De novo heterochromatin establishment in 2C-like cells and early embryonic development

Ruben Sebastian Perez

TESI DOCTORAL UPF 2021

DIRECTOR DE LA TESI

Dr. Maria Pia Cosma

DEPARTMENT

Gene Regulation, Stem Cells and Cancer
Centre for Genomic Regulation (CRG), Barcelona

Department of Experimental and Health Sciences
Universitat Pompeu Fabra

“Las cosas podían haber sucedido de cualquier otra manera y, sin embargo, sucedieron así”.

El camino. Miguel Delibes

ACKNOWLEDGEMENTS

This Doctoral Thesis represents my hard work and commitment over the last four years in the laboratory. However, it is obvious that this Thesis would have never been possible without the unconditional support of my family. Then, the first big THANK YOU goes to them for all their affection, comprehension and intelligent advice!

I would like to continue by thanking my PhD supervisor, Pia Cosma, for giving me the possibility to develop professionally during my time in her laboratory. It has been an enriching experience since it has allowed me to grow independently as researcher. This journey has also taught me the basis of many fundamental aspects related to the research environment, some of which I will surely exploit in the next steps of my professional career within the life sciences arena.

To develop challenging projects one should always try to be surrounded by motivated peers. Along these lines, I would also like to thank Luciano Di Croce and Sergi Aranda for being more than excellent collaborators. You have helped both the project and myself at different levels, scientifically and personally, so I would always be grateful for your suggestions and contributions. Next, I would like to thank the group of people from the lab that have been involved in one way or another with the final product of my PhD project. Thank you Shoma, Martina, Umberto, Pablo, Marc, Jose Luis and Davide for your

help with some experiments and/or analysis, and for your contributions to the scientific discussions.

I am profoundly grateful to the members of my Thesis Committee. Thank you Bernhard Payer, Verena Ruprecht and Manuel Serrano for your involvement in my PhD project, for challenging my hypothesis with rigorous, founded comments during our great discussions, and, overall, for positively influencing the quality of my research year after year.

Many of the experiments presented in this Thesis would have hardly been achieved without the technical support of the CRG Core Facilities. Thus, I would like to extend my appreciation to the Proteomics Unit, in particular, to Eva Borràs and Eduard Sabidó, who provided great advise on the analysis of our chromatome dataset; to all the members of the Flow Cytometry Unit for their kindness and willingness to always provide assistance; to the Advanced Light Microscopy Unit; and to the Histology Unit.

Finally, I would like to thank my colleagues from the Technology and Business Development Office (TBDO), specially Anabel Sanz and Silvia Tórtola. Thank you for welcoming me as TBDO fellow during several months and for generously providing me with the possibility of joining the team. I have been exposed to a complete new world that has profoundly reshaped the way I currently conceive biomedical research.

A mis abuelos

ABSTRACT

Chromocenters are established after the 2-cell (2C) stage during mouse embryonic development, but the factors that mediate chromocenter formation remain largely unknown. To identify regulators of 2C heterochromatin establishment, we generated an inducible system to convert embryonic stem cells (ESCs) to 2C-like cells, thus modeling early embryogenesis *in vitro*. This conversion is marked by a global reorganization of H3K9me3-heterochromatin foci, which are then reversibly formed upon re-entry into pluripotency. Profiling the chromatin-bound proteome of ESCs transitioning to 2C-like cells, we uncover chromatin regulators involved in *de novo* heterochromatin formation and a relationship between cell cycle regulation and the establishment of the 2C-like state. We identified SMARCAD1, which associates with H3K9me3-heterochromatin in ESCs, but its nuclear localization is lost in 2C-like cells. SMARCAD1 depletion in mouse embryos leads to developmental arrest and loss of H3K9me3. Collectively, our findings contribute to comprehending the establishment of chromocenters during early development, a key step to instruct the embryonic totipotent program toward pluripotency.

RESUMEN DE LA TESIS

Los cromocentros se forman después del estadio de 2 células (2C) durante el desarrollo embrionario de ratón, pero los factores que median su formación siguen siendo en gran medida desconocidos. Para identificar los reguladores del establecimiento de la heterocromatina 2C, generamos un sistema inducible capaz de convertir células madre embrionarias (ESCs) en células similares a embriones 2C (células 2C-like), modelando así el desarrollo temprano *in vitro*. Esta conversión está caracterizada por una reorganización global de las regiones heterocromáticas marcadas por H3K9me3, las cuales se forman de nuevo al volver a la pluripotencia. Describimos el proteoma unido a la cromatina de ESCs reprogramándose a células 2C-like donde descubrimos reguladores de la cromatina implicados en la formación *de novo* de heterocromatina así como una relación entre la regulación del ciclo celular y el establecimiento del estado 2C-like. Identificamos el factor SMARCAD1, el cual se asocia con la heterocromatina marcada por H3K9me3 en ESCs, pero su localización nuclear se pierde en las células 2C-like. Finalmente, eliminamos SMARCAD1 en embriones de ratón observando una detención en su desarrollo y una pérdida de H3K9me3. Estos hallazgos contribuyen a comprender el establecimiento de los cromocentros durante el desarrollo temprano, un paso crucial para instruir el programa totipotente embrionario hacia la pluripotencia.

PREFACE

The unicellular totipotent zygote represents the beginning of a sophisticated journey from a molecular standpoint that ends up forming a complex and self-sustained new entity. During the first hours of this developmental adventure when the zygote starts the transformation toward an organized and instructed mass of cells called the blastocyst, the 1-cell embryo experiments a number of tightly regulated and interconnected phenomena that shapes its developmental fate. The focus of the current Thesis is on the mechanisms underlying the establishment of some nuclear structures, the chromocenters, and the factors that mediate their reorganization in the early embryo. In brief, chromocenters assemble a portion of the mammalian genome that should remain transcriptionally silent in normal homeostatic conditions, and whose aberrant expression might generate, in most of the cases, detrimental effects for the individual.

In the work presented in this Thesis, we have exploited the 2C-like cell reprogramming system, which induces early embryo-like cells reassembling the 2-cell stage embryo from ESCs, to investigate the *de novo* formation of the heterochromatic chromocenters. By combining the 2C reprogramming system with chromatin proteomics, advanced imaging and mouse embryo manipulations, among other complementary technologies, we have identified that SMARCAD1, a

nucleosome remodeler, contributes to the establishment of chromocenters in early mammalian embryos.

Collectively, our findings shed light on how repressive nuclear structures are formed *de novo* during developmental transitions, a key step to instruct the embryonic totipotent program toward pluripotency. In particular, we have captured an additional layer of information and complexity in the 2C-like system, the chromatin bound proteome, which helped us discover SMARCAD1, and which we envision would contribute to move the field forward. We hope that this study also exemplifies the necessity of generating and exploring *in vitro* systems that mimic embryonic development to advance in the comprehension of fundamental developmental processes and in the identification of novel regulators.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

2C	2-cell
2C-like	2-cell-like
2i	Two inhibitors
3C	Chromosome confirmation capture
3D	Three-dimensional
APH	Aphidicolin
ATAC-seq	Assay for transposase accessible chromatin followed by sequencing
ATR	Ataxia-telangiectasia mutated and Rad3-related
ATRX	Alpha-thalassemia/mental retardation X-linked syndrome protein
Bp	Base pairs
C/EBPα	CCAAT/enhancer-binding protein alpha
CA	Codon altered
CAF-1	Chromatin assembly factor 1
Cas9	CRISPR associated protein 9
CBX	Chromobox proteins
CDX2	Caudal-type homeodomain transcription factor
CENPA	Centromere protein A
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CPT	Camptothecin
CRISPR	Clustered regularly interspaced short palindromic repeats
DAXX	Death-domain associated protein
DDR	DNA damage response
DEG	Differentially expressed gene
Dm-ChP	DNA-mediated chromatin pull-down
DNMT1	DNA methyltransferase 1
Dox	Doxycycline
DPPA2	Developmental pluripotency associated factor 2
DPPA4	Developmental pluripotency associated factor 4
E	Embryonic day

EB	Embryoid body
EdC	5-ethynyl-2’-deoxycytidine
EdU	5-ethynyl-2’-deoxyuridine
Epi	Epiblast
EpiSC	Epiblast stem cell
EPSC	Extended potential stem cell
ERK1/2	Extracellular signal-regulated kinase 1/2
ERV	Endogenous retrovirus
ESC	Embryonic stem cell
FGF2	Fibroblast growth factor 2
FGF4	Fibroblast growth factor 4
FRAP	Fluorescent recovery after photobleaching
GATA3	GATA binding protein 3
GATA6	GATA binding protein 6
GO	Gene ontology
GSK3	Glycogen synthase kinase 3
H3K27me3	Trimethylation of lysine 27 on histone H3
H3K4me2	Dimethylation of lysine 4 on histone H3
H3K9ac	Acetylation of lysine 9 on histone H3
H3K9me3	Trimethylation of lysine 9 on histone H3
H3R17me	Methylation of arginine 17 on histone H3
H3R26me	Methylation of arginine 26 on histone H3
H4K20me3	Trimethylation of lysine 20 on histone H3
HDAC	Histone deacetylase
HERVL	Human endogenous retrovirus
HEX	Haematopoietically expressed homeobox
Hi-C	High-throughput chromosome confirmation capture
HIRA	Histone cell cycle regulator
Hox	Homeotic gene
HP1	Heterochromatin protein 1
HU	Hydroxyurea
IAP	Intercisternal A-particle
ICM	Inner cell mass
iPSC	Induced pluripotent stem cell

JmjC	Jumonji C
K	Lysine
KAP1	KRAB-associated protein 1
KDM1A	Lysine demethylase 1A
KDM4E	Lysine demethylase 4E
KLF2	Krüppel-like factor 2
KLF4	Krüppel-like factor 4
KO	Knock out
KOSR	Knockout serum replacement
KRAB	Krüppel-associated box
LCDM	hLIF, CHIR99021, DiM or (S)-(+)-dimethindene maleate, and MiH or minocycline hydrochloride
LIF	Leukemia inhibitory factor
LTR	Long terminal repeat
Luc	Luciferase
MajSat	Major satellites
MAPK	Mitogen-activated protein kinase
MAX	Myc-associated factor X
MEF	Mouse embryonic fibroblast
MEK	Mitogen-activated protein kinase
MERVL	Mouse endogenous retrovirus with leucine tRNA primer
miR	microRNA
MO	Morpholino antisense oligo
MS/MS	Tandem mass spectrometry
MyoD	Myogenic differentiation protein
NELFA	Maternal negative elongation factor A
NT	Non transfected
NuRD	Nucleosome remodeling and deacetylase complex
OCT4	Octamer-binding transcription factor 4
OSKM	OCT4, SOX2, KLF4 and c-MYC
pB	PiggyBac transposon system
PCA	Principal component analysis
PCC	Pearson’s correlation coefficient

PEV	Position effect variegation
PRC	Polycomb repressive complex
PrE	Primitive endoderm
PSC	Pluripotent stem cell
pSer10	Histone H3 phosphorylated at serine 10
R	Arginine
RB	Retinoblastoma
REX1	Reduced expression protein 1
RNAi	RNA interference
RRR	Reprogramming resistant region
S	Serine
SAINT	Significance analysis of interactome
SAM	Synergistic activation mediator
SCNT	Somatic cell nuclear transfer
scRNA-seq	Single-cell RNA sequencing
sgRNA	Single guide RNA
sLIF	Serum/LIF
SOX2	SRY-Box transcription factor 2
STORM	Stochastic optical reconstruction microscopy
Su(var)	Suppressors of PEV
SUMO	Small ubiquitin-like modifier
TAD	Topologically associating domain
TALE	Transcription activation-like effector
TE	Trophectoderm
TGF-β	Transforming growth factor-beta
TOPBP1	DNA topoisomerase 2-binding protein 1
TPBPA	Trophoblast-specific protein alpha
TRIM28	Tripartite motif-containing 28
TSA	Trichostatin A
T-SCE	Telomere sister chromatid exchange
Wnt	Wingless and Int-1
WT	Wild type
ZFP	Zinc-finger protein
ZGA	Zygotic genome activation
ZMYM2	Zinc finger MYM-type containing 2

Zscan4	Zinc finger and SCAN domain-containing protein 4

PART I - INTRODUCTION and AIMS

INTRODUCTION

CHAPTER 1 - Mammalian embryonic development

The life of all mammalian species starts at the moment of fertilization. Following fertilization, two highly specialized gametes, the oocyte and the sperm, fuse and combine their genetic information to create a one-cell embryo, the totipotent zygote (Leung and Zernicka-Goetz 2015) (Figure 1).

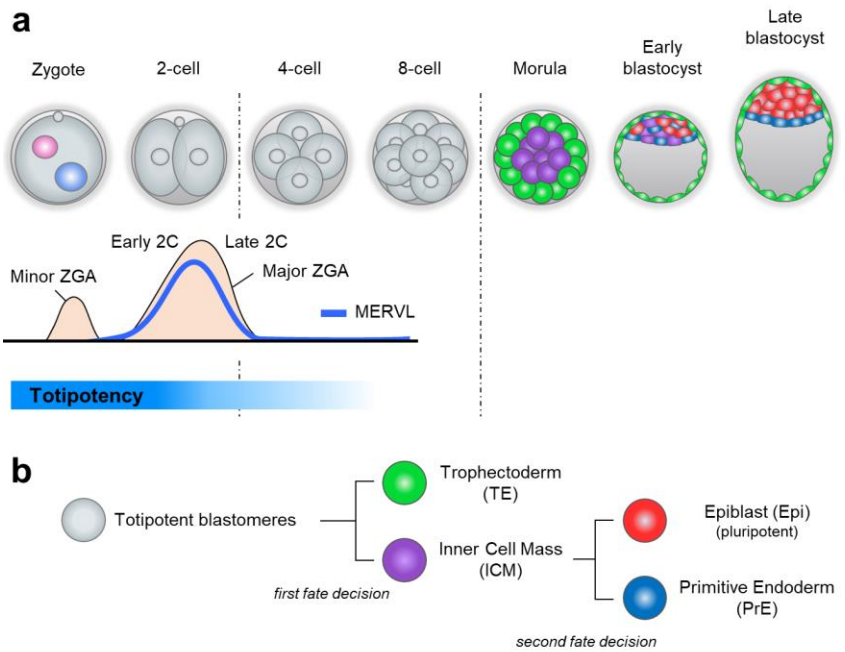

Figure 1. Mammalian embryo development. a, Overview of the mouse preimplantation development from zygote to late blastocyst. Maternal (pink) and paternal (blue) pronuclei are represented in the zygotic nucleus. Embryonic lineages are highlighted with colors. 2C, 2-cell. MERVL, mouse endogenous retrovirus with leucine tRNA primer. ZGA, zygotic genome activation. **b**, Lineage segregation in the mouse embryo is ensured by two rounds of cell fate decisions. The first fate decision specifies TE and ICM, whereas the second fate decision differentiates the ICM into Epi and PrE. Adapted from Schrode et al., 2013.

The zygote contains separate haploid paternal and maternal pronuclei until the first round of DNA replication. Right after fusion, a process denominated pronuclear syngamy come into play to coalesce both pronuclei. Shortly, the zygotic genome is activated which, in mice, occurs within the first few divisions at the 2-cell stage (Flach et al. 1982) (Figure 1). Then, the zygote is subjected to a series of cleavage divisions, which occurs in the absence of cell growth, until forming a mass of > 100 cells named the blastocyst (Figure 1).

