RTE clade [99]. Another anomalous situation has been reported in bats of the suborder Megachiroptera [100], and rodents of the taxa Sigmodontinae [101] and *Spermophilus* [102], where L1 transposition appears to have ceased, along with that of its associated SINEs. Overall, eutherian genomes are characterized by a LINE landscape comprising a single, nonbranching lineage of successively dominant L1 families [103], although some taxa contain additional clades and some others have lost LINE activity altogether.

mammalian evolution and are often lineage specific [47]. Having arisen only once in an ancestor of rodents and primates (Figure 2), SINE1 elements constitute 10% of the human genome in the form of the primate Alu family, and 2.7% of the mouse genome in the form of the rodent B1 family [48]. Among the numerous lineage-specific SINE2 families found in mammalian genomes, the MIR family stands out for its wide distribution among mammals. Amounting to more than 2% of the human and mouse genomes, this family is still active in monotremes [35] but became extinct in metatherians and eutherians following the demise of L2 (Box 3) [19]. While most eutherian SINEs rely on L1 for their transposition, RTE-dependent SINEs (e.g., AfroSINEs [49]) are also present in the genomes where the RTE clade is active. Finally, SINE3 elements have only been described in megabats [50] (Figure 2) and springhare [51], suggesting their emergence is a rare event. Thus, eutherian SINEs are a diverse group of sequences that have repeatedly emerged *de novo* over mammalian evolution and, therefore, are generally restricted to a particular set of taxa.

While, to the best of our knowledge, no functional roles have been ascribed to SINEs during early mammalian development, timely activation of L1 transcription was recently reported to be essential for mouse development [10]. Below, we discuss the implications that LINE transcription poses for the regulation of mammalian preimplantation development.

Co-option of Transposable Elements by Host Regulatory Programs

Transposable elements are thought to have been co-opted into numerous regulatory roles in mouse and human cells. Much of our current understanding of how transposable elements regulate gene expression builds upon early concepts of gene regulation [3,5], which have only recently been tackled thanks to the growth of massively parallel DNA-sequencing techniques, which enable a full overview of the genome and the functional features within transposable elements themselves. Faulker and Carninci note that an important distinction must be made when referring to transposed elements versus transposable elements [52], given that most transposable elements are truncated and unable to transpose. Examples of how transposable

element-derived sequences impact gene expression are multiple, and operate at the following levels: (i) by creating new or alternative promoters [53]; (ii) by providing or disrupting binding sites for regulatory factors [54–56]; (iii) by facilitating silencing of nearby genes [57]; (iv) through the action of noncoding RNAs [58]; (v) by generation of antisense transcripts, typically downstream of gene promoters [53]; (vi) by generating dsRNA, a mechanism recently described in mouse [59] and *Drosophila* oocytes [60]; (vii) by regulating chromatin accessibility [10]; and (viii) by contributing RNA regulatory sequences within introns and UTRs [53,61].

Whenever they are co-opted into providing functional promoters, transposable elements typically confer tissue-specific regulation as opposed to housekeeping activity. In agreement with their role as tissue-specific regulators, transposable elements initiate transcription from sharp promoters [53] (i.e., from a single, well-defined site), which is a typical feature of single-copy gene promoters that display a high degree of tissue specificity [62]. Depending on the cell type analyzed, between 6% and 30% of transcripts initiate within repetitive elements [53]. Notably, tissues from embryonic origin and transformed tissues tend to contain the highest proportion of transposable element-derived sequences in their transcriptomes. In the placenta, specific ERVs of the RLTR13D5 family have been suggested to act as species-specific enhancers in mouse and rat [63]. In addition, pioneering work by Knowles and colleagues [41] demonstrated that ~12% of mouse oocyte transcripts contain transposable element-derived sequences, in particular MaLR LTRs of the MT family, and that MERV-L LTRs often function as alternative promoters of genes activated during embryonic genome activation (EGA).

However, host factors can also be exploited by transposable elements and, therefore, co-option is a two-sided process. Indeed, transposable elements themselves are also subject to selective pressures, among which the most important is probably the ability to transpose within the germline. Transposable elements have been found to carry DNA-binding motifs for numerous transcription factors [11–13,42,54,56], and it is possible that some transcription factors exert a positive effect on the replicative fitness of transposable elements. When bound by a host transcription factor, transposable elements might achieve higher levels of expression and, consequently, increase their chances for transposition. The recent description of the DUX transcription factor binding to, and driving the expression of, ERV-L elements [11–13] is an excellent example of this phenomenon, although a link to actual retrotransposition has not yet been established. All in all, the integration of transposable elements into host regulatory programs probably reflects a complex interplay between the co-option of transposon-derived sequences (e.g., as binding sites) by host factors and the selective pressure over transposable elements themselves to promote their own replication through the acquisition of DNA-binding motifs.

Transposable Element Expression Coincides with Highest Cellular Plasticity

Regardless of the evolutionary forces that have enabled transposable elements to acquire regulatory functions, transposable elements must be able to efficiently replicate within the germline for co-option and transposon expansion within a host genome. In mammals, the germline emerges soon after implantation, upon specification of the primordial germ cells (PGCs). PGC specification is accompanied by their segregation from the epiblast and, in the mouse, this process occurs ~7.5 days after fertilization. Therefore, the developmental period before PGC specification provides an ideal setting for transposons to replicate. Indeed, in both mouse [7,41] and human [8] preimplantation embryos, transposable elements have been shown to exhibit high levels of transcriptional activity. Preimplantation development provides an excellent window of opportunity for the transcriptional activation of transposable elements

for several reasons. In mice, the remodeling of heterochromatic marks [64], the relaxed chromatin structure [65,66], and the temporary uncoupling of chromatin assembly and DNA synthesis after fertilization [67] are now well established, and may promote the robust transcriptional activation of transposable elements that occurs after fertilization, for example by favoring an open chromatin environment [7]. Transcriptional activation of transposable elements is observed as early as the zygote in the mouse and persists throughout preimplantation development (Figure 3, Key Figure). Slightly different kinetics have been documented for other species [8], yet transcriptional activation of transposons is a common theme. During this period of development, the zygote undergoes a series of cleavages, which lead to the first cell fate decision upon emergence of the blastocyst, which segregates the pluripotent inner cell mass and the trophectoderm. The latter is an epithelium of restricted plasticity, which will form the placenta, while the former will give rise to the embryo proper. While there are only a few cell divisions during this period, the chromatin and epigenomic landscape are heavily remodeled [64,68]. This remodeling is thought to mediate, at least in part, the transition from the totipotent blastomeres to the pluripotent cells of the inner cell mass.

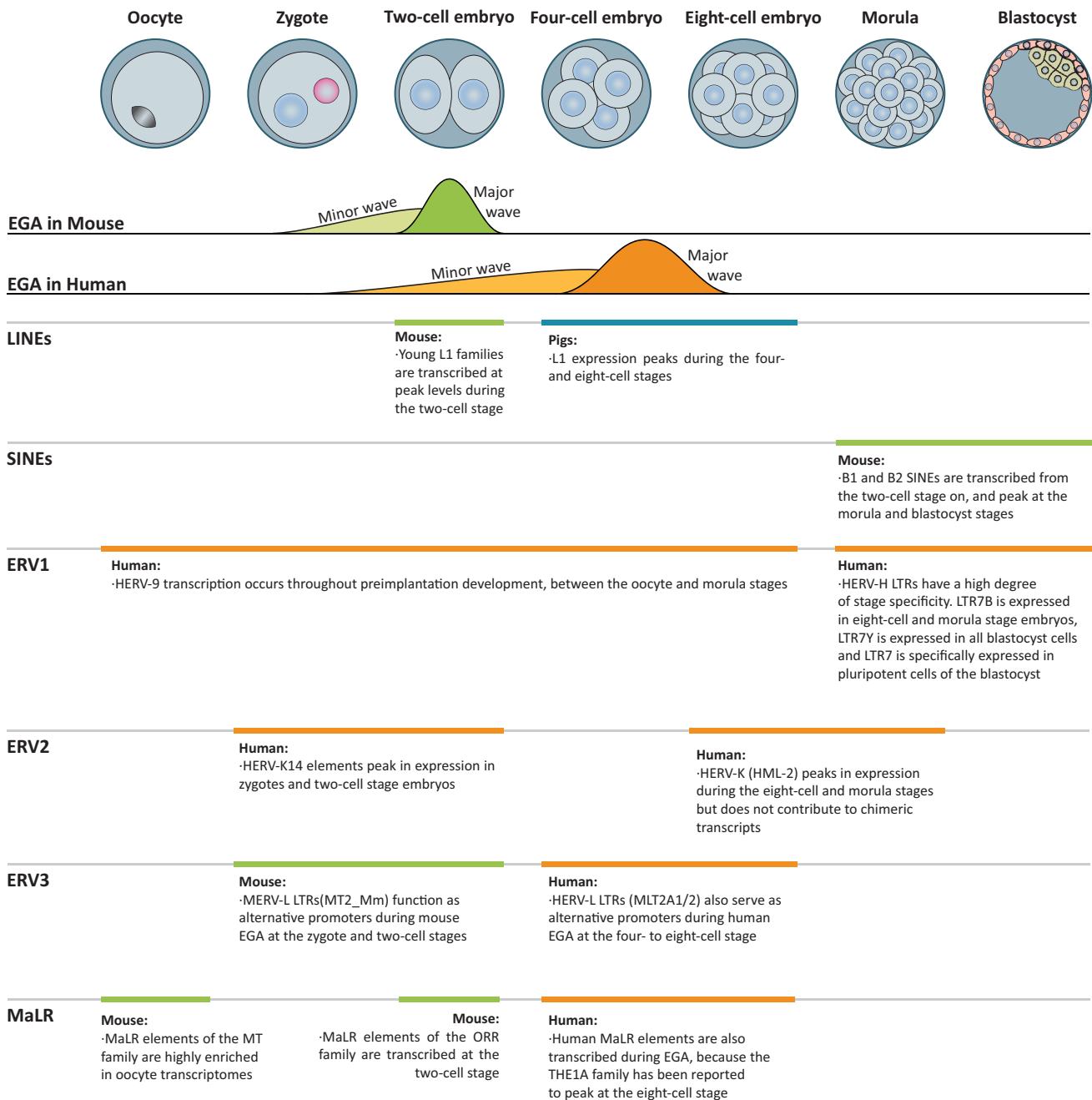
LINE Control of Chromatin Accessibility in Early Development

It is unknown whether the unique chromatin features observed in preimplantation embryos are a consequence of the transcriptional activation of transposable elements, the other way around, or a combination. For example, young L1 elements are transcribed as early as the one-cell zygote in the mouse (Figure 3). Using a DNA-sequence-targeting strategy to manipulate L1 expression *in vivo*, recent work [10] showed that preventing L1 silencing after the 2-cell stage or, conversely, enforcing premature silencing of L1 at the zygote stage, significantly decreased developmental progression. L1 elements appear to regulate global chromatin accessibility during preimplantation development, suggesting that transposable elements also contribute to shaping chromatin architecture during these stages. This is likely to be the result of a general effect of L1, since the authors found no evidence for the regulation of nearby genes after transcriptional manipulation of L1 in preimplantation embryos [10]. In fact, genes that are activated early during EGA in the mouse are generally depleted of L1 [69]. Given the abundance of L1 elements, one could envision that they have a general role in maintaining chromatin structure, in line with previous reports indicating that repetitive, Cot-1 RNA associates with chromatin and can affect compaction and chromatin organization in cells in culture [70]. In agreement, a role for SINEs in the regulation of chromatin organization has been previously documented, whereby various SINE families have driven the spread of CTCF-binding sites across the mouse, dog, and opossum genomes [55]. While in-depth analyses of transposable element distribution in the context of chromatin conformation capture data sets have not yet been documented during preimplantation development, we anticipate that transposable elements may play a key role in the organization of higher order chromosomal structures during the initial establishment of nuclear architecture in the early embryo.

As we previously discussed (Box 3), L1 is no longer active in some rodents and bats and has been gradually superseded by other LINE clades in species such as cows (Figure 2). The extent to which the co-option of L1 elements observed in mice is conserved, particularly in species with highly divergent LINE content or activity, and whether the other three LINE clades observed in mammals have been co-opted into similar roles, are both outstanding questions. In porcine embryos [71], L1 elements are also actively transcribed and their transcription peaks at the eight-cell stage, coincident with EGA (Figure 3). Interestingly, antisense transcription of L1 at this stage prevents retrotransposition, providing a mechanism whereby, despite the robust upregulation of L1 during this developmental window, the retrotransposition events are rare. Thus, while there is evidence

Key Figure

The High Potency Habitats of Transposable Elements



Trends in Genetics

Figure 3. Known windows of transposable element expression for mice, humans, and pigs are shown. Abbreviations: EGA, embryonic genome activation; ERV, endogenous retrovirus; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element.

that L1 elements in species with similar LINE content to mouse have similar expression dynamics, other species with more divergent LINE landscapes remain unexplored.