During this period, embryos must balance accumulation of cells with fate specification and differentiation, first into the trophectoderm (TE) at the morula stage and then into the primitive endoderm (PrE) in the blastocyst (Figure 1). The blastocyst-stage embryo represents a key developmental stage since it contains pluripotent stem cells (PSCs) that constitute the epiblast (Epi) located in the inner cell mass (ICM) (Figure 1). PSCs have the potential to differentiate into the three germ layers: ectoderm, endoderm and mesoderm. These events are mainly achieved thanks to the sequence of regulatory events such as the zygotic genome activation, the clearance of maternally and paternally inherited information in the form of mRNA, proteins and epigenetic modifications and the reorganization of zygotic chromatin. Remarkably, in mice, preimplantation development, which is the unidirectional process that entails the developmental stages from fertilization to the implantation of the blastocyst embryo in the maternal uterus, lasts approximately four and a half days after

fertilization (normally indicated as embryonic days (E), i.e. E4.5) (Wennekamp et al. 2013; Eckersley-Maslin et al. 2018) (Figure 1).

Totipotency

The unicellular zygote is a totipotent entity from which the process of embryogenesis starts and, therefore, multicellular organisms originates. The term "*cellular totipotency*" was originally coined by the German plant physiologist Göttlieb Haberlandt in 1902. It was the observation that, in plants, terminally differentiated cells retain the potential to revert to an undifferentiated state and form entire new plants what provided the foundation for this concept. However, with our current knowledge, the term totipotency could be define by two criteria of increasing complexity and stringency (Ishiuchi and Torres-Padilla 2013; Wu et al. 2017; Torres-Padilla 2020).

As a first approximation, totipotency is the ability of a single cell to contribute to all cell lineages, both from embryonic and extraembryonic origin, in an organism. Based on this more permissive definition, some embryonic cells (also named blastomeres) up to the 4-cell and 8-cell stage embryos could fulfill the requirements of being totipotent (Figure 1). Transplantation experiments demonstrated that single blastomeres of these stages contribute to the development of all lineages in chimeric mice (Rossant 1976; Kelly 1977;

Balakier and Pedersen 1982). However, it was reported that not all blastomeres at the 4-cell and 8-cell stages display complete totipotency due to some bias toward specific lineages (Fujimori et al. 2003; Piotrowska-Nitsche et al. 2005; Tabansky et al. 2013). For example, work done by Nicolas Plachta and colleagues identified the existence of an intraembryonic variation in the kinetics of OCT4, appearing after the 4-cell stage. The slower kinetics of OCT4, which were postulated as tighter chromatin binding, were found to correlate with specification of the ICM (Plachta et al. 2011).

Nevertheless, the definition of totipotency can be revised and add an additional degree of stringency (Edwards and Beard 1997; Ishiuchi and Torres-Padilla 2013). We can then define totipotency as the ability of a single cell to develop into a complete organism. Thus, in the mouse embryo, totipotency is restricted to the zygote and to each of the blastomeres from the 2-cell stage embryo, as demonstrated by Andrzej K. Tarkowski in pioneering work done in the late 1950s (Tarkowski 1959; Tsunoda and McLaren 1983; Papaioannou et al. 1989) (Figure 1). Irrespectively of the criteria we apply to define totipotency, what is unequivocal is the fact that the intrinsic cellular potency of the individual blastomeres decreases as the embryo develops further (Tarkowski et al. 2010; Wu et al. 2016b). This gradual decay of cellular potency is linked to the fact that individual blastomeres embarks into cell fate decision choices (Figure 1).

Unlike the case for pluripotency, where embryonic stem cells (ESCs) can be derived and maintained in culture as we will discuss later in this Chapter, the mechanism underlying the molecular regulation of totipotency remains largely unknown. In mice, only the zygote and 2-cell stage blastomeres can generate an entire organism on their own, and are therefore regarded as truly totipotent cells as previously defined by the most stringent criteria (Tarkowski 1959). In brief, 2-cell embryos are characterized by the lack of DAPI-stained chromocenters in the nucleus (Probst et al. 2007), which are interchromosomal clusters of centromeric heterochromatin that come together in the three-dimensional (3D) nuclear space, and high chromatin mobility, which decreases with development (Boskovic et al. 2014). The transcriptional profile of 2-cell embryos is characterized by activation of major satellites, mouse endogenous retrovirus with leucine tRNA primer (MERVL), and 2-cell specific genes, such as Eif1a-like and Zscan4 (zinc finger and SCAN domain containing protein 4) genes, among many other that we will deeply cover in Chapter 2 (Figure 1).

Importantly, the laboratory of Samuel Pfaff identified in 2012 a small transient population with fluctuating expression of a particular retrotransposon, MERVL, arising in pluripotent stem cell cultures (Macfarlan et al. 2012). This rare population has a transcriptome profile that closely resembles the observed in the blastomeres of the totipotent 2-cell embryos, thus leading to be named 2-cell-like (2C-like) cells (Macfarlan et al. 2012).

The results presented in this study indicated that some features of totipotent cells can be regained occasionally in pluripotent cells. The discovery of this phenomenon provided the community with a novel way of studying certain aspects of totipotency that were extremely challenging to examine *in vivo*. Although to date, no study has reported the establishment of a totipotent cell line *in vitro*, the identification of 2C-like cells has moved us closer to solving the fundamental question in developmental biology of how totipotency is established. We will cover all the history, characteristics and advances done in the field of 2C-like cells in Chapter 2 since this is the main model used in the project presented in the current Thesis.

Zygotic genome activation

To start the developmental journey, mammalian embryos inherit a pool of maternal transcripts that play a leading role in epigenetically resetting the paternal DNA and histones across the genome of the zygote. These maternal factors are progressively degraded and replaced by the products of zygotic transcription (Zernicka-Goetz et al. 2009). In particular, it is in these early stages of mammalian development when a burst of zygotic transcription is initiated in a coordinated, multistep process referred to as zygotic genome activation (ZGA, also known as embryonic genome activation) (Figure 1). Until the 2-cell stage, embryos exhibit the highest degree of developmental plasticity and flexibility since they can switch

their fate toward one of the three cell types that will be originated during preimplantation development: the pluripotent epiblast (Epi) and the extraembryonic primitive endoderm (PrE) and trophectoderm (TE) (Zernicka-Goetz et al. 2009).

The ZGA is fully completed after two successive transcriptional bursts, the so-called minor and major phase of ZGA, which begins at the late 1-cell stage and peaks at the 2-cell stage (Flach et al. 1982; Braude et al. 1988; Schultz 1993; Bouniol et al. 1995) (Figure 1). The first or minor wave happens at low levels in the S to G2 cell cycle phase, predominantly in the male pronucleus of the zygote (Bouniol et al. 1995). Then, the major wave of ZGA occurs in the G1 to S phase of the 2-cell embryo (Ram and Schultz 1993; Bouniol et al. 1995; Aoki et al. 1997; Hamatani et al. 2004; Zeng et al. 2004; Park et al. 2013). It is noteworthy that endogenous retroviruses or repetitive elements are generally not expressed in oocytes, with a few exceptions like the intercisternal A-particle (IAP) retrotransposon, but are later expressed in a restricted time window at the onset of the ZGA (Gifford et al. 2013; Friedli and Trono 2015). For example, mouse endogenous retrovirus with leucine tRNA primer (MERVL) expression peaks in 2-cell stage embryos (Svoboda et al. 2004; Gifford et al. 2013) (Figure 1). The derepression of retroelements is in line with the fact that hyperaccessibility of chromatin by transcriptional machinery is an indispensable requisite to initiate ZGA. Chromatin accessibility is largely determined by chemical modifications located on the N-terminal tails of histone proteins (commonly

referred to as histone modifications or marks), which acts as a fundamental epigenetic regulator to control the gene expression during embryo development in mammals (epigenetics will be extensively covered in Chapter 3). In brief, there are two major types of histone modifications involved in the regulation of gene expression during the ZGA: lysine acetylation and lysine trimethylation. H4 acetylation makes pronucleus permissive for active transcription (Adenot et al. 1997). To emphasize the importance of a tight control over the chromatin status for a successful ZGA, the laboratory of Terry Magnuson identified in 2006 that the loss of the maternal BRG1, the catalytic subunit of the ATP-dependent chromatin remodeling SWI/SNF complex, results in reduced levels of zygotic genes and arrest at the 2-cell stage (Bultman et al. 2006). This study demonstrated that chromatin remodelers that induce acetylation of lysine 4 on histone H3 are required for mouse embryogenesis.

Recently, a series of groundbreaking studies reported that Dux genes play a master regulating role in ZGA in mammalian embryos, with Dux and DUX4 in mouse and human, respectively, acting upstream of earliest transcribed genes (De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017) (Figure 2). Both Dux and DUX4 encode double homeodomain transcription factors that have a tandem repeated arrangement, with Dux close to a fusion point that resembles a subtelomeric structure (Leidenroth et al. 2012; De Iaco et al. 2017) and DUX4 lying within the subtelomeric region (Clapp et

al. 2007). In terms of expression, Dux and DUX4 are either no expressed or have relatively low expression in the oocyte and appear to accumulate after fertilization (Deng et al. 2014; Hendrickson et al. 2017) (Figure 2).

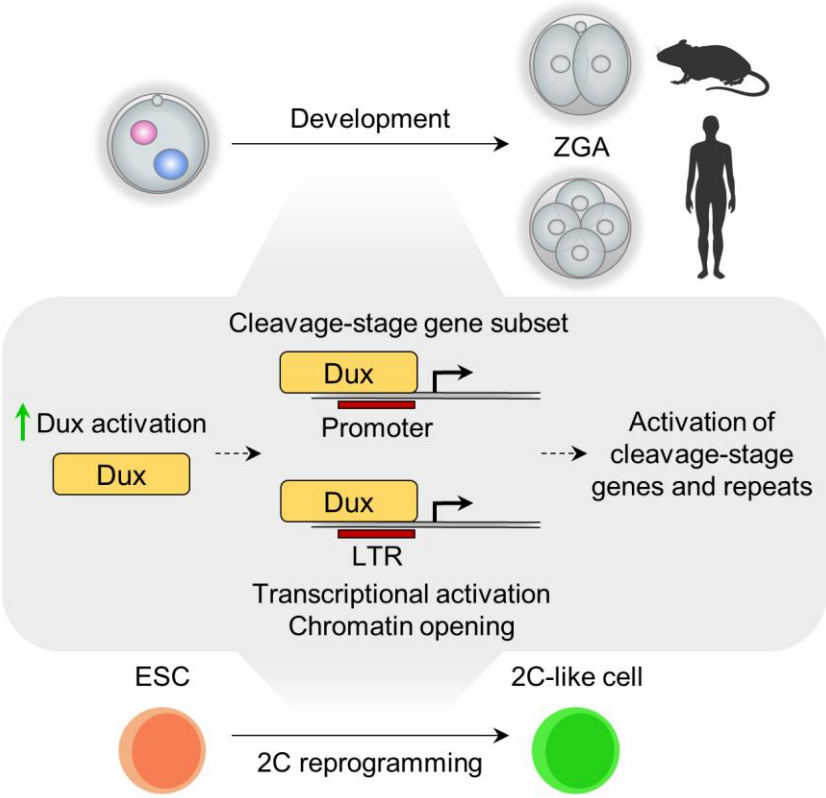

Figure 2. Function of Dux in mouse and human cells. Dux binds to promoters and long terminal repeats (LTRs) on cleavage-stage genes and repetitive elements, leading to transcriptional activation, chromatin opening and further activation of more cleavage-stage genes. In the embryo, Dux regulates the ZGA, while Dux reactivation in ESCs induces the generation of 2C-like cells. ESCs, embryonic stem cells. ZGA, zygotic genome activation. Adapted from Iturbide & Torres-Padilla, 2017.

The reported expression profile of Dux and DUX4 is in contrast to previous observations of master regulators of ZGA in other species, for example, Zelda in *Drosophila melanogaster* (Drosophila) (Liang et al. 2008; Harrison et al. 2011).

In mouse ESCs, the KRAB-associated protein 1, KAP1 (also known as TRIM28, tripartite motif-containing 28) induces repressive trimethylation of lysine 9 on histone H3 (H3K9me3) at the 5' end of Dux (De Iaco et al. 2017), indicating transcriptional silencing of Dux in the pluripotent state. Along these lines, knockout of KAP1 induces Dux expression converting ESCs to the 2C-like state (Macfarlan et al. 2012; De Iaco et al. 2017). Interestingly, in human embryos, DUX4 activates the lysine demethylase 4E (KDM4E) resulting in loss of H3K9me3 (Hendrickson et al. 2017), suggesting a positive feedback.

The program of genes activated by Dux contributes broadly to reset the epigenetic and transcriptional state of early embryos. Dux and DUX4 target long terminal repeat (LTR) promoters of endogenous retrovirus (ERVs), and Dux and DUX4 have evolutionarily diverged to interact with MERVLs and human endogenous retroviruses (HERVLs), respectively (De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017) (Figure 2). Within the cohort of genes transcribed by both Dux and DUX4 are the Zscan4 cluster and Tcstv1/3, both act together to elongate the short telomeres of mouse oocytes (Liu et al. 2007) via telomere sister chromatid exchange (T-SCE). The

Zscan4 cluster of six paralogous genes (from Zscan4a to Zscan4f) activates T-SCE (Zalzman et al. 2010). Tcstv1 and Tcstv3 contribute to telomere extension by stabilizing Zscan4 (Zhang et al. 2016). Shortly after fertilization, the activation of members of the Dux family induces a dramatic reset of gene expression state, contributing to the loss of DNA and histone methylation, resulting in a significant shift in chromatin accessibility (Hendrickson et al. 2017) (Figure 2).

Cell fate specification

Following ZGA, and to ultimately acquire a given fate, blastomeres undergo two rounds of cell fate specification that drive them toward one of the three first cell lineages (Figure 3). During the early morula stage, the first wave of cell fate specification starts, leading to the establishment of the two initial cell populations, the inner cell mass (ICM) and the trophectoderm (TE) (Schrode et al. 2013) (Figure 3). This cell fate decision between ICM and TE is thought to be orchestrated by the expression of lineage-specific transcription factors and partially resolved by the spatial and polarization queues integrated by the Hippo signaling pathway (Chazaud and Yamanaka 2016). The polar cells commonly remain at the exterior and adopt a TE cell fate, while the apolar cells will be localized inside the embryo and will become ICM (Figure 1). Indeed, perturbation of cell cycle polarity affects TE/ICM allocation (Plusa et al. 2005).

These lineage-specific transcription factors have a restricted expression pattern, and their absence impairs embryonic development. The caudal-type homeodomain transcription factor CDX2 and the GATA binding protein 3 (GATA3) are expressed in the TE (Figure 3). Remarkably, the laboratory of Janet Rossant demonstrated in 2005 that CDX2 mutant embryos fail to form a functional TE (Strumpf et al. 2005; Ralston and Rossant 2008). Additionally, Gata3 knockdown impairs the progression of morula toward the blastocyst stage (Home et al. 2009). Conversely, ICM cells specifically express several transcription factors, such as the octamer-binding transcription factor 4 (OCT4), the SRY-Box transcription factor 2 (SOX2), the homeobox protein NANOG and the GATA binding protein 6 (GATA6) (Figure 3). These ICM-specific transcription factors have been suggested not to be essential for ICM specification during the first wave but rather for the second cell lineage decision via gene inactivation studies (Messerschmidt and Kemler 2010; Frankenberg et al. 2011; Frum et al. 2013; Bessonnard et al. 2014; Schrode et al. 2014; Wicklow et al. 2014).

In brief, cells of the ICM are maintained undifferentiated and, therefore, retain pluripotency, and will continue to divide until reaching the second round of cell fate specification. On the other hand, cells on the outside develop into extraembryonic TE, which will later support the development of the embryo in the uterus and provide signaling sources crucial for axis patterning (Rodriguez et al. 2005; Soares et al. 2008; Zernicka-

Goetz et al. 2009). However, this is a very simplistic and incomplete view of a key developmental process, where more regulatory layers come into play.

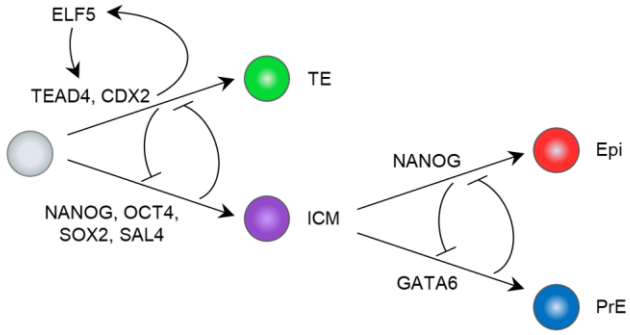

Figure 3. Transcription factors controlling cell fate decisions. ICM-specific transcription factors repress TE-specific genes and vice versa. Similar antagonistic effect is observed between the Epi and the PrE transcription program. Colored circles represent embryonic lineages: Totipotent blastomeres (grey), inner cell mass (ICM, purple), trophectoderm (TE, green), epiblast (Epi, red) and primitive endoderm (PrE, blue). Adapted from Zernicka-Goetz et al., 2009.

The output of the first cell fate decision is a combination of several factors such as (i) cell polarization generated via (a)symmetric divisions; (ii) an inside-outside location of the cells, as a result of the latter; and (iii) early epigenetic asymmetries identified at the 4-cell stage. For instance, the balance between symmetric and asymmetric divisions would lead to an unequal inheritance of the TE-specific transcription factor Cdx2 mRNA, thus predisposing the fate of the recipient daughter cell toward TE (Jedrusik et al. 2008). With respect to the epigenetic asymmetries, the laboratory of Magdalena Zernicka-Goetz discovered in 2007 that there was a differential methylation of arginine 26 and arginine 17 on histone H3

(H3R26me and H3R17me) in this early stages (Torres-Padilla et al. 2007). Along these lines, the presence of H3R26me is associated with blastomere pluripotency as the same group demonstrated in chimera experiments some years earlier (Piotrowska-Nitsche et al. 2005). Ultimately, all these aspects influence the regulation of cell fate gene expression, and vice versa, to specify cell fate.

Later in development, at the blastocyst stage, the second cell fate decision takes place (Figure 3). In this process, cells of the ICM that are in contact with the blastocyst cavity are set aside to form the second extraembryonic tissue, the primitive endoderm (PrE), whereas deeper ICM cells express pluripotency markers and form the epiblast (Epi) (Figure 3). Epiblast cells will give rise to all the somatic tissues and the germline of the proper embryo (Brook and Gardner 1997). Here, it is important to reinforce the notion that both tissues are derived from the ICM. Thus, the undifferentiated identity of the cells that will form the Epi is actively maintained by pluripotency transcription factors. Indeed, the specification of the PrE is associated with the loss of expression of several pluripotency transcription factors.

These lineages are spatially and molecularly distinct and have restricted developmental potential by the time of implantation. This lineage decision will occur without any apparent spatial order in the first instance, and then, cells of the ICM will sort into a layer of Epi and PrE by migration and apoptosis (Plusa

et al. 2008; Meilhac et al. 2009). Initially, PrE, marked by the lineage factors GATA6 and SOX17, and Epi, marked by NANOG, have an overlapping expression within the blastomeres until the 32-cell stage (Plusa et al. 2008). By the 64-cell stage, ICM cells express only Epi or PrE markers in a "*salt and pepper*" pattern, as originally identified by the laboratory of Janet Rossant in 2006 (Chazaud et al. 2006; Plusa et al. 2008) (Figure 1).

These two mixed populations are eventually reorganized, where PrE cells form a single epithelial cell layer at the surface of the blastocoel cavity (Plusa et al. 2008; Meilhac et al. 2009). The previous observation of the "*salt and pepper*" expression pattern of NANOG and GATA6 led to postulate the "*cell sorting*" model for Epi and PrE specification (Figure 1). In fact, NANOG is required for Epi specification and GATA6 is detected in all ICM cells in NANOG mutant embryos (Frankenberg et al. 2011). Also, it has been proposed that GATA6 and NANOG repress each other (Figure 3). Conversely, GATA6 is required for PrE, and, in its absence, all ICM adopt an Epi fate (Bessonnard et al. 2014; Schrode et al. 2014). The current model postulates that the pluripotency factors OCT4, SOX2, and NANOG activate fibroblast growth factor 4 (FGF4) expression, driving some cells to downregulate the activity of the pluripotency network. This triggers the specification of the PrE (Ambrosetti et al. 2000; Messerschmidt and Kemler 2010; Bessonnard et al. 2017), leading to its

maturation via a sequential activation of PrE transcription factors (Plusa et al. 2008; Artus et al. 2011).

Embryonic stem cells

The ICM contains pluripotent embryonic stem cells (ESCs) with the unique ability to generate all the cells of the adult organism and, therefore, to differentiate into the three germ layers (endoderm, ectoderm and mesoderm) as well as into primordial germ cells (Nichols et al. 1998; Hayashi et al. 2011; Wu et al. 2016b; Zhou et al. 2016) (Figure 1). They accomplish this through complex regulation of gene expression and epigenetic factors. We should always remember that ESCs are a static *in vitro* representation of a dynamic developmental system. Researchers have attempted to capture the preimplantation and post-implantation stages of embryonic development in culture, as represented by naïve and primed ESCs, respectively. Notably, the growth conditions for naïve and primed ESCs are mostly agreed upon throughout the stem cell field, although naïve exits in two distinct states as we will discuss shortly. Also, defining the naïve and primed states in mouse ESCs has been more direct than in humans because the cells can be subjected to the gold standard assay: reincorporation into a developing blastocyst and contributing to all the tissues of a new organism, including germline transmission.

Mouse ESCs were isolated and expanded from the ICM of mouse preimplantation blastocysts for the first time in 1981 (Evans and Kaufman 1981; Martin 1981). ESCs can be maintained *in vitro* under controlled culture conditions. Initial culture conditions were chosen from earlier experiments performed on embryonic carcinoma cells that originate from germ cell tumors. Indeed, ESCs were firstly co-cultured on a feeder layer of mitotically inactivated mouse embryonic fibroblasts (MEFs) in a serum-containing media, which promoted the self-renewal capacity and inhibited differentiation cues, thus providing the essential conditions for an almost infinite proliferation cycle (Evans and Kaufman 1981; Martin 1981). Those early ESCs were shown to be able to generate embryoid bodies (EBs) when grown in suspension and teratomas with tissues from all three germ layers when subcutaneously injected into immunocompromised mice (Evans and Kaufman 1981; Martin 1981). These represented the first pluripotent ESCs ever reported to grow and to be manipulated *in vitro* and set a standard for pluripotency testing.

The culturing system was later simplified since feeder cells were substituted with the cytokine leukemia inhibitory factor (LIF) (Smith et al. 1988; Williams et al. 1988; Nichols and Ying 2006), resulting in the media formulation referred to as serum/LIF (referred to as sLIF). The cytokine LIF is important for self-renewal by activating Stat3 signaling (Smith et al. 1988; Williams et al. 1988; Niwa et al. 1998).

Blocking of extracellular signal-regulated kinase 1/2 (ERK1/2) signaling and inhibition of the glycogen synthase kinase 3 (GSK3) (referred to as 2i/LIF) generated ESCs similar to the preimplantation epiblast in what is also known as the "*ground state*" of pluripotency (Silva and Smith 2008; Ying et al. 2008). Each of the components of the 2i/LIF medium was shown to be important in maintaining pluripotency in ESCs. In particular, mitogen-activated protein kinase (MEK) inhibition, which is normally achieved by PD0325901, suppresses ERK activation by FGF4 and keeps ESCs in a preimplantation state (Kunath et al. 2007). GSK3 inhibition, which is commonly achieved by CHIR99021, was found to maintain the undifferentiated phenotype by activating the Wnt (Wingless and Int-1) pathway through β-catenin-mediated transcriptional activation of Wnt genes (Sato et al. 2004; Ying et al. 2008). The addition of the two inhibitors (2i) to block key pathways made possible that ESCs could be grown in the absence of serum in the medium (Ying et al. 2008).

Both sLIF and 2i/LIF cultured ESCs are able to contribute to the ICM of developing preimplantation embryos and thus appear to have equal pluripotent potential. Cells can interconvert between the two conditions simply through media change and fully adopt the characteristics of the new growth condition (Marks et al. 2012). Cultures of ESCs grown in sLIF exhibit more heterogeneous morphology, with dense, domed colonies and more fattened colonies existing in the same culture system (Wray et al. 2010). sLIF colonies also show

heterogeneous expression of the key pluripotency factor NANOG (Wray et al. 2010), indicating that within this culture condition cells exist at different points on the pluripotency spectrum. Notably, other key pluripotency factors such as OCT4, SOX2, the Krüppel-like factor 2 (KLF2), the Krüppel-like factor 4 (KLF4), and the estrogen related receptor beta (ESRRB) are expressed at roughly the same level in both cell states (Marks et al. 2012).

Notably, transcriptomic analysis revealed that ESCs cultured in 2i/LIF are most similar to the preimplantation epiblast of the E4.5 embryo (Boroviak et al. 2015). This stage is closer to the developmental point when epiblast cells are specified and separated from the primitive endoderm (Figure 1). Thanks to these and many other advances, ESCs have been a great model to investigate the mechanisms that control and regulate pluripotency and early mouse embryo development in a dish as reflected by more than 40 years of research in the stem cell field.

In 2007, two research groups led by Ludovic Vallier and Ronald McKay derived a new set of mouse ESCs from a later stage of embryonic development. ESCs from the post-implantation epiblast were isolated and cultured in the presence of activin and fibroblast growth factor 2 (FGF2) (Brons et al. 2007; Tesar et al. 2007) (Figure 1). Preimplantation ICM cells could not grow in these conditions, but post-implantation grew well (Brons et al. 2007). The cells were termed epiblast stem cells

(EpiSCs) and had a striking resemblance to human ESCs (Brons et al. 2007; Tesar et al. 2007). EpiSCs in culture were morphologically similar to human ESCs with large flattened colonies compared with mouse ESCs that had smaller domed colonies (Brons et al. 2007; Tesar et al. 2007). EpiSCs are dependent on activin/Nodal signaling via transforming growth factor-beta (TGF-β) pathway (Brons et al. 2007; Tesar et al. 2007). EpiSCs are very inefficient at producing chimeric mice when injected in preimplantation blastocysts (Brons et al. 2007; Tesar et al. 2007), likely due to more limited potential since they were derived from later stage of development. However, EpiSCs still prove to be pluripotent cells. EpiSCs are able to generate teratomas and EBs with differentiation to all three germ layers (Brons et al. 2007; Tesar et al. 2007). They express OCT4, SOX2 and NANOG at comparable levels to ESCs (Tesar et al. 2007) but do not express the reduced expression protein 1 (REX1) (Brons et al. 2007), a naïve ESC marker.

In this Chapter, we have presented and discussed how mammalian species develops, with a particular focus on the first stages of embryo development that include the awakening of the zygotic genome and the initial cell fate decisions toward lineage specification. Importantly, we have also introduced and discussed the concept of totipotency and the derivation of ESCs since it will be relevant for Chapter 2 and for results presented in this Thesis.

CHAPTER 2 - Cell fate reversibility to model early mammalian development in culture

Cell fate and iPS cell reprogramming

In the classical view of development, the differentiated state of a cell was believed to be terminal and irreversible. As development progresses, cells become increasingly specialized to perform defined functions, a commitment that is accompanied by a restriction in the range of potential fates of those cells. This concept has been exemplified by Conrad H. Waddington's description of the epigenetic landscape in 1957.

John Gurdon's seminal paper on nuclear reprogramming of cell identity provided a remarkable challenge to this dogma, and formed the basis for today's cell-reprogramming field (Gurdon et al. 1958). In 1958, Gurdon, Elsdale and Fischberg successfully cloned a frog via transplantation of a somatic nucleus from the animal's intestine into an oocyte using a technique called somatic cell nuclear transfer (SCNT) (Gurdon et al. 1958; Gurdon 1962). Gurdon and colleagues showed for the first time that the identity of differentiated cells could be reversed and the nuclei of differentiated cells retain the capacity to orchestrate the development of a fully functioning organism. Almost 40 years after the experiments of Gurdon and colleagues, Ian Wilmut and Keith Campbell reported in 1996 the generation of the first cloned mammal, Dolly the sheep (Campbell et al. 1996), which confirmed that the

epigenetic state of mammalian somatic cells could be reprogrammed to an embryonic pluripotent state. This pioneering work formed the foundations for the reprogramming field.

Early work in reprogramming revealed the instructive role of transcription factors in lineage specification. We will highlight some studies that exemplified that it is not only possible to reset cell identity to the blank state of early embryonic development, but also to switch a cell's identity. In 1983, work done by Helen Blau showed that fusion of a differentiated mouse muscle cell with a human fetal cell derived from the amniotic fluid produce a cell with both a human and a mouse nucleus resulted in the rapid expression of human muscle-specific genes (Blau et al. 1983). Thus, this was the first evidence showing that factors produced in a differentiated cell can induce the expression of genes that are repressed in another cell type. Another remarkable study was published in 1987, when the laboratory of Harold Weintraub discovered that forced expression of a single transcription factor, myogenic differentiation protein (MyoD), can convert fibroblast cells into contracting muscle cells (Davis et al. 1987). This study provided the first evidence indicating that a single factor was capable of reprogramming cell identity. In the early 2000s, work led by Thomas Graf showed that the single overexpression of the transcription factor CCAAT/enhancer-binding protein alpha (C/EBPα) can convert committed B- and T-cell progenitors into functional macrophages with high

efficiencies, demonstrating that even fully differentiated cells can switch identities (Xie et al. 2004; Laiosa et al. 2006).

The previously mentioned studies reported that cell identity can be manipulated and, therefore, that cells retain certain degree of plasticity. However, none of the discussed approaches, with the exception of nuclear transfer, had the potential to revert the cellular identity to an embryonic state. In 2006, Kazutoshi Takahashi and Shinya Yamanaka reported in a groundbreaking study that differentiated cells could be reset to a pluripotent state through the expression of only four transcription factors (Takahashi and Yamanaka 2006). After transducing a specific cocktail of four transcription factors, namely OCT4, SOX2, KLF4 and c-MYC (commonly referred to as OSKM or the Yamanaka factors), differentiated cells can be artificially induced to revert to a pluripotent state resulting in the generation of the so-called induced pluripotent stem cells (iPSCs).

Since the establishment of ESC cultures in the 1980s (Evans and Kaufman 1981; Martin 1981) and, more recently, the discovery of iPS cell reprogramming (Takahashi and Yamanaka 2006), several studies exploiting transcription-factor-mediated reprogramming to pluripotency and direct fate conversions have allowed for extensive *in vitro* investigation of mechanisms controlling pluripotency, contributing to the unravelling of puzzling concepts such as potency, lineage-commitment and cell fate (Cohen and Melton 2011;

Vierbuchen and Wernig 2011; Morris and Daley 2013). Moreover, cell reprogramming strategies offer the potential to produce patient-derived cells for modeling diseases *in vitro* (Passier et al. 2016) and form the basis of various proposed regenerative therapies (Mandai et al. 2017; Pesaresi et al. 2019).

The emergence of 2C-like cells

Building upon the initial discoveries made in the field of cellular reprogramming, the laboratory of Samuel Pfaff discovered in 2012 that cells resembling the embryonic 2-cell (2C) stage blastomeres appeared in ESC cultures (Macfarlan et al. 2012) (Figure 4). This early embryonic-like cells were named as 2-cell-like or 2C-like cells (Macfarlan et al. 2012). This discovery brought the first cellular model with the possibility to investigate some features of the totipotent embryo, the molecular mechanisms regulating totipotency *in vitro*, and how the exit from this state is regulated (Figure 4). We should reinforce the notion that, until now, the study of totipotency has primarily been done through investigations on the mammalian embryo, which can be limiting in amount to the applications of many molecular and biochemical approaches. Therefore, the emergence of a cellular model to study totipotent-like features has created an enormous research interest in the last years.