In addition, LINE elements might also be shaping the regulatory landscape of early development through indirect mechanisms. In humans, L1 has been shown to be responsible for the retrotransposition of host transcripts and the formation of more than 8000 processed, intronless retrogenes [72]. Several retrogenes are transcribed during preimplantation development, including those encoding some important regulatory factors, such as the DUX transcription factor [73], which is in turn believed to regulate transcription of murine ERV-L. We speculate that L1-dependent retrogene formation has shaped the regulatory landscape of early development and perhaps contributed to some of the uncoupling of splicing from transcription that is observed at the earliest stages of genome activation [74].

Gene Transcription Initiates within LTRs during EGA in Primates and Rodents

In mice, another interesting example is the MERV-L family of transposable elements and their MT2 LTRs, which are actively transcribed in two-cell embryos [41]. During EGA, these LTRs serve as powerful promoters enabling the transcription of neighboring genes, resulting in 'chimeric' LTR-host transcripts. In fact, a significant proportion of genes activated during mouse EGA are located in proximity to an MT2 LTR, which contains a binding motif for the pioneer transcription factor DUX [11–13]. Ectopic expression of *Dux* in mouse ESCs results in the transcriptional activation of MERV-L LTRs, as well as of a substantial proportion of the murine EGA transcriptional program. Indeed, it appears that *Dux* is expressed at the early stages of EGA in the mouse, similarly to its human ortholog *DUX4* [11,12]. In analogy to its mouse ortholog, ectopic expression of *DUX4* in human induced pluripotent stem cells (iPSCs) activates a subset of the human EGA transcriptional program, as well as HERV-L. Nonetheless, human *DUX4* fails to activate murine MERV-L LTRs when overexpressed in mouse cells, reflecting the partial divergence of the DNA-binding motif of the two DUX proteins. This is an extraordinary example of strong functional conservation in the absence of sequence conservation, because both transcription factor orthologs bind the homologous transposon families despite their limited sequence conservation. In addition, it is a nice illustration of LTRs containing species-specific transcription factor-binding sites. DUX is probably not the exception, and we anticipate that additional factors that can drive developmental programs with the help of transposable elements may exist.

Indeed, Class III ERVs, such as ERV-L and MaLR, appear to significantly contribute to both oocyte and embryo transcriptomes in both mice and humans [41]. In the mouse zygote, downregulation of MERV-L through RNA interference can result in developmental arrest at the two-cell stage [75]. However, it is unclear whether this phenotype is a consequence of the downregulation of the transcript itself, of the protein products, or of the chimeric transcripts that initiate within MERV-L LTRs. Determining whether MERV-L itself is required for totipotency has yet to be addressed. In addition, a systematic characterization of ERV-L co-option beyond mice and humans is lacking and it remains unclear whether ERV-L elements have undergone similar co-options in other eutherian lineages.

Expression of transposable elements appears to be a widespread feature of mammalian early development (Figure 3) [8]. Human preimplantation development is characterized by sequential waves of transcriptional changes, each of them with distinctive signatures of transposable element transcription [8]. Namely, HERV-K (HML-2), one of the youngest LTR insertions in the human genome, undergoes DNA hypomethylation, becomes activated during human EGA,

and remains expressed at least until the blastocyst stage [76]. In this case, however, there is no evidence of chimeric transcripts arising from LTR5HS and instead retroviral ORFs were observed to regulate host mRNA translation and/or stability of host mRNAs in the cytoplasm.

In addition to the increasing evidence documenting roles of transposable elements during development *in vivo*, they also have a clear role in promoting cellular plasticity or differentiation of stem cells *in vitro*. Approximately 231 HERV-H LTRs are expressed in human ESCs, where they function as enhancers and affect the expression of long noncoding (lncRNAs) and neighboring protein-coding genes [9]. Depletion of HERV-H not only reduced the efficiency of iPSC reprogramming, but also induced early differentiation markers, suggesting that they are important for maintaining cellular potency. In contrast to the work on L1 in mice discussed above, where no evidence of regulation of expression of neighboring genes was observed in embryos, HERV-H may regulate stem cell identity through different mechanisms. Approximately one-third of HERV-H LTRs affect nearby lncRNAs, while the remaining two-thirds affect protein-coding genes. In addition, HERV-H transcripts themselves may directly interact with chromatin modifiers to potentially recruit them to enhancers. Thus, transposable elements can regulate cellular plasticity through a variety of mechanisms.

Chromatin Regulation of Transposable Elements during Early Mammalian Development

Despite the wealth of data on transposable element transcription during human and mouse preimplantation development, relatively little is known about the epigenetic and/or chromatin-mediated regulation of these elements during this period. There have been many important contributions to our understanding of the chromatin regulation of transposons in ESCs, but *in vivo* studies in embryos have only recently started to emerge. While transposable elements are typically repressed in somatic cells by H3K9me3 and/or DNA methylation, preimplantation mouse embryos appear to resort to alternative mechanisms to regulate transposon expression.

In mouse and humans, the parental genomes undergo major DNA demethylation following fertilization [77]. Different families of LTR elements exhibit different levels of DNA methylation during preimplantation development. While ERV-L and MaLR elements are mostly hypomethylated after fertilization in both species, ERV2 elements generally maintain high levels of DNA methylation [78]. By contrast, ERV1 elements are mostly hypomethylated in humans, but exhibit intermediate levels of methylation in mouse embryos. LINEs are highly methylated in human sperm but become demethylated following fertilization, particularly in the case of young L1 families [77]. Finally, while SINEs exhibit intermediate levels of DNA methylation in human sperm and also become hypomethylated upon fertilization [77], their DNA-methylation dynamics are mostly comparable to intergenic sequences.