The initial work describing a rare subpopulation of cells resembling the blastomeres of the 2-cell stage embryo was reported by Todd Macfarlan and colleagues in 2012, as mentioned before (Macfarlan et al. 2012) (Figure 4).

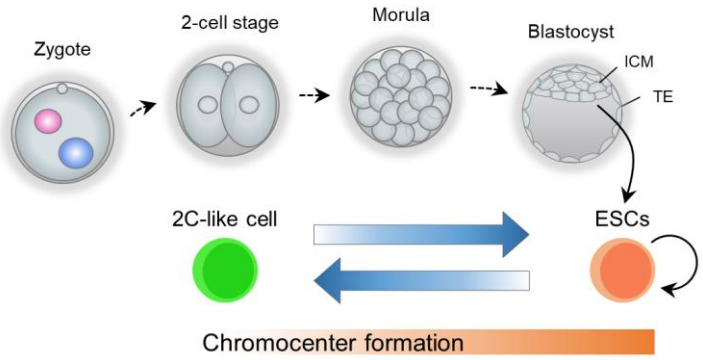

Figure 4. 2C-like cells arise from embryonic stem cell cultures. The zygote embryo develops until the blastocyst stage, prior to implantation in the maternal uterus. ESCs are derived from the ICM of blastocyst embryos and show self-renewing capacities *in vitro*. 2C-like cells have been identified to arise from ESC cultures in a transient and fluctuating manner. Chromocenters are formed *in vivo* after the 2-cell stage, making the 2C-like system suitable for their study. ESCs, embryonic stem cells. ZGA, zygotic genome activation. Adapted from Surani & Tischler, 2012.

2C-like cells were identified upon monitoring the transcriptional activity of MERVL in ESCs. The MERVL transposable elements were known to peak in expression levels at the 2-cell stage in embryos (Peaston et al. 2004) (Figure 1). Actually, these MERVL sequences act as promoters regulating gene expression of distant genes, in particular totipotency-related genes (Peaston et al. 2004; Macfarlan et al. 2012). 2C-like cells have been studied and characterized using a "*2C reporter*", which contains the LTR from MERVL fused to a

fluorescent protein, thus reflecting their transcriptional activation. 2C-like cells are present in both ESC and iPSC cultures and share several key features with the 2-cell stage embryo, which we will extensively discuss shortly in this Chapter. Importantly, these 2C-like cells were present *in vitro* in different ESC culture conditions (2i/LIF, sLIF or knockout serum replacement, KOSR), but more were observed in KOSR growth conditions, and almost all cells in culture were able to flow in and out of the 2C-like state over a period of time (Figure 4). However, although pluripotent stem cells can spontaneously revert their fate to resemble early embryogenesis, this process happens at very low frequency, accounting for less than 1-0.5 % of the cells in culture. Additionally, Macfarlan and colleagues showed that 2C-like cells have high levels of acetylated H3 and H4 histones, and an increase of active histone marks like dimethylation of lysine 4 on histone H3 (H3K4me2) (Macfarlan et al. 2012). Complementing this observation, treatment with a histone deacetylase (HDAC) inhibitor or the absence of the silencing histone modifiers lysine demethylase 1A (KDM1A), KAP1 or G9A, facilitates the emergence of 2C-cell like cells marked by MERVL (Macfarlan et al. 2012). Indeed, KAP1 depletion leads to upregulation of MERVL (Rowe et al. 2010). These findings raise the interesting possibility that 2C-like cells can be induced and/or maintained through modifying the epigenetic state and/or expression of retrotransposons.

Remarkably, a year later, in 2013, the laboratory of Joshua Brickman identified totipotent cells in ESC cultures using a reporter for the extraembryonic endodermal marker haematopoietically expressed homeobox (HEX) (Morgani et al. 2013). They found that within a ESC population, grown in sLIF or in the presence of 2i, cells were either positive or negative for the presence of HEX protein and, if isolated, the cells could regenerate a heterogeneous population of HEX-expressing cells (Morgani et al. 2013). These results suggested that indeed, this subpopulation of cells was fluctuating in and out of the particular state from pluripotent cultures, as previously demonstrated for 2C-like cells (Macfarlan et al. 2012; Morgani et al. 2013). Gene expression data also showed that HEX+ cells also expressed trophoblast-specific genes, extraembryonic markers such as Eomes, Gata3, and Tcfap2a/Tfap2a, and upregulated ERV elements. The latter was also observed in 2C-like cells (Macfarlan et al. 2012; Ishiuchi et al. 2015). In fact, HEX+ cells could differentiate *in vitro* to trophoblast and endoderm cells and, when introduced into mouse embryos, could contribute to both embryonic and extraembryonic tissues (Morgani et al. 2013). This work conducted by Sophie Morgani nicely demonstrate that transient, rare subpopulations of totipotent-like cells can be identified in pluripotent cultures when careful analysis over extraembryonic markers is performed.

More recently, in 2017, a series of groundbreaking studies reported that Dux genes play a master regulating role in ZGA

in mammalian embryos (De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017) (Figure 2). In particular, the work led by the laboratories of Bradley Cairns, Douglas Carrell and Stephen Tapscott elegantly demonstrated that reprogramming of pluripotent ESCs toward the 2C-like state can be enhanced by the single overexpression of the double homeodomain murine transcription factor Dux (Hendrickson et al. 2017) (Figure 5). The controlled expression of the cleave-stage-specific transcription factor Dux is necessary and sufficient to provide a robust and efficient *in vitro* system to induce 2C-like cells, reaching > 60 % of 2C-like cells in culture (Hendrickson et al. 2017) (Figure 5).

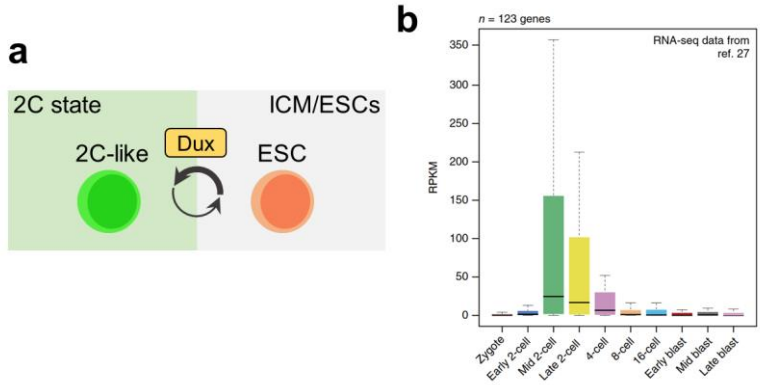

Figure 5. Dux overexpression in ESCs activates the 2C-specific transcriptional program. a, Mouse Dux expression is both necessary and sufficient to convert ESCs into 2C-like cells with high efficiency. **b**, Relative expression of Dux-induced genes in the preimplantation mouse embryo. Plot in **b** was extracted from Hendrickson et al., 2017.

This landmark study was the first in reporting the generation of "*induced*" 2C-like cells at large scale, opening the possibility of performing experiments that were previously impossible due to

the scarcity of cells. Many of the highly upregulated genes after Dux overexpression are transcripts exclusively expressed in the mid and late 2-cell stage (Hendrickson et al. 2017) (Figure 5). Similarly to the 2-cell embryo, MERVL repeats and chimeric transcripts derived from the MERVL LTR are speculated to play a role in activating the 2C transcriptome. MERVL are directly bound and induced by DUX itself. The chromatin landscape is also largely remodeled as characterized by changes in chromatin accessibility, which resemble that of the 2-cell stage embryo (Hendrickson et al. 2017). Of note, this Dux-dependent 2C-like system is the one that will be used in the current Thesis.

Interestingly, it should be mentioned that although the activity of Dux on some ZGA genes has been clearly demonstrated *in vivo* and in 2C-like cells, its role during embryonic development seems not to be essential and still is a matter of debate in the field. Several studies have provided evidences pointing toward slightly different conclusions (De Iaco et al. 2017; Chen and Zhang 2019; Guo et al. 2019), presumably due to differences in genetic background and experimental design. For example, the laboratory of Didier Trono reported that zygotic depletion of Dux leads to impaired early embryonic development and defective ZGA (De Iaco et al. 2017). On the contrary, Zhiyuan Cheng and Yi Zhang generated knockout mice for the entire Dux cluster and demonstrated that both Dux zygotic and maternal/zygotic deleted embryos survive to adulthood although with reduced developmental potential (Chen and

Zhang 2019). However, RNA-seq analysis on Dux-null embryos showed a clear transcriptional defect during ZGA, even though to a different extent depending on the study. In sum, Dux was identified as a key positive regulator of the 2C-like state. Some additional positive and negative regulators of the 2C-like state have been identified and we will discuss them in detail later in this Chapter.

The journey from pluripotency to 2C-like cells

ESCs can fluctuate back to a 2C-like state. Yet, the dynamics of the reprogramming process from pluripotency to 2C-like cells are still unclear. Given the recent advances in single cell expression technologies, such as single-cell RNA sequencing (scRNA-seq), some progress has been made in the understanding of how ESCs are reprogrammed toward 2C-like cells.

Several studies have already utilized scRNA-seq in the 2C-like reprogramming system. The laboratory of Maria-Elena Torres-Padilla uncovered an intermediate cell population expressing ZSCAN4, but which has not activated MERVL elements yet, through the analysis of single cell expression profiles of spontaneous 2C-like cells (Rodriguez-Terrones et al. 2018). This ZSCAN4+ cell population is ten times more abundant that 2C-like cells and expresses some 2C genes, overall showing an intermediate transcription profile between ESCs and 2C-like

cells (Rodriguez-Terrones et al. 2018). Importantly, they demonstrated that 2C-like cells arise primarily from the ZSCAN4+ population as shown by live cell imaging using a reporter cell line for both Zscan4 and MERVL (Rodriguez-Terrones et al. 2018).

Following a similar approach, the laboratory of Yi Zhang performed scRNA-seq analysis on the Dux-dependent reprogramming system toward 2C-like cells, also uncovering an intermediate cellular state (Fu et al. 2019). Similarly, this intermediate population shows downregulation of pluripotency factors. However, it does not show upregulation of 2C genes nor of Zscan4 (Fu et al. 2019). Remarkably, Xudong Fu and colleagues showed that the mechanistic logic of the Dux-mediated ESC to 2C-like transition occurs in a stepwise manner, controlled by the transcriptional regulator MYC and the maintenance DNA methyltransferase 1, DNMT1 (Fu et al. 2019). The model they propose is that, first, the pluripotent transcriptional network needs to be downregulated, followed by the activation of totipotent genes required for entering the 2C-like state. Additionally, the same laboratory characterized a year later the transcriptional dynamics of the reverse transition, from the 2C-like state to pluripotency (Fu et al. 2020). Importantly, they reported that the nonsense-mediated Dux mRNA decay plays an important role in the 2C-like state exit (Fu et al. 2020). In summary, ESC transition toward the 2C-like state involves changes in gene expression that are also accompanied by a dramatic remodeling of the spatial

organization of the genome. Despite the mentioned efforts in understanding the mechanistic details governing the 2C-like cell fate transition, the process by which specific genomic structures are being shaped during early development still needs to be fully unravel.

Features of 2C-like cells

2C-like cells share many features with blastomeres of the 2-cell stage embryos (Genet and Torres-Padilla 2020; Iturbide and Torres-Padilla 2020). We will next summarize the known characteristics of 2C-like cells reported to date.

Transcriptional features

2C-like cells downregulate pluripotency factors. As introduced in Chapter 1, pluripotent ESCs in culture, similar to cells of the ICM, express OCT4. However, immunostaining analysis indicate that 2-cell stage embryos lack OCT4 protein (Macfarlan et al. 2012; Do et al. 2013). SOX2 and NANOG are also expressed at very low levels in the 2-cell embryo, but their expression levels increase as development proceeds and reach highest levels at the blastocyst stage, specifically in the epiblast (Avilion et al. 2003; Silva et al. 2009; Komatsu and Fujimori 2015). 2C-like cells, mimicking the 2-cell stage embryo, show undetectable levels of OCT4, SOX2 and NANOG proteins (Macfarlan et al. 2012; Rodriguez-Terrones

et al. 2018), even though their mRNAs are expressed to a certain extent.

2C-like cells express genes from the embryonic program at the 2-cell stage (Figure 5). 2C-like cells were named as such due to the significant overlap of their transcriptome with that of 2-cell stage embryos (Macfarlan et al. 2012; Ishiuchi et al. 2015; Eckersley-Maslin et al. 2016; Rodriguez-Terrones et al. 2018). Particularly, the Zscan4 cluster of genes, which are expressed during ZGA in late 2-cell mouse embryos (Falco et al. 2007), are also markers of 2C-like cells both at the mRNA and protein level (Macfarlan et al. 2012; Rodriguez-Terrones et al. 2018). In fact, embryos depleted of ZSCAN4 display reduced rates of blastocyst formation, indicating a role for ZSCAN4 during early development. In ESCs, ZSCAN4 regulates telomere elongation and genomic stability by preventing telomere shortening in ESCs (Falco et al. 2007; Zalzman et al. 2010). It is noteworthy that ZSCAN4 alone is not a marker of 2C-like cells by itself, especially if this occurs in the absence of MERVL reactivation, as demonstrated in the work of Diego Rodriguez-Terrones and colleagues (Rodriguez-Terrones et al. 2018). Furthermore, the chromatin accessibility landscape of Zscan4+ cells is distinct from that of 2C-like cells (Zalzman et al. 2010; Eckersley-Maslin et al. 2016; Rodriguez-Terrones et al. 2018).

2C-like cells express repetitive elements characteristic of 2-cell stage embryos. MERVL transcripts are drastically more abundant in the 2-cell stage embryo than in the oocyte

(Svoboda et al. 2004; Macfarlan et al. 2012). MERVL reactivation at the 2-cell stage is thought to drive the expression of some ZGA genes (Peaston et al. 2004; Macfarlan et al. 2012). Then, the appearance of 2C-like cells can be identified and monitored using a "*2C reporter*", which consists of a MERVL-LTR driving the expression of a fluorescent protein. Notably, 2C-like cells can be identified by positive immunostaining for GAG, which is a marker gene that is present in nearly half of all MERVL repeats. 2C-like cells also reactivate the transcription of major satellite sequences that constitute pericentromeric chromatin (Ishiuchi et al. 2015), which is a feature of the zygote and of 2-cell embryos (Puschendorf et al. 2008; Probst et al. 2010; Santenard et al. 2010) (Figure 6). In the embryo, major satellites are necessary for pericentromeric heterochromatin reorganization at the late 2-cell stage (Casanova et al. 2013) (Figure 6). In the case of ESCs, it is unusual to detect major satellite transcription foci, whereas 2C-like cells show a severe increase in the number of major satellite foci (Ishiuchi et al. 2015; Tosolini et al. 2018).

Chromatin and nuclear organization features

2C-like cells display high histone mobility. Histones are quite immobile proteins once incorporated into chromatin. However, fluorescent recovery after photobleaching (FRAP) experiments performed in totipotent 2-cell embryos has shown that they display unusually high core-histone mobility (Boskovic et al. 2014; Ooga et al. 2016). Interestingly, the reported increase in

mobility decreases in pluripotent cells. 2C-like cells also display high core-histone mobility, as it was the case for 2-cell embryos, suggesting a link between histone mobility and greater cellular plasticity (Boskovic et al. 2014; Ishiuchi et al. 2015).

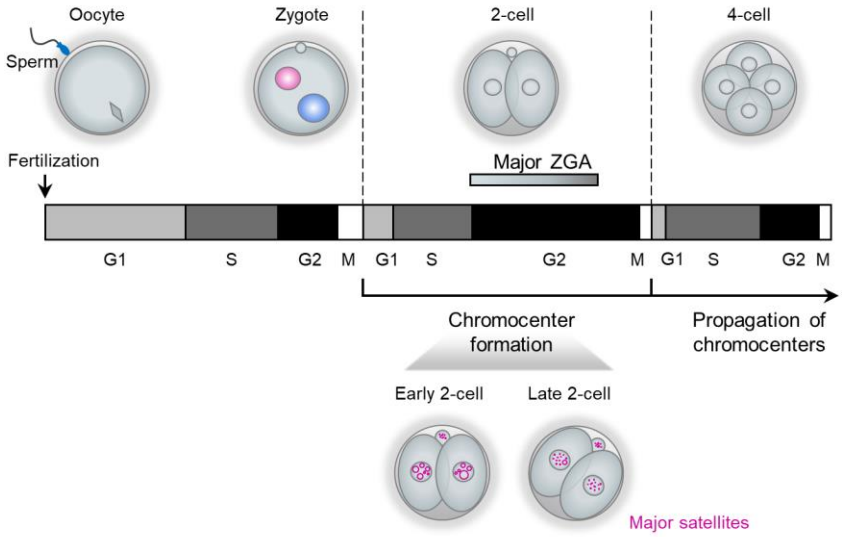

Figure 6. Cell cycle progression in mouse embryos could be coupled with chromocenter formation. Schematic showing the distribution of the cell cycle phases during the first two cleavage divisions. In the 2-cell stage, which is the embryonic stage with the longest G2 phase and coinciding with the ZGA, chromocenters are formed as illustrated by the reorganization of major satellites. From the 4-cell stage onward, chromocenters are propagated to all the cells. ZGA, zygotic genome activation. Adapted from Gu et al., 2018, Casanova et al., 2013 and Probst et al., 2010.

2C-like cells do not have DAPI-stained chromocenters (Figure 4). Chromocenters are composed of the centromeric heterochromatin regions from several chromosomes that cluster together in the 3D nuclear space. They typically appear as dense nuclear puncta when visualized by DNA staining. In

mice, chromocenters are established after ZGA, from the late 2-cell stage onwards, and become clearly detectable by the 8-cell stage (Probst et al. 2007; Aguirre-Lavin et al. 2012). Therefore, the zygote and early 2-cell stage embryo do not have defined chromocenters. In the same way, 2C-like cells do not have DAPI-stained chromocenters, in contrast to ESCs (Ishiuchi et al. 2015). Of note, ZSCAN4+ cells also have some degree of decondensed pericentromeric chromatin (Akiyama et al. 2015). Interestingly, the loss of chromocenters alone cannot induce the transition of ESCs to the 2C-like state (Ishiuchi et al. 2015).

2C-like cells have a chromatin landscape similar of that of the 2-cell embryo. Firstly, 2C-like cells have an increased chromatin accessibility. Genome-wide assay for transposase-accessible chromatin using sequencing (ATAC-seq) profiling demonstrated that 2C-like cells have increased chromatin accessibility at MERVL elements, which clearly distinguishes them from ZSCAN4+ cells (Eckersley-Maslin et al. 2016; Rodriguez-Terrones et al. 2018). The open chromatin landscape of 2C-like cells resembles that of the 2-cell stage embryo (Wu et al. 2016a; Hendrickson et al. 2017). Also, loss of the activity of the chromatin assembly factor 1 (CAF-1) complex leads to a more open chromatin structure and more 2C-like cells (Ishiuchi et al. 2015), as we will discuss next in the 2C-like regulators section.

Thus, 2C-like cells are characterized by accessible chromatin at MERVL elements and acquire an open/closed chromatin landscape resembling that of the 2-cell stage embryo. Secondly, it is also becoming evident that 2C-like cells adopt a higher-order chromatin organization similar of that of the 2-cell embryo. Thirdly, 2C-like cells present global DNA hypomethylation compared to ESCs grown in sLIF (Eckersley-Maslin et al. 2016). Finally, and in line with a globally more plastic chromatin as illustrated with the many similarities with 2-cell embryos, 2C-like cells display higher levels of some histone modifications associated with transcriptional activation (Macfarlan et al. 2012; Ishiuchi et al. 2015). In particular, 2C-like cells have more H3K4me2 and acetylated H3 and H4 histones (Macfarlan et al. 2012; Ishiuchi et al. 2015).

2C-like cells recapitulate some metabolic features of 2-cell embryos. Complementary work done by the laboratories of Maria-Elena Torres-Padilla and Wee-Wei Tee identified that 2C-like cells exhibit decreased glycolytic and respiratory activity, lower levels of reactive oxygen species and increased glucose uptake (Hu et al. 2020; Rodriguez-Terrones et al. 2020).

Cellular potency features

In the field of stem cell and developmental biology, the word "*totipotency*" has been associated with different meanings, as we exemplified in Chapter 1 (Condic 2014; Morgani and Brickman 2014; Baker and Pera 2018; Torres-Padilla 2020). In

brief, totipotency can be firstly defined as the ability of a single cell to contribute to all cell lineages, both from embryonic and extraembryonic origin, in an organism (zygote up to 4/8-cell stage). Then, in the more stringent scenario, totipotency can be viewed as the ability of a single cell to develop into a complete organism (zygote and 2-cell blastomeres). What it is important to highlight here is that 2C-like cells are thought to possess an expanded cell fate potential based on several observations, although this particular aspect is a matter of active debate in the field. Of note, 2C-like cells must not be confused with extended potential stem cells (Yang et al. 2017b). In 2016, the laboratories of Deng, Izpisua Belmonte and Shen reprogrammed mouse and human ESCs to extended potential stem cells (EPSCs), using the complex "*LCDM*" culture cocktail (Yang et al. 2017b) (LCDM stands for hLIF, CHIR99021, DiM or (S)-(+)-dimethindene maleate, and MiH or minocycline hydrochloride). When injected into mouse embryos, EPSCs could contribute to the embryonic and extraembryonic tissues, including the placenta (Yang et al. 2017b), as we will later discuss for 2C-like cells. However, the most important difference is their transcriptome. 2C-like cells do not express extra-embryonic markers, but they do have the ability to form extra-embryonic tissues in chimeras (Macfarlan et al. 2012). Contrarily, EPSCs already express markers of embryonic and extra-embryonic lineages in culture conditions (Yang et al. 2017b).

There have been many efforts in the field to address whether 2C-likes cells display truly totipotent features *in vivo*. We will summarize here most of the studies that have provided insights in this particular, somehow controversial, aspect.

In the initial study where 2C-like cells were discovered, aggregation of four 2C-like cells with morula-stage embryos to form chimeric mice was used as a test for totipotency (Macfarlan et al. 2012). Spontaneous 2C-like cells contributed to both embryonic and extraembryonic lineages, such as placenta and yolk sac, and even primordial germ cells (Macfarlan et al. 2012). By contrast, ESCs were only be detected in the ICM but not the TE as expected. Overall, this would suggest that 2C-like cells might have a bipotential competency, based on their potential to contribute to both the TE and ICM lineages.

In 2013, the laboratory of Manuel Serrano reported the first evidence that full cell reprogramming to iPSCs can be achieved within a living organism, using a "*reprogrammable mice*" in which OSKM expression could be induced with doxycycline (Abad et al. 2013). Importantly for our discussion, *in vivo*-generated iPSCs are significantly different from *in vitro*-generated ones, showing an expression profile remarkably more similar to that of true ESCs. Actually, like ESCs, *in vivo*-derived iPSCs can efficiently contribute to the ICM when injected into morulas. However, it was not only that since *in vivo* iPSCs can also undergo trophectoderm lineage

differentiation, showing some totipotency features that even ESCs do not display (Abad et al. 2013).

In the stem cell field, it is often discussed that while the chimera assays are used as the "*gold standard*" to test for pluripotency, they may not be best suited to assess totipotency. Usually, some of the arguments are that (i) chimera assays are performed at a stage at which the recipient cells (e.g. morula embryos) are already beyond the totipotency stage, lacking the environmental factors of totipotent cells; (ii) previous findings report that teratomas and EBs from pluripotent ESCs also express extra-embryonic markers, although at low levels (Spangler et al. 2018). In an ideal scenario, one should assess the *in vivo* potential of a single cell. However, performing this using a single cultured cell can be technically challenging and, until now, no study has ever reported any experiment of this kind using 2C-like cells.

In 2015, work done by Takashi Ishiuchi and colleagues used nuclear transfer to test the reprogramming efficiency of 2C-like cells (Ishiuchi et al. 2015). The rational of this experiment was to measure cellular plasticity since the reprogramming efficiency depends on the potential of the donor cell and, accordingly, early embryo blastomeres are better donors than ESCs and terminally differentiated cells (McGrath and Solter 1984). This experimental design cannot directly determine the potency state of a cell given that earlier work already demonstrated that somatic cells can be reprogrammed via

SCNT (Gurdon 1962; Wilmut et al. 1997). 2C-like cells assessed in this study showed a success rate of reprogramming almost 4-fold higher compared with ESCs (Ishiuchi et al. 2015). These data demonstrate that 2C-like cells display higher plasticity than ESCs, but they cannot formally inform us whether or not 2C-like cells are in fact totipotent.

In the following years, several other studies assessed the potential of 2C-like cells induced after some perturbations. In 2017, a study from the laboratory of Lin He showed that ESC cultures deficient for the microRNA miR34a contain more 2C-like cells (Choi et al. 2017). ESCs that do not express miR34a (miR34a$^{-/-}$) were compared with WT ESCs using teratoma and EB formation assay. Of note, 2C-like cells arising from miR34a$^{-/-}$ cultures were not analyzed for developmental potential. miR34a$^{-/-}$ ESCs highly expressed markers of TE and extra-embryonic endoderm, a derivative of the ICM (Choi et al. 2017). Chimeric embryos containing miR34a$^{-/-}$ ESCs showed contribution to TE and ICM even after injecting one single cell in recipient morula-stage embryos. miR34a$^{-/-}$ ESCs also contributed to differentiated cell lineages in the embryo, yolk sac and placenta *in vivo*, driving to similar conclusion of previous reports and supporting the notion that 2C-like cells might have a higher potential (Macfarlan et al. 2012; Choi et al. 2017). More recently, in 2020, work done by the laboratory of Wee-Wei Tee described the maternal negative elongation factor A (NELFA) as a positive regulator of 2C-like cells (Hu et al. 2020). Here again, using chimera assays, they

demonstrated that NELFA high expressing cells have a higher probability of contributing to the TE and the ICM, compared with NELFA low expressing cells, suggesting bipotential competency as in previous occasions. Yet, no study has ever studied the *in vivo* potential of 2C-like cells generated after the expression of the potent inductor Dux.

Regulators of 2C-like cells

Recent efforts in the field have focused on identifying novel strategies to induce 2C-like cells by modulating master regulators or repressors of this process. As previously indicated, spontaneously arising endogenous 2C-like cells represent less than 1 % of the cells in culture, thus limiting the range of downstream technologies that can be applied. We have already introduced the discovery of Dux, CAF-1 and miR34a, and the major effects that their modulation produce on the reprogramming of ESCs toward the 2C-like state. We will now discuss several other studies reporting some of the most promising regulators of 2C-like cells.

In 2018, a team led by Miguel Ramalho-Santos identified another type of transposable element with a role in regulating the 2C transcriptome: the LINE1 elements (Jachowicz et al. 2017; Percharde et al. 2018). LINE1 retrotransposon constitutes the most abundant class and, although being thought to be deleterious for cells, it is highly expressed during

development (Percharde et al. 2018). In fact, ESCs express LINE1 repeats, whose RNA is localized in the nucleus. Michelle Percharde and colleagues reported that nuclear LINE1 RNA can recruit Nucleolin and KAP1 to sequester the Dux locus and keep it transcriptionally silent, therefore inhibiting the transition to the 2C-like state (Percharde et al. 2018).

Later that year, the laboratory of Anne Dejean identified that the small ubiquitin-like (SUMO) modification also plays a role in repressing Dux by tethering the non-canonical polycomb repressive complex PRC1.6 (Cossec et al. 2018). Although SUMO activity is important for Dux regulation and 2C-like reprogramming, it plays a more general and pleiotropic role in repressing large genomic regions containing 2C-like genes and reprogramming resistant regions (RRRs) (Cossec et al. 2018). Of note, RRRs mainly consist of chromatin regions decorated by the repressive trimethylation of lysine 9 on histone H3 (H3K9me3) histone modification. Overall, SUMO was suggested to have a role as a repressor of reprogramming or a "*gatekeeper*" of cell fate, rather than as a specific regulator of the 2C-like cell transition, since it was also important in regulating other cell fate transitions (Cossec et al. 2018).

Although Dux can activate many 2C-related genes and drive reprogramming to a 2C-like state in cell culture, in the embryo Dux itself is activated during ZGA (Eckersley-Maslin et al. 2018). Thus, an active line of research was trying to identify

potential maternally inherited factors upstream of Dux that could induce its expression in the embryo. Recent studies have identified the developmental pluripotency associated factor 2 (DPPA2) and 4 (DPPA4), and NELFA as activators of Dux transcription (De Iaco et al. 2019; Eckersley-Maslin et al. 2019; Hu et al. 2020). The laboratories of Wolf Reik and Didier Trono showed that Dppa2/Dppa4 double knockout ESCs are devoid of 2C-like without exerting any effect on pluripotency or self-renewal. Importantly, DPPA2/DPPA4 overexpression lead to an increase in the proportion of 2C-like cells, which they demonstrated to be through the binding on the Dux. Nicely, Dux overexpression rescues the emergence of 2C-like cells in the Dppa2/Dppa4 double KO cell line, but not vice versa, thus indicating that DPPA2/DPPA4 act upstream of Dux (De Iaco et al. 2019; Eckersley-Maslin et al. 2019). This finding was accompanied by the discovery of NELFA as a key 2C-like regulator, demonstrated as well to be acting upstream of Dux (Hu et al. 2020). Similar to DPPA2/DPPA4, NELFA overexpression triggers 2C-like cell reprogramming, through direct binding to the Dux locus to activate its transcription.

In 2020, work led by the laboratory of Vincenzo Costanzo identified aphidicolin as an unexpected positive regulator of 2C-like cells (Atashpaz et al. 2020). Aphidicolin (APH) is a reversible inhibitor of DNA polymerases that activates ataxia-telangiectasia mutated and Rad3-related (ATR) by stalling replication forks progression, which was one of the aspects they were initially investigating in their research (Aze et al.

2016; Atashpaz et al. 2020). ATR activation was shown to induce 2C-like cell reprograming, but without activating the DNA Damage Response (DDR) pathway (Atashpaz et al. 2020). Among other interesting observations, Sina Atashpaz and colleagues also reported a post-transcriptional regulation of the Dux gene in the ATR-induced transition to the 2C-like state, which increases Dux protein levels without altering Dux transcription (Atashpaz et al. 2020).

More recently, the laboratory of Jianlong Wang described that the microRNA miR-344 activates 2C-like cell induction (Yang et al. 2020). miR-344 targets and silences the zinc finger MYM-type containing 2 (ZMYM2) protein, a factor that binds to and represses MERVL transcription. Remarkably, Fan Yang and colleagues showed that activation of MERVL alone is sufficient to increase the proportion of 2C-like cells in culture via a CRISPR-Cas9 synergistic activation mediator (SAM) approach (Yang et al. 2020). However, even if MERVL activation alone was enough to reprogram ESCs to the 2C-like state, it was not sufficient to maintain 2C-like cells stably in culture.

To sum, since the identification of 2C-like cells, many relevant features and key regulators of the 2C-like cell reprograming have been identified. Still, many questions remain to be solved. Nevertheless, what it is clear is that 2C-like cells deserve their name based on solid evidences reporting many features shared with their *in vivo* counterparts, the blastomeres of the 2-cell stage embryo. Additionally, 2C-like cells offer a

tremendous opportunity to unravel pivotal rules governing the initial phases of mammalian life.