Using ChIP-qPCR, one of the first characterizations of histone modifications over transposable elements in preimplantation embryos [69] revealed little changes in the levels of H3K9me3 on L1 and IAP between the two-cell and eight-cell stages. Instead, a stark decrease in the levels of H3K4me3 during the same developmental period was observed. Thus, it was suggested that downregulation of L1 after the four-cell stage is associated with the reduction of H3K4me3, rather than with the acquisition of heterochromatic marks. In agreement, maternal depletion of *Lsd1* resulted in increased L1 transcription [79]. More recently, ChIP-sequencing approaches largely confirmed these observations and significantly expanded our knowledge of the distribution of H3K27me3 [80,81], H3K4me3 [80,82], and H3K9me3 [83] across transposons in the early embryo. For example, they revealed that the number of LTR loci marked by H3K9me3 gradually increases between the zygote and blastocyst stages. Notably, at least some MERV-L

elements lack H3K9me3 in oocytes and zygotes [83], but acquire H3K9me3 at the four-cell stage concomitant with their transcriptional repression. In addition, some LTR families appear to gain H3K9me3 concomitant with their loss of DNA methylation after fertilization, in what has been proposed to be an epigenetic switch of transposon control for a reduced group of LTR families [83].

However, globally, the dynamics of transcriptional activation and repression of transposons during preimplantation development cannot be solely explained by their DNA methylation levels or H3K9me3 enrichment. Indeed, a specific class of LTRs (IAPz) are regulated by the arginine methyltransferase PRMT5 in the embryo [84]. Overall, little is known about the molecular players that regulate transposable element expression during preimplantation development, and much work remains to be done to dissect the molecular mechanisms underlying their regulation by specific chromatin pathways.

Concluding Remarks: Towards a Systematic Understanding of Transposable Element Activity in Early Development Across Mammals

Given the broad co-option of transposable elements into regulatory functions in mice and humans, we can expect that similar and/or additional functional roles are widespread across mammals. While the human and mouse models have proven exceedingly useful in characterizing the contribution of transposable elements to the regulation of early development, it is important to note the many parallels in repetitive element content that both species have and that might inadvertently paint a false picture of conservation. Indeed, characterization of additional mammalian species with more divergent repetitive landscapes is necessary to fully appreciate the repertoire of regulatory roles and the evolutionary impact of transposable elements (See Outstanding Questions). For example, while DNA transposons have long been extinct in both primates and rodents, they have dramatically expanded in bats in the form of numerous families across at least two distinct transposon orders. It will be interesting to address whether DNA transposons are active during early development in this lineage and whether they have been co-opted for similar purposes as their retrotransposon counterparts in mice and humans. Given the limited conservation of transposable element content across species, we anticipate that further characterization of early embryonic transcriptomes across mammals will uncover conserved regulatory mechanisms driven by divergent repetitive landscapes. Transcriptomic analyses capable of distinguishing transposable element-driven transcription from read-through transcription during early development are particularly lacking, regardless of the species, and will certainly contribute to elucidating the evolutionary and regulatory roles of transposable elements during preimplantation development.

References

- Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603
- Orgel, L. and Crick, F. (1980) Selfish DNA: the ultimate parasite. *Nature* 284, 604–607
- McClintock, B. (1956) Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216
- Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. *Science* 165, 349–357
- Davidson, E.H. and Britten, R.J. (1979) Regulation of gene expression: possible role of repetitive sequences. *Science* 204, 1052–1059
- Finnegan, D.J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5, 103–107
- Fadloun, A. *et al.* (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* 20, 332–338
- Göke, J. *et al.* (2015) Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 16, 135–141
- Lu, X. *et al.* (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* 21, 423–425
- Jachowicz, J.W. *et al.* (2017) LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* 49, 1502–1510
- Hendrickson, P.G. *et al.* (2017) Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934
- De Iaco, A. *et al.* (2017) DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945

Outstanding Questions

How can the highly divergent repetitive landscapes present in mammalian genomes regulate such a conserved developmental process as preimplantation development?

Is transposable element transcription a conserved feature of early development across mammals? Is it a fundamental property of early development across Metazoa?

Is the transcriptional regulation of specific transposable element families by transcription factors a widespread property of these elements?

Does the presence of transcription factor-binding sites in transposable elements promote their expansion within the host or do host factors evolve an affinity for these motifs to exploit these integrations?

Are the unique chromatin features observed in preimplantation embryos a consequence of the transcriptional activation of transposable elements or vice versa?