In this Chapter, we have extensively presented and discussed fundamental aspects of cell identity and approximations that demonstrate the degree of cellular plasticity that reside in many potency states, exemplified from the initial observations through nuclear transfer to more recent discoveries such as iPS cell reprogramming or the identification of 2C-like cells.

CHAPTER 3 - Chromatin structure and epigenetics

Chromatin organization in the mammalian genome

Genomes are more than linear sequences of DNA. Two meters of mammalian DNA might be compacted and packed to fit in the cell nucleus, each of which having a diameter of 5 to 10 µm, depending on the species and the cell type (Figure 7). Therefore, chromosomes must achieve highly folded structures. DNA associates with histone proteins to form a higher-order complex known as chromatin. Early observations done primarily by Emil Heitz in 1928 already distinguished differentially stained regions of interphase chromatin, indicating that chromosome structure is heterogeneous yet highly organized (Figure 7). These early descriptions termed the chromatin-dense fraction as "*heterochromatin*" and the more open configuration as "*euchromatin*". We will extensively cover the heterochromatin component later in this Chapter. Based on many subsequent studies, the community has built up a hierarchical model of chromosome folding, where different architectural features are present at multiple scales ranging from higher-level structures to the basic repeated unit, the nucleosome. (Sexton and Cavalli 2015; Furlong and Levine 2018).

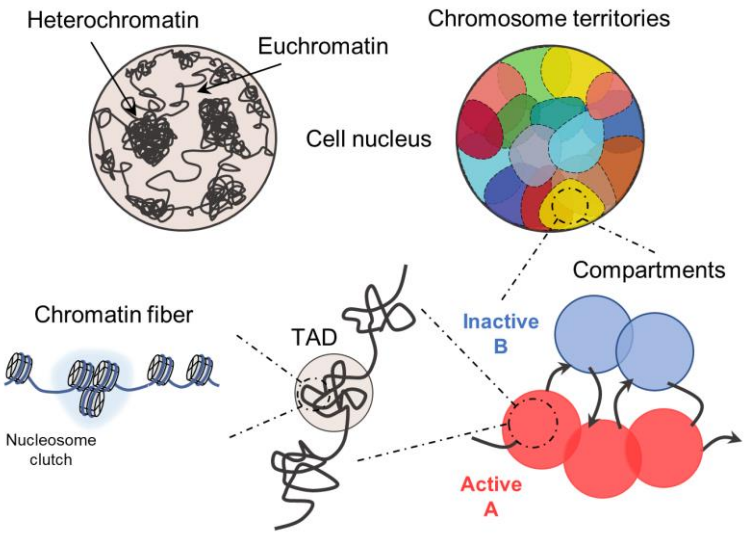

Figure 7. Hierarchical folding of the mammalian genome. Overview of the different scales of chromatin organization inside the cell nucleus, spanning from heterochromatin/euchromatin segregation and chromosome territories to compartments, TADs and the chromatin fiber. TAD, topologically associating domain.

Many of the discoveries done in the chromatin structure and epigenetics field have been mediated by technology advancements. Microscopy has been a powerful tool to observe chromatin organization in mammalian cells since early days, and, indeed, we will discuss some studies exploiting advance microscopy technologies shortly (Ricci et al. 2015; Ou et al. 2017; Lakadamyali and Cosma 2020). However, the development of chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Bernstein et al. 2006; Barski et al. 2007; Johnson et al. 2007; Mikkelsen et al. 2007) and chromosome confirmation capture (3C) (Dekker et al. 2002), and its high-throughput derivatives such as Hi-C (Dostie et al. 2006; Simonis et al. 2006; Fullwood et al. 2009; Lieberman-

Aiden et al. 2009; Schoenfelder et al. 2010; Hughes et al. 2014; Rao et al. 2014; Mifsud et al. 2015; Grob and Cavalli 2018), have revolutionized the field of chromatin biology. These methods have allowed mapping DNA contacts, epigenetic modifications and chromatin-binding proteins in a genome-wide fashion and integrate the spatial information of the three-dimensional (3D) organization of the genome (Furey 2012; Grob and Cavalli 2018).

Chromosomes occupy distinct territories within the interphase nucleus, with a non-random segregation that is linked to gene density and activity (Cremer and Cremer 2010) (Figure 7). Particularly, the nuclear positioning of genes within the nuclei can inform about their transcriptional status. Active genes tend to localize more often in the nuclear interior, whereas repressed regions are found closer to the periphery (Bickmore 2013; van Steensel and Belmont 2017). Chromosomes later segregate into regions of preferential long-range interactions that form two mutually excluded types of chromatin, the so-called A and B compartments (Lieberman-Aiden et al. 2009; Rao et al. 2014; Denker and de Laat 2016; Wang et al. 2016) (Figure 7). A compartments correspond to gene-rich and active chromatin (euchromatin-like), while B compartments are mostly enriched in repressive chromatin (heterochromatin-like) (Lieberman-Aiden et al. 2009; Rao et al. 2014).

Zooming in, at the scale of tens to hundreds of kilobases, chromosomes appear to be organized into discretely folded

modules named topologically associating domains (TADs) (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012; Rowley and Corces 2018) (Figure 7). These domains have preferential intradomain interactions compared to interdomain interaction with the neighboring chromatin domains (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012; Bintu et al. 2018). TADs tend to form long-range interaction networks with regions of similar transcriptional activity or chromatin state, which has been proposed to allow coordinated control of gene expression programs. TADs have been described in many species and are thought to be highly conserved, showing cell type-specificity, and a functional feature of the genome (Lupianez et al. 2015), although this is a matter of active research and controversy in the field.

Distal regulatory elements such as enhancers are brought into physical proximity with their related gene promoters via the formation of chromatin loops (Deng et al. 2012). Then, crystallographic, electron and light microscopy studies have given detailed structural insight into the primary folding of DNA into 10 nm nucleosome fibers, which exist at different local compaction states (Fussner et al. 2012; Ricci et al. 2015; Ou et al. 2017; Lakadamyali and Cosma 2020). Chromatin fibers are formed by the repetition of nucleosomes, which are composed of a core particle with 147 base pairs (bp) of double stranded DNA wrapped around octamers containing two copies of the histone proteins H2A, H2B, H3 and H4, and a linker histone H1 (Luger et al. 1997) (Figure 7). Our laboratory

discovered that nucleosomes assembled in heterogeneous groups of varying sizes, named "*nucleosome clutches*", along the chromatin fiber (Ricci et al. 2015). Indeed, the number of nucleosomes inside the clutches and their compaction are cell-type specific and the clutch size correlates with the pluripotency potential of stem cells (Ricci et al. 2015). More recently, our laboratory has also identified the clutch-associated DNA, providing a direct link between the epigenetic state of nucleosome clutches and their DNA packing density (Otterstrom et al. 2019). Ultimately, nucleosomes functionally regulate the access of DNA-binding proteins and serve as platform for epigenetic modifications to regulate gene expression, as we will introduce in the following sections.

Epigenetics and histone modifications

The term "*epigenetics*" was first coined in the 1940s, by the developmental biologist Conrad H. Waddington, and referred to the study of processes by which genotypes give rise to phenotypes. Currently, epigenetics is more often used to describe heritable changes in gene expression caused by mechanisms other than changes in the underlying DNA sequence. DNA methylation, chromatin remodeling, histone variants, histone post-translational modifications, and noncoding RNAs have been identified as the major forms of epigenetic regulation (Graf and Enver 2009; Gaspar-Maia et al. 2011; Stadhouders et al. 2019). These features contribute

to generate specialized genomic domains for a versatile and plastic chromatin landscape (Groth et al. 2007; Filipescu et al. 2014; Yadav et al. 2018). The exquisite combination of all the above, in addition to other contributing factors, orchestrate which set of genes are expressed at a given developmental time, therefore, leading to define cellular identity. In this Chapter, we will discuss some of these aspects although with an emphasis on histone post-translational modifications.

Although the nucleosomal unit appears homogeneous, variation in histone composition can also introduce different functionalities. Different variants of histones can be incorporated, mainly for histone H3 and histone H2A since histones H2B and H4 appear to be predominantly canonical (Henikoff and Smith 2015; Henikoff and Greally 2016). In mammals, there are several examples of histone variants such as the histone H3.1 and H3.2, which are deposited during DNA replication in S-phase. On the contrary, the histone H3.3 is involved in histone exchange at active transcription units and at pericentric and telomeric heterochromatin. The proteins that interact with, process and deposit histones on the core nucleosomal particle are called histone chaperones. Histone chaperones, or loaders, mediate the assembly and disassembly of nucleosomes in different chromatin contexts, together with another dedicated class of proteins named chromatin or nucleosome remodelers. For example, during S-phase, chromatin assembly factor 1 (CAF-1) incorporates histones H3.1 and H3.2 into DNA, whereas histone cell cycle

regulator (HIRA) has a role in incorporating H3.3 into transcribed genes (Martire and Banaszynski 2020). At pericentric and telomeric heterochromatin, the complex formed by the alpha-thalassemia/mental retardation X-linked syndrome protein (ATRX) and the death-domain associated protein (DAXX) is required for histone H3.3 incorporation (Henikoff and Smith 2015; Henikoff and Greally 2016; Martire and Banaszynski 2020). At the centromere of mammalian cells, the canonical histone H3 is replaced by the centromere protein A (CENPA). Among histone H2A variants in mammals, it is noteworthy to highlight macroH2A since it accumulates in silent chromatin. MacroH2A is enriched at the inactive X chromosome of female mammals, consistent with its association with transcriptionally repressed chromatin.

Changes in histone composition are not the only way by which information can be added to nucleosomes. Histone proteins are subjected to post-translational modifications (Kouzarides 2007) (Figure 8). Much like DNA, which can be chemically modified (e.g. DNA methylation), modifications of histones are a way to regulate gene expression and chromatin structure. Histone post-translational modifications can occur at many amino acids in the body but most of the modifications studied occur on the unstructured N-terminus of histones (commonly referred to as histone tails) that protrude from the nucleosome (Figure 8). Of note, the tail of histone H3 interacts with the major groove of DNA (Luger et al. 1997), making an ideal candidate for altering DNA-histone association. Remarkably,

histone tails can be modified in a variety of ways, expanding the chemical spectrum of modifications and their combinatorial complexity (Figure 8).

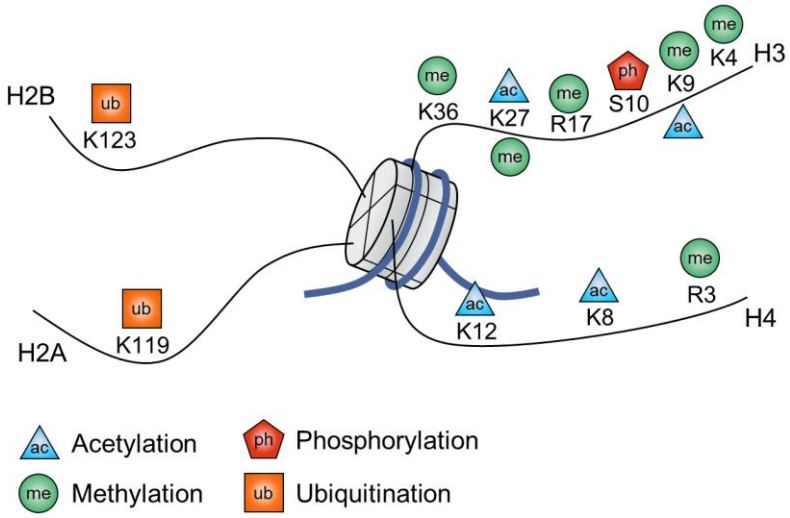

Figure 8. Histone proteins can be post-translationally modified. Overview of some of the most studied histone post-translational modifications (PTMs) occurring in histone N-terminal tails. Histone modifications are indicated as colored geometrical forms. Ac, acetylation. Me, methylation. Ph, phosphorylation. Ub, ubiquitination. K, lysine. R, arginine. S, serine. Adapted from Bannister & Kouzarides, 2011.

Specific residues on histones can be post-translationally modified via the covalent addition of chemical groups. Phosphorylation of serine (S), acetylation and methylation of lysine (K), and methylation of arginine (R) residues are the most prominent histone modifications, together with ubiquitination and SUMOylation (small ubiquitin-like modifier), although to a lesser extent (Bannister and Kouzarides 2011) (Figure 8). Notably, multiple methyl groups can be added to lysines and arginines, thereby increasing the complexity.

Overall, the function of histone modifications is to regulate the accessibility of DNA to the transcriptional machinery and to serve as marks to recruit effector proteins, triggering diverse functional outcomes.

Histone modifications are somewhat dynamic in the sense that they are not usually permanent and static, particularly when cells are challenged to different stimuli or exposed to changing environments or developmental contexts (Bannister and Kouzarides 2011). Enzymatic "*writers*" carry out the activity to deposit these modifications, which is counterbalanced by "*eraser*" enzymes that can remove them. A third player in the equation are factors that can specifically recognize the given histone modification named as "*readers*". As recently introduced, histones can suffer different rewiring processes in different residues, but we will mainly focus on modifications that lead to repression and, in particular, on the deposition of three methyl groups at lysine 9 on histone H3 (H3K9me3) since it will result in heterochromatin formation (Bannister and Kouzarides 2011).

Heterochromatin and mechanisms of transcriptional silencing

There are two broad types of heterochromatin. Constitutive heterochromatin is present in all cell types and is typically composed of repeat-rich and gene-poor regions around

centromeres and telomeres, exerting stable and almost irreversible transcriptional silencing. In contrast, facultative heterochromatin is established in a cell type specific manner on genomic regions that generally have normal gene density and that are developmentally regulated, leading normally to transcriptional repression (Elgin and Reuter 2013) (Figure 9).

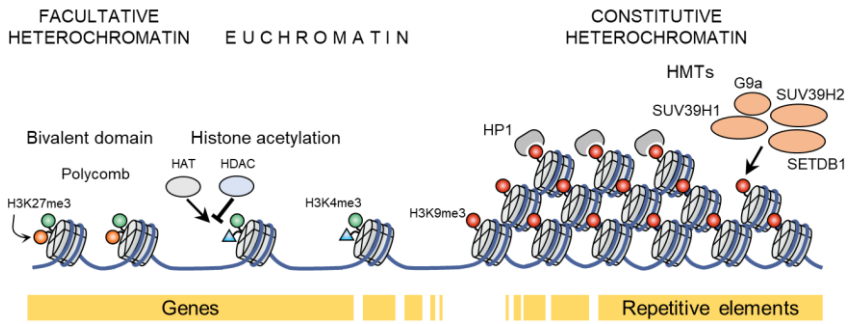

Figure 9. The chromatin landscape. Overview of the open, active euchromatin, the developmentally regulated facultative heterochromatin and the silent constitutive heterochromatin. Histone modifications are indicated as colored geometrical forms. HAT, histone acetyltransferases. HDAC, histone deacetylases. HMTs, histone methyltransferases. HP1, heterochromatin protein 1. Adapted from Gaspar-Maia et al., 2010.

Trimethylation of lysine 9 (H3K9me3) (Nakayama et al. 2001; Wang et al. 2018) and lysine 27 (H3K27me3) (Wiles and Selker 2017) on histone H3 are the best known histone modifications associated with repressive constitutive and facultative heterochromatin, respectively. These marks are established and recognized by distinct writer and reader complexes and are typically localized to different genomic regions, suggesting that they associate with distinct types of chromatin. H3K27me3 is found on many regions silenced in a

cell-specific manner, such as the X chromosome and the developmentally regulated homeotic (Hox) genes (Beuchle et al. 2001; Plath et al. 2003; Ringrose and Paro 2004; Schuettengruber et al. 2017). The role of H3K27me3 in controlling expression of developmentally regulated genes has been extensively studied (Schuettengruber et al. 2007; Aloia et al. 2013; Schuettengruber et al. 2017) (Figure 9).

H3K9me3 is enriched in constitutive heterochromatin, which is usually located in repetitive DNA sequences at centromeric and telomeric regions from yeast to human (Richards and Elgin 2002) (Figure 9). H3K9me3-marked constitutive heterochromatin is best known for its role in chromosome architecture, genome stability and stable repression of transposable elements (Mikkelsen et al. 2007; Matsui et al. 2010; Rowe et al. 2010; Karimi et al. 2011; Klenov et al. 2011; Riddle et al. 2011; Wang and Elgin 2011; Sienski et al. 2012; Le Thomas et al. 2013; Rozhkov et al. 2013; Pezic et al. 2014). Interestingly, pericentric heterochromatin clusters in compacted structures named chromocenters, which vary in number and size depending on developmental stage and cell type (Jones 1970; Probst and Almouzni 2011) (Figure 10). Chromocenters can be observed as DNA-dense clusters under the microscope (Figure 10). However, how these structures emerge for the first time during early embryogenesis is poorly understood, thus, we will investigate it thoroughly in this Thesis. Nonetheless, we should first revise how H3K9me3 is

established and which factors are responsible for its deposition and recognition in order to promote a silent chromatin status.

Factors involved in heterochromatin formation first emerged from genetic screens for factors affecting position effect variegation (PEV) in Drosophila, a phenomenon by which the relocation of protein-coding genes that normally reside in euchromatin next to heterochromatin leads to their variegated repression (Elgin and Reuter 2013).

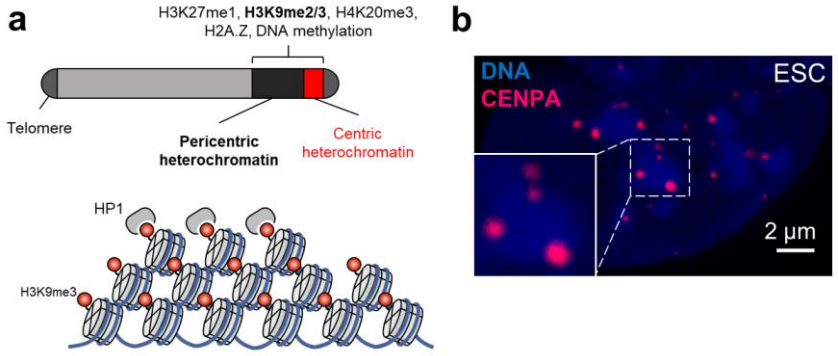

Figure 10. Pericentric heterochromatin clusters in structures named chromocenters. a, Typical murine acrocentric chromosome (top) highlighting some of the epigenetic modifications often present in (peri)centric heterochromatin. Schematic representation of chromocenters at the finest layer of chromatin folding (bottom). **b**, Chromocenters can be visualized in embryonic stem cell (ESC) as DAPI-dense clusters composed of centromeric heterochromatin from several chromosomes. The centromeric-specific histone H3 variant (CENPA) indicates the interchromsomal nature of these clusters in the 3D nuclear space. Scale bar, 2 µm.

Many suppressors of PEV (also referred to as Su(var) genes) encode factors that are necessary for heterochromatin-induced gene repression. Their molecular characterization identified proteins that establish and maintain heterochromatin

structure as well as remove heterochromatin marks, including then writers, readers and erasers of the H3K9me3 mark. The Su(var)3-9 gene family was found to be the first known H3K9-specific histone methyltransferase in human (SUV39H), Drosophila (Su(var)3-9) and yeast (Clr4) thanks to pioneering work done in the laboratories of Thomas Jenuwein, Shiv Grewal and Axel Imhof (Rea et al. 2000; Czermin et al. 2001; Nakayama et al. 2001) (Figure 9).

Other complementary studies identified two additional conserved families showing H3K9-specific histone methyltransferase activity, including SETDB1/ESET (Matsui et al. 2010; Rowe et al. 2013a; Rowe et al. 2013b), which correspond to Setdb1/eggless in Drosophila (Schultz et al. 2002), and G9A (Tachibana et al. 2001) (Figure 9). Despite similar activities *in vitro*, these histone methyltransferases differ in their *in vivo* specificity, especially in the genomic regions where they are more active and their bias toward a specific methyltransferase activity (mono-, di- or tri-methylation). For example, G9A predominantly regulates H3K9me1 and H3K9me2 and is essential for embryogenesis (Tachibana et al. 2001; Tachibana et al. 2002). Indeed, lack of G9A results in a dramatic loss of H3K9me2, except at constitutive heterochromatic regions, and an increase in active modifications like acetylation of lysine 9 on histone H3 (H3K9ac) and dimethylation of lysine 4 on histone H3 H3K4me2 (Tachibana et al. 2002). Of note, G9A has been reported to have methyltransferase activity for both lysine 9

and lysine 27. Then, the SUV39 family proteins are typically associated with H3K9me2 and H3K9me3 at centromeric and telomeric regions, while SETDB1 primarily acts on euchromatin regions such as transposable elements (Aagaard et al. 1999; Schotta et al. 2002; Schultz et al. 2002; Peters et al. 2003; Rice et al. 2003; Karimi et al. 2011). SETDB1 is also required for embryogenesis as null mice embryos die shortly after implantation (Dodge et al. 2004).

Besides, it should be noted that heterochromatin can be also established by non-coding RNA, components of the RNA interference (RNAi) machinery and even transcription factor-mediated recruitment of H3K9me3 histone methyltransferases, depending on the species and cellular context (Volpe et al. 2002; Bulut-Karslioglu et al. 2012; Becker et al. 2016; Allshire and Madhani 2018; Janssen et al. 2018). For example, in the fission yeast *Schizosaccharomyces pombe* (S. pombe) and in plants, constitutive heterochromatin formation requires components of the RNAi machinery and transcription of the locus targeted for silencing (Selker 1998; Djupedal et al. 2005; Kato et al. 2005; Martienssen et al. 2008).

Proteins from the highly conserved heterochromatin protein 1 (HP1, also referred to as chromobox (CBX) proteins in Drosophila or Swi6 in yeast) family are H3K9me3 readers (Iyengar and Farnham 2011; Groh and Schotta 2017) (Figure 9). Of note, drosophila and mammalian genomes encode several paralogs of the H3K9me3 reader HP1 family that

exhibits distinct localization patterns and post-translational modifications (Lomberk et al. 2006a; Lomberk et al. 2006b). HP1 proteins consist of an N-terminal chromodomain, which is required for their specific interaction with the methylated modification in H3K9 (Bannister et al. 2001; Jacobs et al. 2001; Lachner et al. 2001), a hinge region and a C-terminal chromoshadow domain. The current accepted model suggests that two HP1 proteins bound to H3K9me3 residues can dimerize via interaction through their chromoshadow domains, bringing their nucleosomes in closer proximity and causing chromatin condensation (Canzio et al. 2011; Hiragami-Hamada et al. 2016; Machida et al. 2018). The chromoshadow domain also mediates HP1 interaction with additional proteins (Platero et al. 1995), serving as a platform to recruit other chromatin remodeling and modifying complexes. Notably, HP1 mediates the recruitment at pericentric regions of the additional histone methyltransferases SUV420H1 and SUV420H2, which catalyze the deposition of the repressive trimethylation of lysine 20 on histone H3 (H4K20me3) (Schotta et al. 2004).

Since we are now discussing the H3K9me3 reader protein HP1, it is important to highlight that the laboratories of Geeta Narlikar and Gary Karpen reported different physical states of heterochromatin inferring a novel role for HP1 protein in directing heterochromatin formation by liquid phase separation (Larson et al. 2017; Strom et al. 2017). Based on biophysical parameters, they suggested a model where heterochromatin exists in multiple physical states: a soluble one, representing

the least repressed form and allowing DNA access; a liquid-droplet like state, regulating gene expression; and a gel-like state characterizing constitutive heterochromatin at centromeres and exerting structural functions during chromosome segregation (Larson and Narlikar 2018).

Importantly, H3K9me3 reader and writer activities can be coupled since members of the SUV39 family have a conserved chromodomain that can mediate binding to H3K9me3, implying direct binding of the mark by its own writer. This is an example of the complex and tightly regulated feedback loop that exist to ensure H3K9me3 maintenance and propagation (Vermaak and Malik 2009; Allshire and Madhani 2018). In this regard, studies in S. pombe highlighted the importance of an adequate balance between reader, writer and eraser factors for H3K9 methylation maintenance and inheritance (Audergon et al. 2015; Ragunathan et al. 2015). Moreover and as previously introduced, the activity of enzymatic writers of histone modifications is counterbalance by eraser proteins. The jumonji C (JmjC) domain-containing demethylase families, with the JMJD2 (also known as KDM4) family displaying activity towards H3K9me2 and H3K9me3 residues and JMJD1 (also referred to as KDM3) proteins displaying activity toward H3K9me1 and H3K9me2 (Cloos et al. 2008; Nottke et al. 2009). To reinforce even more the exquisite regulation that H3K9me3 has, Danny Reinberg and Lynne Vales proposed that indeed repressive marks represent a true epigenetic mechanism where changes are inherited, in opposition to

transcription activating modifications, which require the continuous presence of initiation factors to establish and maintain active states (Reinberg and Vales 2018).

To ensure H3K9me3 deposition at appropriate genomic regions avoiding ectopic silencing, additional factors are required to recruit the above mentioned silencing factors since neither H3K9 histone methyltransferases nor HP1 proteins can recognize or bind specific DNA sequences. DNA-binding proteins such as members of the large vertebrate specific family of Krüppel-associated box (KRAB)-containing zinc-finger proteins (KRAB-ZFPs) are involved in the recruitment of the silencing machinery. The majority of KRAB-ZFPs target sites are located within transposable elements, and specific KRAB-ZFPs have been shown to induce SETDB1/H3K9me3-dependent silencing at endogenous retrovirus targets (Turelli et al. 2014; Collins et al. 2015; Fasching et al. 2015; Wolf et al. 2015; Ecco et al. 2016; Kato et al. 2018; Tie et al. 2018). Several studies have demonstrated that the KRAB domain interacts with the co-repressor KAP1, which in turn recruits SETDB1 (Friedman et al. 1996; Peng et al. 2000; Schultz et al. 2002; Thompson et al. 2015). There are also studies implicating the retinoblastoma (RB) protein and the myc-associated factor X (MAX) to SUV39H1- and SETDB1-mediated silencing of specific genes in mammals (Nielsen et al. 2001; Maeda et al. 2013; Tatsumi et al. 2018). The regulation of silent chromatin can be even more complex since it has been reported the crosstalk between different silencing

pathways. For instance, H3K9-related factors can interact and share some targets with other chromatin modifiers, such as the DNA methyltransferase 1 (DNMT1) and the nucleosome remodeling and deacetylase (NuRD) complex (Robertson et al. 2000; Schultz et al. 2001; Lehnertz et al. 2003; Uchimura et al. 2006; Ivanov et al. 2007; Tatsumi et al. 2018). It should be noted that ATP-dependent chromatin remodelers shape the chromatin landscape by determining the composition, position, and density of nucleosomes, thus regulating the chromatin access for other proteins (Narlikar et al. 2013; Hota and Bruneau 2016; Clapier et al. 2017).

Importantly, small RNA-based targeting mechanisms have also been identified and proposed to recruit H3K9 histone methyltransferases to chromatin, studied mainly in S. pombe (Hall et al. 2002; Volpe et al. 2002), *Caenorhabditis elegans* (C. elegans), plants (Zilberman et al. 2003; Ashe et al. 2012; Chalker et al. 2013) and Drosophila (Gu and Elgin 2013; Akkouche et al. 2017).

In this Chapter, we have revised and discussed how the mammalian genome is organized inside the cell nucleus and some of the principles that dictate how gene expression is regulated in different contexts. For the purpose of the work presented in this Thesis, we have centered our attention on histone modifications and, in particular, in how certain histone modifications (e.g. H3K9me3) lead to transcriptional silencing and repression.

AIMS OF THE STUDY

Aims of the study

Chromocenters are established after the 2-cell stage during mouse embryonic development, but the factors that mediate chromocenter formation remain largely unknown. Previous reports focused either on the *in vivo* scenario, where limited manipulations can be performed, or exclusively on the transcriptional output, making the *in vitro* model necessary to overcome the scarcity of material available in early embryos and expand the array of experimental approaches.

Therefore, the present PhD Thesis proposed to address the following objectives:

1. Profile the chromatin-bound proteome during the 2C-like cell reprograming.
2. Characterize the remodeling of the repressive H3K9me3-heterochromatin foci during the 2C-like cell reprograming.
3. Identify and characterize factors controlling the remodeling of chromocenters in 2C-like cells and early mouse embryos.

PART II - RESULTS

2C-like cells are a suitable model system to investigate key features of early embryonic development

To explore the molecular events underlying the establishment of constitutive heterochromatin during embryo development, we generated stable ESC lines carrying doxycycline-inducible cassettes that drive the expression of either the cleavage-specific homeodomain transcription factor Dux (Dux-codon altered, CA) or luciferase as control (Figure 11a).

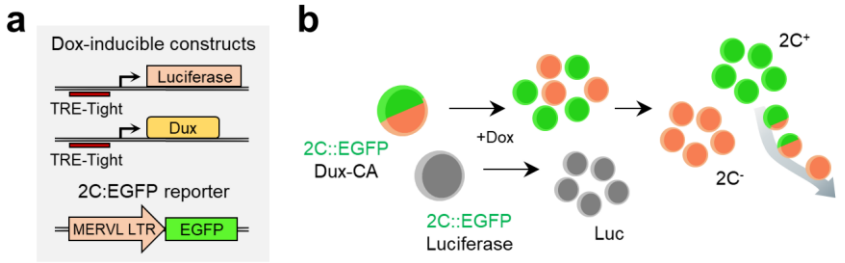

Figure 11. Modeling early development with 2C-like cells. a, Scheme of the constructs used to generate stable ESC lines for the induction and identification of 2C-like cells arising from ESC cultures. Doxycycline-inducible constructs containing a tetracycline responsive element (TRE) tight promoter were used for ectopically expressing luciferase or Dux transgenes (upper). An EGFP reporter under the control of the endogenous retroviral element MERVL long terminal repeat (2C::EGFP) was used to monitor MERVL-expressing 2C-like cells (lower). **b**, Schematic representation of the 2C-like state conversion system and the cellular states generated after Dux overexpression ($2C^-$ cells, which correspond to the GFP negative population, and $2C^+$ cells, corresponding to GFP positive 2C-like cells). 2C::EGFP, EGFP reporter under the control of the MERVL LTRs. Dux-CA, tet-inducible mouse Dux-codon altered (CA) sequence. Dox, doxycycline. Luc, luciferase.

These ESC lines also carry an EGFP reporter under the control of the endogenous retroviral element MERVL long terminal repeat (2C::EGFP) (Ishiuchi et al. 2015) to identify and purify the distinct cellular states arising in the course of the Dux-mediated $2C^+$ cell fate conversion process (Figure 11a). The EGFP reporter allowed the identification of *bona fide* 2C-like cells (hereinafter named $2C^+$), as those with expression of the 2C::EGFP reporter, and cells expressing low levels of endogenous Dux and that are negative for MERVL reporter expression ($2C^-$ cells) (Figure 11b). $2C^-$ cells have been recently postulated to be an intermediate population generated during the 2C-like reprogramming (Rodriguez-Terrones et al. 2018; Fu et al. 2019). After purification and enrichment steps, we studied the ability of $2C^+$ cells to revert to an ESC-like state and the accompanied chromatin rearrangements associated with H3K9me3 heterochromatin (Figure 11b).

During the $2C^+$ reprogramming process, in contrast to $2C^-$ cells, $2C^+$ cells do not show DAPI-dense chromocenters (Ishiuchi et al. 2015; Hendrickson et al. 2017). Therefore, we can study *de novo* chromocenter formation by following the transition of $2C^+$ cells toward an ESC-like state, thus modelling in culture the epigenetic reprogramming that occurs during mouse early development.