13. Whiddon, J.L. *et al.* (2017) Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* 49, 935–940
14. Wicker, T. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982
15. Bao, W. *et al.* (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6, 11
16. Ros, F. and Kunze, R. (2001) Regulation of activator/dissociation transposition by replication and DNA methylation. *Genetics* 157, 1723–1733
17. Engels, W.R. *et al.* (1990) High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* 62, 515–525
18. Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368
19. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
20. Pace, J.K. and Feschotte, C. (2007) The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* 17, 422–432
21. Waterston, R.H. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
22. Lindblad-Toh, K. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819
23. Huang, C.R.L. *et al.* (2012) Active transposition in genomes. *Annu. Rev. Genet.* 46, 651–675
24. Ray, D.A. *et al.* (2008) Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* 18, 717–728
25. Mitra, R. *et al.* (2013) Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc. Natl. Acad. Sci. U. S. A.* 110, 234–239
26. Chalopin, D. *et al.* (2015) Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* 7, 567–580
27. Kojima, K.K. (2018) Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob. DNA* 9, 2
28. Kapitonov, V.V. and Jurka, J. (2006) Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4540–4545
29. Pritham, E.J. *et al.* (2007) Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390, 3–17
30. Pritham, E.J. and Feschotte, C. (2007) Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1895–1900
31. Thomas, J. *et al.* (2014) Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. *Genome Biol. Evol.* 6, 2595–2610
32. Mager, D.L. and Stoye, J.P. (2015) Mammalian endogenous retroviruses. *Microbiol. Spectr.* 3, MDNA3-0009-2014
33. Eickbush, T.H. and Jamburuthugoda, V.K. (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* 134, 221–234
34. Cook, J.M. *et al.* (2000) Systematic screening of *Anopheles* mosquito genomes yields evidence for a major clade of Pao-like retrotransposons. *Insect Mol. Biol.* 9, 109–117
35. Warren, W.C. *et al.* (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453, 175–183
36. Xiong, Y. and Eickbush, T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362
37. Hayward, A. (2017) Origin of the retroviruses: when, where, and how? *Curr. Opin. Virol.* 25, 23–27
38. Johnson, W.E. (2015) Endogenous retroviruses in the genomics era. *Annu. Rev. Virol.* 2, 135–159
39. Chong, A.Y. *et al.* (2014) Evolution and gene capture in ancient endogenous retroviruses - insights from the crocodylian genomes. *Retrovirology* 11, 71
40. Vargiu, L. *et al.* (2016) Classification and characterization of human endogenous retroviruses mosaic forms are common. *Retrovirology* 13, 7
41. Peaston, A.E. *et al.* (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606
42. Kunarso, G. *et al.* (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634
43. Wei, W. *et al.* (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439
44. Eickbush, T.H. and Malik, H.S. (2002) Origins and evolution of retrotransposons. In *Mobile DNA II* (Craig *et al.*, eds), pp. 1111–1144, American Society of Microbiology
45. Kojima, K.K. (2015) A new class of SINEs with snRNA gene-derived heads. *Genome Biol. Evol.* 7, 1702–1712
46. Longo, M.S. *et al.* (2015) Identification of a recently active mammalian SINE derived from ribosomal RNA. *Genome Biol. Evol.* 7, 775–788
47. Kramerov, D.A. and Vassetzky, N.S. (2011) Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 107, 487–495
48. Garcia-Perez, J.L. *et al.* (2015) The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol. Spectr.* 3, MDNA3-0061-2014
49. Gilbert, C. *et al.* (2008) Target site analysis of RTE1_{LA} and its AfroSINE partner in the elephant genome. *Gene* 425, 1–8
50. Gogolevsky, K.P. *et al.* (2009) 5S rRNA-derived and tRNA-derived SINEs in fruit bats. *Genomics* 93, 494–500
51. Gogolevsky, K.P. *et al.* (2008) Bov-B-mobilized SINEs in vertebrate genomes. *Gene* 407, 75–85
52. Faulkner, G.J. and Carninci, P. (2009) Altruistic functions for selfish DNA. *Cell Cycle* 8, 2895–2900
53. Faulkner, G.J. *et al.* (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571
54. Bourque, G. *et al.* (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762
55. Schmidt, D. *et al.* (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348
56. Imbeault, M. *et al.* (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543, 550–554
57. Morgan, H.D. *et al.* (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318
58. Espinoza, C.A. *et al.* (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat. Struct. Mol. Biol.* 11, 822–829
59. Watanabe, T. *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543
60. Brennecke, J. *et al.* (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103
61. Gong, C. and Maquat, L.E. (2011) LncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470, 284–290
62. Carninci, P. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626–635
63. Chuong, E.B. *et al.* (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* 45, 325–329
64. Burton, A. and Torres-Padilla, M.-E. (2014) Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 722–734
65. Bošković, A. *et al.* (2014) Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* 28, 1042–1047

66. Wu, J. *et al.* (2016) The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* 534, 652–657
67. Ishiuchi, T. *et al.* (2015) Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* 22, 662–671
68. Eckersley-Maslin, M.A. *et al.* (2018) Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* 19, 436–450
69. Fadloun, A. *et al.* (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* 20, 332–338
70. Hall, L.L. *et al.* (2014) Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* 156, 907–919
71. Zhang, H. *et al.* (2017) Identification and characterization of L1-specific endo-siRNAs essential for early embryonic development in pig. *Oncotarget* 8, 23167–23176
72. Zhang, Z. *et al.* (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13, 2541–2558
73. Leidenroth, A. *et al.* (2012) Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma* 121, 489–497
74. Abe, K. *et al.* (2015) The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing. *EMBO J.* 34, 1523–1537
75. Huang, Y. *et al.* (2017) Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *eLife* 6, e22345
76. Grow, E.J. *et al.* (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221–246
77. Smith, Z.D. *et al.* (2014) DNA methylation dynamics of the human preimplantation embryo. *Nature* 511, 611–615
78. Wang, L. *et al.* (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell* 157, 979–991
79. Ancelin, K. *et al.* (2016) Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. *eLife* 5, e08851
80. Liu, X. *et al.* (2016) Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* 537, 558–562
81. Zheng, H. *et al.* (2016) Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Mol. Cell* 63, 1066–1079
82. Zhang, B. *et al.* (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* 537, 553–557
83. Wang, C. *et al.* (2018) Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* 20, 620–631
84. Kim, S. *et al.* (2014) PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol. Cell* 56, 564–579
85. Bannert, N. and Kurth, R. (2006) The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.* 7, 149–173
86. Stocking, C. and Kozak, C.A. (2008) Murine endogenous retroviruses. *Cell. Mol. Life Sci.* 65, 3383–3398
87. Wildschutte, J.H. *et al.* (2016) Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci. U. S. A.* 113, E2326–E2334
88. Baust, C. *et al.* (2003) Structure and expression of mobile ETnII retroelements and their coding-competent MusD relatives in the mouse. *J. Virol.* 77, 11448–11458
89. B nit, L. *et al.* (1999) ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J. Virol.* 73, 3301–3308
90. Gallus, S. *et al.* (2015) Evolutionary histories of transposable elements in the genome of the largest living marsupial carnivore, the Tasmanian devil. *Mol. Biol. Evol.* 32, 1268–1283
91. Zhuo, X. *et al.* (2013) Genome-wide characterization of endogenous retroviruses in the bat *Myotis lucifugus* reveals recent and diverse infections. *J. Virol.* 87, 8493–8501
92. Costas, J. (2003) Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L. *J. Mol. Evol.* 56, 181–186
93. Blanco-Melo, D. *et al.* (2018) Reconstruction of a replication-competent ancestral murine endogenous retrovirus-L. *Retrovirology* 15, 34
94. Kordi , D. *et al.* (2006) Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. *Syst. Biol.* 55, 886–901
95. Gentles, A.J. *et al.* (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* 17, 992–1004
96. Walsh, A.M. *et al.* (2013) Widespread horizontal transfer of retrotransposons. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1012–1016
97. Sotero-Caio, C.G. *et al.* (2017) Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* 9, 161–177
98. Denli, A.M. *et al.* (2015) Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* 163, 583–593
99. Adelson, D.L. *et al.* (2009) Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12855–12860
100. Cantrell, M.A. *et al.* (2008) Loss of LINE-1 activity in the megabats. *Genetics* 178, 393–404
101. Grahn, R.A. *et al.* (2005) Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. *Cytogenet. Genome Res.* 110, 407–415
102. Platt, R.N. and Ray, D.A. (2012) A non-LTR retroelement extinction in *Spermophilus tridecemlineatus*. *Gene* 500, 47–53
103. Furano, A.V. *et al.* (2004) L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* 20, 9–14
104. Tarver, J.E. *et al.* (2016) The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol.* 8, 330–344

Annex II – Curriculum vitae

Diego Rodriguez-Terrones | Curriculum vitae

Date of birth: November 12, 1992, in Mexico City, Mexico (age 26) · Nationality: Mexican
Marchioninstr. 25, 81377, Munich, Germany · +49 176 5770 5021 · diego.rodriguez@helmholtz-muenchen.de
[Google Scholar](#) · [Research Gate](#) · [LinkedIn](#) · [Twitter](#)

Professional and educational history

Education

PhD: March 2016 – March 2019

Degree: PhD in Biology. Fakultät für Biologie, Ludwig-Maximilians-Universität München.

Supervisors:

Main supervisor: Prof. Dr. Maria Elena Torres-Padilla, Helmholtz Zentrum München

Co-supervisor: Dr. Juan Manuel Vaquerizas, Max Planck Institute of Molecular Biomedicine

Thesis title: *On the molecular basis of mammalian totipotency.*

Undergraduate studies: August 2011 – March 2016

Degree: Licenciado en Ciencias Genómicas con Mención Honorífica (equivalent to a Bachelor in Genomics and Bioinformatics with Honors). Center for Genomic Science of the National University of Mexico.

Supervisor:

Dr. Maria Elena Torres-Padilla, Institut de Génétique et de Biologie Moléculaire et Cellulaire

Thesis title: *Single cell expression profiling of Embryonic Stem Cells during reprogramming to the 2-cell-like state identifies key intermediate and limiting steps of the transition.*

Research activity

March 2016 – March 2019. PhD student in the lab of Maria-Elena Torres-Padilla. Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, Munich, Germany.

September 2014 – December 2015. Undergraduate student in the lab of Maria-Elena Torres-Padilla. Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg, France.

January 2012 – August 2014. Undergraduate student in the lab of Isabel López Lara. Center for Genomic Science of the National University of Mexico, Cuernavaca, México.

Publications

First author peer-reviewed publications

1. **Rodriguez-Terrones D.*** & Torres-Padilla M.E. (2018). Nimble and ready to mingle: transposon outbursts of early development. *Trends in Genetics*, 10: 806-820. ([Link](#))
2. **Rodriguez-Terrones D.***, Gaume X.*, Ishiuchi T., Weiss A., Kopp A., Penning A., Brino L., Kruse K., Vaquerizas J.M. & Torres-Padilla M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nature Genetics*, 50: 106-119. ([Link](#))

Co-authorships in peer-reviewed publications

1. Burton A*., **Rodriguez-Terrones D.**, Brochard V., Galan C., Eid A., Beaujean N., Jenuwein T. & Torres-Padilla M.E. Heterochromatin remodelling is essential for reprogramming at fertilization. (**Under review**)
2. Eid A.*, **Rodriguez-Terrones D.**, Burton A., & Torres-Padilla M.E. (2016). SUV4-20 activity in the pre-implantation mouse embryo controls timely replication. *Genes and Development*, 30: 2513-2526. ([Link](#))
3. Ishiuchi T.*, Enriquez-Gasca R., Mizutani E., Boškovic A., Ziegler-Birling C., **Rodriguez-Terrones D.**, Wakayama T., Vaquerizas J. M. & Torres-Padilla M.E. (2015). Early embryonic-like cells are induced by down-regulating replication-dependent chromatin assembly. *Nature Structural & Molecular Biology*, 22: 662-71. ([Link](#))

*First authors

Fellowships and awards

Fellowships

September 2014 – August 2015. International Mobility fellowship from the National University of Mexico to perform a research internship in Maria Elena Torres-Padilla's lab at the Institut de Génétique et de Biologie Moléculaire et Cellulaire in Strasbourg, France.

Scientific meetings and conferences

Conferences and invited oral presentations

November 19, 2018 @ Cornell University, Ithaca, USA. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

October 30, 2018 @ NYU Langone Health Center, New York City, USA. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

July 10, 2018 @ Broad Center for Regenerative and Stem Cell Research, Los Angeles, USA. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

April 20, 2018 @ LCG-UNAM Symposium, Paris, France. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

January 24, 2018 @ Helmholtz Zentrum München, Munich, Germany. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

August 1, 2017 @ Osaka University, Osaka, Japan. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

July 21, 2017 @ Max Planck Institute for Molecular Biomedicine, Münster, Germany. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

May 9, 2017 @ Helmholtz Zentrum München, Munich, Germany. **Presented talk:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

July 21, 2016 @ BioMedizinisches Zentrum, Munich, Germany. **Presented talk:** *Modelling of single cell expression data identifies molecular and epigenetic roadmaps during the emergence of totipotent-like cells.*

March 12, 2015 @ Max Planck Institute for Molecular Biomedicine, Münster, Germany. **Presented talk:** *Modeling single-cell expression data to understand cell fate decisions in early development.*

February 13, 2015 @ Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg, France. **Presented talk:** *Modeling single-cell expression data to understand cell fate decisions in early development.*

Poster presentations at scientific conferences

November 1 – 4, 2018 @ Transposable elements. Cold Spring Harbor Laboratories, USA. **Presented poster:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

April 23 – 26, 2017 @ Awakening of the genome: The maternal to zygotic transition. Dresden, Germany. **Presented poster:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

January 7-9, 2014 @ Commemorative Meeting of the 10th Anniversary of the Undergraduate Program in Genomic Sciences. Cuernavaca, México. **Presented poster:** *A biotechnological atmosphere generation system based on nitrogen fixation.*

October 3-7, 2013 @ Third Mexican Meeting on Biochemistry and Molecular Biology of Bacteria. Cuatro Ciénegas, México. **Presented poster:** *SMc02490 is an acyltransferase involved in Sulfolipid biosynthesis.*

Attended scientific conferences

January 15 – 17, 2018 @ Epigene2sys. Munich, Germany. **Attended.**

April 05 – 07, 2017 @ Abcam meeting Chromatin and Epigenetics Meeting: From Mechanism to Function. Munich, Germany. **Attended.**

October 14-15, 2014 @ Abcam meeting Chromatin and Epigenetics: From Omics to Single Cells. Strasbourg, France. **Attended.**

Courses and workshops

November 5 – 17, 2018 @ Advance sequencing technologies and their applications. Cold Spring Harbor Laboratories, USA.

March 8 – 15, 2017 @ 13th course on epigenetics: Chromatin dynamics and epigenetics. Institute Curie, Paris, France. **Presented poster:** *A molecular roadmap for the emergence of early-embryonic-like cells in culture.*

August 8 – 14, 2016 @ EMBO lecture course: Chromatin and the environment. Spetses, Greece. **Presented poster:** *Modelling of single cell expression data identifies molecular and epigenetic roadmaps during the emergence of totipotent-like cells.*

Public outreach activities

Newspaper communications and articles

1. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2014, February 1). ¿Cómo buscar lo desconocido? Curiosity y la búsqueda de vida en Marte. La Unión de Morelos, pp. 28.
2. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2014, January 25). Pequeños polizones: ¿Quién viaja con nosotros al espacio? La Unión de Morelos, pp. 28.
3. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2014, January 18). Meteoritos condriticos: fotografías del momento del origen de la vida en la Tierra. La Unión de Morelos, pp. 28.
4. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2013, November 9). Humanos, calamares, microbios y viajes espaciales. La Unión de Morelos, pp. 28.
5. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2013, November 2). Astrobiología y Plantas: Mecanismos de crecimiento de raíces. La Unión de Morelos, pp. 28.
6. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2013, October 12). Ambientes extremos: una herramienta útil para la búsqueda de vida extraterrestre. La Unión de Morelos, pp. 28.
7. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2013, October 5). El hierro y el ARN en la Tierra primitiva. La Unión de Morelos, pp. 28.
8. **Rodríguez-Terrones D.** and Escalona Meléndez J. (2013, September 21). El código genético, ¿astrobiológicamente universal? La Unión de Morelos, pp. 28.

Core competencies

Experimental experience

Molecular biology: general DNA molecular biology, preparation of single cell transcriptomic libraries and microfluidics-based single cell gene expression profiling

Biochemistry: western blotting, luciferase assays, lipid radioactive labeling and lipid thin layer chromatography.

- Cell culture:** mouse embryonic fibroblast and embryonic stem cell cultures and Escherichia coli, Agrobacterium tumefaciens and Sinorhizobium meliloti bacterial cultures.
- Flow cytometry:** cell sorting in FACS Aria II and III. Formal training and qualification obtained.
- Microscopy:** confocal microscopy experience in Leica SP5 and SP8, Bruker Opterra II and Yokogawa spinning disk microscopes. Immunofluorescence experiments in cell lines. Live cell microscopy with mouse embryonic stem cells.

Computational experience

- Programming languages:** Bash, C, Perl, R, Python, Matlab, MySQL, PHP.
- Research-related applications:** ImageJ, FlowJo and a range of epigenomics-focused bioinformatic skills. In particular, RNA-seq, ChIP-seq and ATAC-seq bioinformatics analyses.
- Graphic design:** Photoshop, Illustrator.

Teaching experience

2014. Teaching assistant of an undergraduate Epigenetics course (4 hours per week, 20 undergraduate students). National University of Mexico.

2013. Initiator and organizer of an Astrobiology and Origin of Life seminar series for undergraduate students (2 hours per week, 30 undergraduate students). National University of Mexico.

Referees

Prof. Dr. Maria Elena Torres-Padilla
(Main PhD supervisor)

Director of the Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München
+49 89 3187 3317
torres-padilla@helmholtz-muenchen.de

Dr. Juan Manuel Vaquerizas
(PhD co-supervisor)

Group leader at the Max Planck Institute of Molecular Biomedicine
+49 251 70365 580
jmv@mpi-muenster.mpg.de

Dr. Esperanza Martinez Romero
(Undergraduate program director)

Director of the Undergraduate Program in Genomic Science, National University of Mexico
+52 777 313 1697
emartine@ccg.unam.mx

Annex III – Copyright statements

**SPRINGER NATURE**

Title: Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly

Author: Takashi Ishiuchi, Rocio Enriquez-Gasca, Eiji Mizutani, Ana Bošković, Celine Ziegler-Birling et al.

Publication: Nature Structural & Molecular Biology

Publisher: Springer Nature

Date: Aug 3, 2015

Copyright © 2015, Springer Nature

LOGIN

If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select no to the question "Are you the Author of this Springer Nature content?".

Ownership of copyright in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:

To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

The author and any academic institution, where they work, at the time may reproduce the contribution for the purpose of course teaching.

To reuse figures or tables created by the Author and contained in the Contribution in oral presentations and other works created by them.

To post a copy of the contribution as accepted for publication after peer review (in locked Word processing file, of a PDF version thereof) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the contribution on the publisher's website.

Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read Springer Nature's online [author reuse guidelines](#).

For full paper portion: Authors of original research papers published by Springer Nature are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v1.0

[BACK](#)[CLOSE WINDOW](#)

**SPRINGER NATURE**

Title: A molecular roadmap for the emergence of early-embryonic-like cells in culture
Author: Diego Rodriguez-Terrones et al
Publication: Nature Genetics
Publisher: Springer Nature
Date: Dec 18, 2017
Copyright © 2017, Springer Nature

LOGIN
If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select no to the question "Are you the Author of this Springer Nature content?".

Ownership of copyright in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:

To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

The author and any academic institution, where they work, at the time may reproduce the contribution for the purpose of course teaching.

To reuse figures or tables created by the Author and contained in the Contribution in oral presentations and other works created by them.

To post a copy of the contribution as accepted for publication after peer review (in locked Word processing file, of a PDF version thereof) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the contribution on the publisher's website.

Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read Springer Nature's online [author reuse guidelines](#).

For full paper portion: Authors of original research papers published by Springer Nature are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v1.0

[BACK](#)[CLOSE WINDOW](#)

1. All articles in *Genes & Development* are accessible online free of charge six months from the full-issue publication date, except for articles that carry the journal's Open Access icon, which are made freely accessible online upon publication in return for a fee paid by their authors.
2. Authors of articles published in *Genes & Development* retain copyright on their articles (except for US Government employees) but grant Cold Spring Harbor Laboratory Press exclusive right to publish the articles. This grant of rights lasts for six months following full-issue publication for all non-Open Access articles and includes the rights to publish, reproduce, distribute, display, and store the article in all formats; to translate the article into other languages; to create adaptations, summaries, extracts, or derivations of the article; and to license others to do any or all of the above.
3. Authors of articles published in *Genes & Development* can reuse their articles in their work as long as *Genes & Development* is credited as the place of original publication. They can also archive the Cold Spring Harbor Laboratory Press PDF version of their article with their institution, immediately on publication if it is an Open Access article and 6 months after publication if it is a non-Open Access article.
4. Beginning six months from the full-issue publication date, articles published in *Genes & Development* that are not designated as Open Access are distributed under the Creative Commons Attribution-Non-Commercial 4.0 International License (CC-BY-NC), as described at <http://creativecommons.org/licenses/by-nc/4.0/>. This license permits non-commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited. Articles that carry the Open Access designation are immediately distributed under one of two Creative Commons Licenses (based on author selection and in response to funding agencies' policies): (a) CC-BY-NC (<http://creativecommons.org/licenses/by-nc/4.0/>) or (b) Creative Commons Attribution 4.0 International License (CC-BY) (<http://creativecommons.org/licenses/by/4.0/>). The CC-BY license permits commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited.
5. Cold Spring Harbor Laboratory Press will deposit articles in PubMed Central where they will be released to the public six months following the full-issue publication date (with the exception of Open Access papers, which are made freely available in PubMed Central immediately upon full-issue publication).
6. Preprint servers: Conference presentations or posting un-refereed manuscripts on community preprint servers will not be considered prior publication. Authors are responsible for updating the archived preprint with the journal reference (including DOI), and a link to the published article on the *Genes & Development* website upon publication. Submission to the journal implies that another journal or book is not currently considering the paper. Submitted manuscripts are subject to press embargo.

Warranties

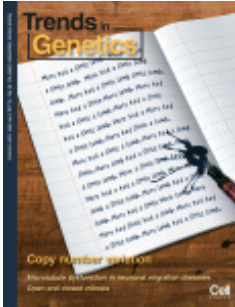
This publication is provided "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

In no event shall Cold Spring Harbor Laboratory Press be liable for any claim for damages including but not limited to any special, incidental, indirect, consequential damages, damages resulting from loss of use, data, or profits or any damages which are claimed to arise from or be in connection with the use or performance of any information published in *Genes & Development*.

Descriptions of, or references to, products or publications do not imply endorsement of those products or publications by *Genes & Development* or Cold Spring Harbor Laboratory Press.

Genes & Development is under continuing development and changes may be made to this Notice at any time.





Title: Nimble and Ready to Mingle:
Transposon Outbursts of Early
Development

Author: Diego Rodriguez-
Terrones, Maria-Elena Torres-
Padilla

Publication: Trends in Genetics

Publisher: Elsevier

Date: October 2018

© 2018 Elsevier Ltd. All rights reserved.

LOGIN

If you're a **copyright.com**
user, you can login to
RightsLink using your
copyright.com credentials.
Already a **RightsLink user** or
want to [learn more?](#)

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

Copyright © 2018 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#)
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Acknowledgements

This thesis is dedicated to my family and in particular to my mother Maria Eugenia, without whose persistent support I most certainly would not be concluding a PhD at this time. I'm particularly grateful for allowing me to borrow her laid-back attitude and stoic sense of humor, both of which have proven to be of enormous convenience while enduring the peculiar vicissitudes of the scientific endeavor. I would also like to dedicate this thesis to my father Ariel, and my brothers Emilio and Santiago, and in particular I would like to thank them for the constant influx of interdisciplinary viewpoints that they provide. Be it rock profiles, reductionist approaches to geopolitics, perspectives on the left as seen from Manhattan or behavioral reports on the modern Mexican bureaucrat, each and every one of their practical and theoretical – although mostly theoretical – expertise are highly regarded.

I would also like to thank everyone both in Munich and in Strasbourg who contributed one way or another to the work I conducted during my PhD and my undergraduate work in the lab. In particular, I would like to express my gratitude to both Takashi and Xavier for the privilege of working with them; it has truly been an honor to learn from such a structured and disciplined pair of scientists. I also want to thank Götz Hartleben, for his tremendous help with the metabolic work and for the valuable insights that fueled this collaboration. I would also like to thank Camille, Kilian, Mayuko, Valeri and Irene for their invaluable help when collecting the 700+ pre-implantation embryos used for the generation of the transcriptional profiles described in this thesis. This project would most certainly not have been possible without them (please consult Celine Ziegler or Camille Noll for photographic evidence of this claim) and I'm really grateful for their tremendous efforts. Fortunately, not all cell sorting relies on mouth pipettes and someone somewhere invented high-throughput FACS machines. I would like to thank Dagmar Pich, Claudine Ebel and Wolfgang Hammerschmidt for providing access to these wondrous machines, and for their help with their troubleshooting. Also in terms of infrastructure, I would like to thank Andreas Ettinger, Elisabeth Graf, Thomas Schwarzmayr and Christelle Thibault for providing access to microscopy, DNA sequencing and single cell isolation instruments over the past years, and for their help with their operation.

Of course, the biggest and most special acknowledgement must undoubtedly go to my supervisor, Maria Elena, who has somehow managed to bear with me for the past four years (no

easy feat!). I'm really, really grateful for all her patience and support over these past years, as well as for the countless extraordinary opportunities she has provided. Be it by encouraging attendance to renowned courses or conferences, by securing access to fantastic scientific resources, by patiently permitting me to explore eccentric side-projects, or even by simply providing her input on how to best arrange panels in a manuscript's figure, she has always manifested her genuine interest in supporting my training and my career and I will be forever grateful for that. I would also like to thank Juanma Vaquerizas, who has provided a wealth of valuable input and advice over the past few years, and without whose initial guidance (not unexpectedly, in a bar) my career would probably have taken a different direction. Lastly, I would also like to thank the members of my thesis advisory committee, namely Didier Trono and Wolfgang Enard, for their valuable and enthusiastic input throughout my PhD. It has certainly been a singular privilege to be supervised – one way or another – by these four brilliant scientists.

Lastly, I would also like to thank all of the past and current members of the Torres-Padilla lab for all of the great moments we've shared over the past 4 years. In particular, I'd like to thank: Ane, Andre and Camille for their fantastic sense of humor, memorable coffee breaks and unforgettable photoshops (of fellow lab mates, that is); Adam, for never letting us stick to all those 'one drink' initiatives; Xavier, for the impeccable teamwork and permanent good mood; Takashi, for his critical thinking and for teaching me the importance of robust foundations in any project; Mate and Joanna, for more than one highly-memorable improvised trip; Tsune and Ken, for being really nice senseis; Manuel, Natasha and Marion, for their fun albeit short time at the lab; Dr. Emily, for being the nicest possible person and always a joy to have around; and Jess, Poonam and Lara for all being extraordinarily fun in their own particular way.