To validate the efficiency of the cell lines we generated, we first tested if Dux overexpression was sufficient to express detectable levels of the 2C::EGFP reporter. After culturing the

Dux-CA line with doxycycline (Dox), the number of 2C-like cells increased as compared to luciferase control cells (Figures 12a-d). We observed individual 2C$^+$ cells appearing in ESC cultures exclusively in the Dux-CA line after Dox administration (Figure 12a). Indeed, we confirmed the expression of the detected GFP signal by immunostaining and flow cytometry (Figures 12b, 12c).

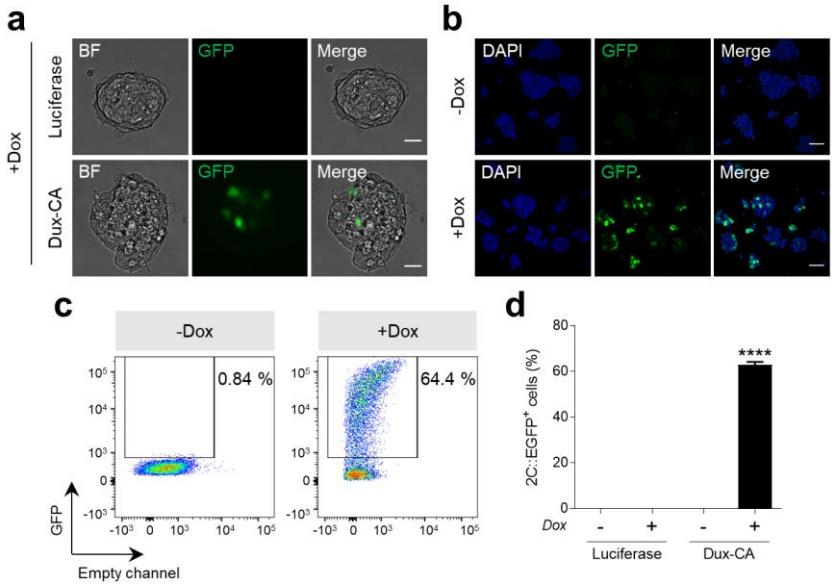

Figure 12. Dux overexpression in ESCs is sufficient to generate 2C-like cells. a, Representative live cell images of the stable ESC lines expressing luciferase or Dux-CA upon doxycycline (Dox) induction. Scale bar, 20 µm. **b**, Representative immunofluorescence images of the 2C::EGFP reporter in control and Dux overexpressing ESC lines showing activation of the 2C::EGFP reporter after 24 hours of Dox administration. Scale bar, 50 µm. **c**, Representative FACS plots showing GFP$^+$ cells in the Dux-CA line without (-Dox) and after 24 hours of Dox treatment (+Dox). **d**, Effect of Dux overexpression on the activation of the 2C::EGFP reporter by flow cytometry. Shown are the mean ± SD (n = 3 independent cultures). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test).

Flow cytometry analysis allowed us to accurately determine that the increase of $2C^+$ cells was exclusively observed in the Dux overexpressing condition (Figures 12c, 12d). Notably, the Dux-CA line achieved > 60 % of $2C^+$ induction (Figure 12d).

The observed induction of $2C^+$ cells was in line with previous reports (Hendrickson et al. 2017). Importantly, the achieved $2C^+$ induction implied a significant improvement, especially if compared with the < 1 % of $2C^+$ cells detected in spontaneous conditions in pluripotent cultures. This fact is also relevant since it opens up the range of downstream experimental procedures that we can carry out using $2C^+$ and $2C^-$ cells.

Next, we decided to validate additional reported features of 2C-like cells in the Dux-CA line. Dux overexpression resulted in loss of DAPI-dense chromocenters and loss of the pluripotency transcription factor OCT4 (Figure 13a), in accordance with previous reports (Macfarlan et al. 2012; Ishiuchi et al. 2015). Of note, even mild 2C::EGFP reporter expression resulted in complete absence of OCT4 at the protein level (Figure 13a). These changes were accompanied by an upregulation of specific genes of the 2-cell transcriptional program such as endogenous Dux, MERVL and major satellites (MajSat) in $2C^+$ cells (Figure 13b). $2C^-$ cells showed higher expression levels of the abovementioned 2-cell genes as compared to ESCs although they did not express the 2C::EGFP reporter (Figure 13b). Overall, these data indicate that the Dux-CA line recapitulates known features of 2C-like cells.

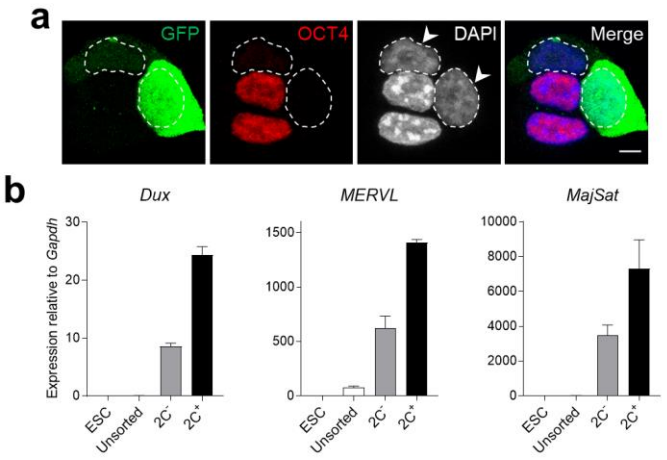

Figure 13. 2C⁺ cells lose OCT4 and express 2-cell genes. a, Representative immunofluorescence images of the 2C::EGFP reporter and the endogenous pluripotency transcription factor OCT4 after 24 hours of Dox induction. Arrowheads indicate 2C⁺ cells. Dashed lines indicate nuclei contour. Scale bar, 5 µm. **b**, qRT-PCR of *Dux*, MERVL and major satellites (MajSat) in WT ESCs, unsorted, 2C⁻ and 2C⁺ sorted cells. Data are presented as mean ± SE (n ≥ 2, 2C⁻ and 2C⁺ samples are technical replicates).

2C-like cells generated via Dux overexpression predominantly contribute to extraembryonic tissues *in vivo*

Having characterized the Dux-CA line, we then assessed the *in vivo* potency of the 2C-like cells generated via Dux overexpression during both pre- and post-implantation development (Figure 14a). We generated chimeric embryos microinjecting either 2C⁺ or luciferase cells to morula-stage (E2.5) embryos to follow their contribution. To track the injected cells and their progeny in the recipient mouse embryos, we

constitutively expressed the DsRed protein via the piggyBac (pB) transposon system in the luciferase and the Dux-CA ESC lines (Figures 14b, 14c).

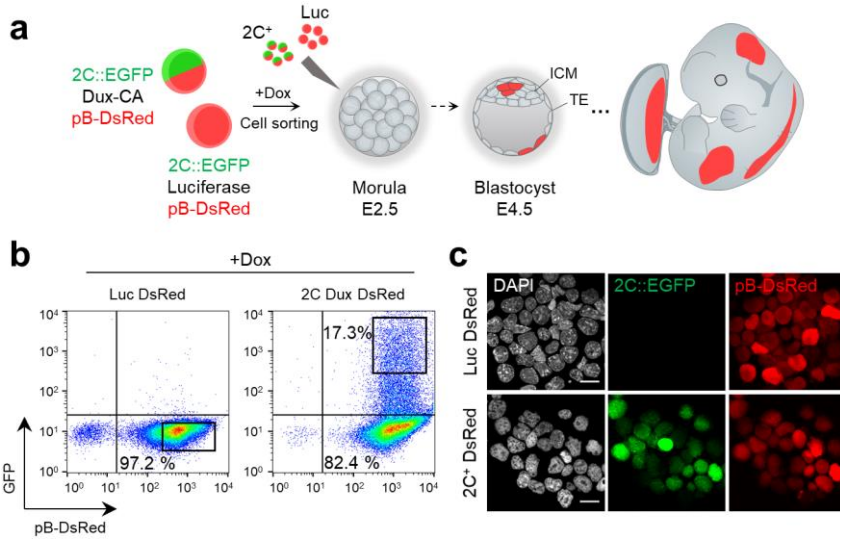

Figure 14. Assessing the potency of 2C⁺ cells *in vivo*. a, Schematic representation of the experimental design to test cellular potency *in vivo* in pre- (E4.5) and post-implantation (E12.5) mouse embryos. pB-DsRed, piggyBac-based vector containing a DsRed protein. ICM, inner cell mass. TE, trophectoderm. **b**, Representative FACS plots showing the flow cytometry gates used for FACS sorting luciferase DsRed⁺ and 2C⁺ DsRed⁺ (GFP⁺ DsRed⁺) cells. **c**, Representative fluorescent images of sorted luciferase (Luc DsRed) and Dux overexpressing (2C⁺ DsRed) DsRed-tagged ESC lines showing the activation of the 2C::EGFP reporter exclusively in the 2C⁺ DsRed line and expression of the piggyBac-DsRed (pB-DsRed) fluorescent protein prior to embryo injection. Scale bar, 20 µm.

Either Luc DsRed or 2C⁺ DsRed freshly sorted cells were injected into morula-stage embryos to follow their contribution to E4.5 and to E12.5 mouse embryos (Figure 14b). We imaged sorted cells to ensure that the 2C⁺ cells displayed chromocenter decompaction as well as 2C::EGFP and pB-

DsRed expression prior to embryo injection (Figure 14c). As expected, Luc DsRed cells had compacted chromocenters and were negative for 2C::EGFP expression (Figure 14c).

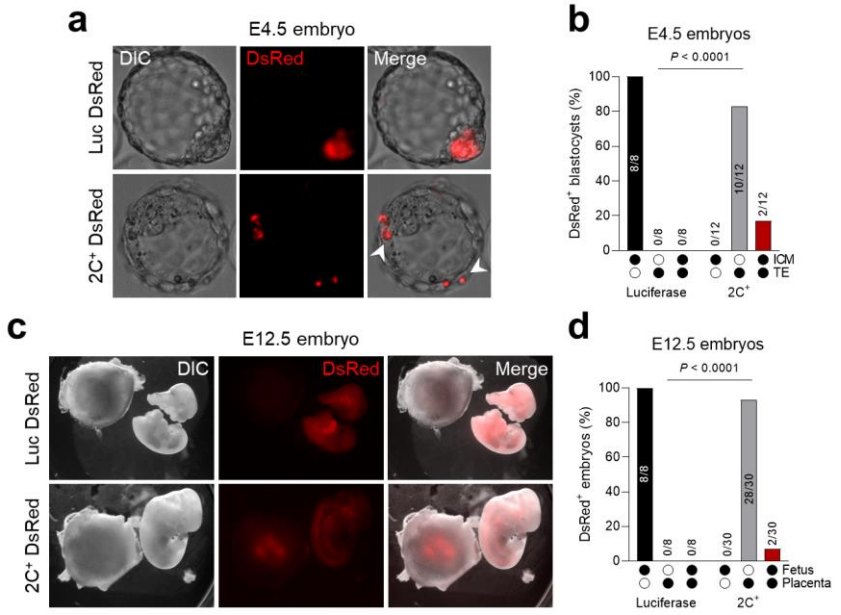

Figure 15. 2C⁺ cells predominantly contribute to extraembryonic tissues *in vivo*. **a**, Representative live chimeric E4.5 blastocysts injected with luciferase DsRed (Luc DsRed) and 2C⁺ DsRed cells. Arrowheads indicate DsRed⁺ cells located in the ICM and TE. DIC, differential interference contrast. **b**, Quantification of the frequency of DsRed⁺ blastocysts with signal localized in ICM, TE or both ICM/TE. $P < 0.0001$ by Fisher's exact test (Luc DsRed group = 8 embryos; 2C⁺ DsRed group = 12 embryos). **c**, Representative live chimeric E12.5 embryos from luciferase DsRed (Luc DsRed) or 2C⁺ DsRed injected blastocysts. DIC, differential interference contrast. **d**, Quantification of the frequency of DsRed⁺ embryos with DsRed contribution in fetus, placenta or both fetus/placenta. $P < 0.0001$ by Fisher's exact test (Luc DsRed group = 8 embryos; 2C+ DsRed group = 30 embryos).

When we analyzed E4.5 embryos, we observed that Luc DsRed cells exclusively contributed to the inner cell mass (ICM) of the blastocysts as expected for pluripotent cells

(Figures 15a, 15b). In contrast, $2C^+$ DsRed cells contributed predominantly to the trophectoderm (TE) and, in a minority of the analyzed blastocysts (16.7 %), to both ICM and TE (Figures 15a, 15b). We then tested whether Luc or $2C^+$ cells could engraft and contribute to the embryo and/or extraembryonic tissues, such as the placenta, at later developmental stages. In principle, this is a more stringent assay to rule out the possibility that those $2C^+$ cells that contributed to the TE were compromised and, therefore, not able to expand in advanced stages.

We transferred either Luc or $2C^+$ cell-injected blastocysts to foster mice and analyzed embryos at E12.5. Luc DsRed cells contributed exclusively to E12.5 fetuses as expected, whereas the large majority of $2C^+$ DsRed cells contributed to the placenta (Figures 15c, 15d). Of note, in some cases, $2C^+$ cells also contributed to fetuses, although with lower frequencies (6.7 %) (Figures 15c, 15d).

For chimeric blastocysts, we performed immunostaining against the DsRed protein to ensure specificity followed by 3D image reconstruction. We could detect a clear cluster of Luc DsRed cells exclusively contributing to the ICM of the blastocysts, as observed in live embryos (Figure 16a). $2C^+$ DsRed cells provided a milder DsRed immunofluorescent signal overall compared to Luc DsRed cells, although it was predominantly located in the TE, in accordance to live embryo visualizations (Figure 16b).

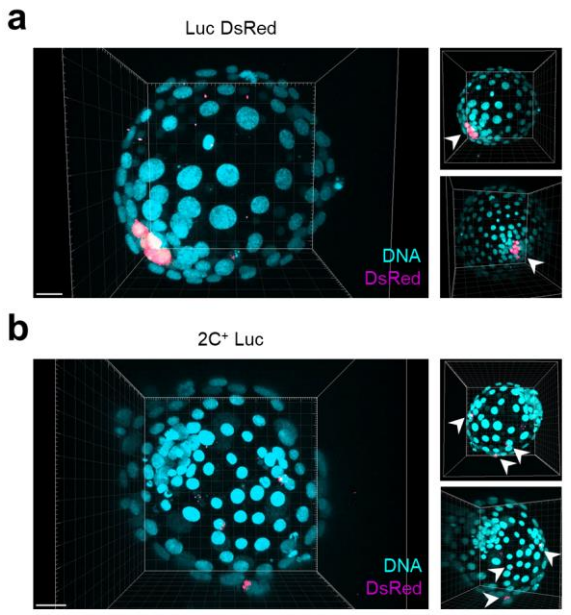

Figure 16. 3D reconstructions of chimeric blastocyst-stage embryos. a, Representative snapshot from a 3D reconstructed image of a Luc DsRed-injected E4.5 embryo. Lateral views are shown on the right. Arrowheads indicate cells expressing DsRed. Scale bar, 20 µm. **b**, Representative snapshot from a 3D reconstructed image of a $2C^+$ DsRed-injected E4.5 embryo. Lateral views are shown on the right. Arrowheads indicate cells expressing DsRed. Scale bar, 20 µm.

In order to ensure that the DsRed signal detected in live chimeric E12.5 embryos was specific, we performed immunostainings against the DsRed protein in fetal and placental sections (Figures 17a, 17b). With these stainings we could also determine whether or not the viability of those $2C^+$ DsRed cells located in the TE was compromised, and thus, whether their progeny could engraft and expand in the mentioned tissues. Indeed, we observed DsRed staining in chimeric fetus containing Luc DsRed and $2C^+$ DsRed cells (Figures 17a, 17c).

Of note, these 2C+ DsRed chimeric fetuses were selected among the ones that were already presenting high DsRed

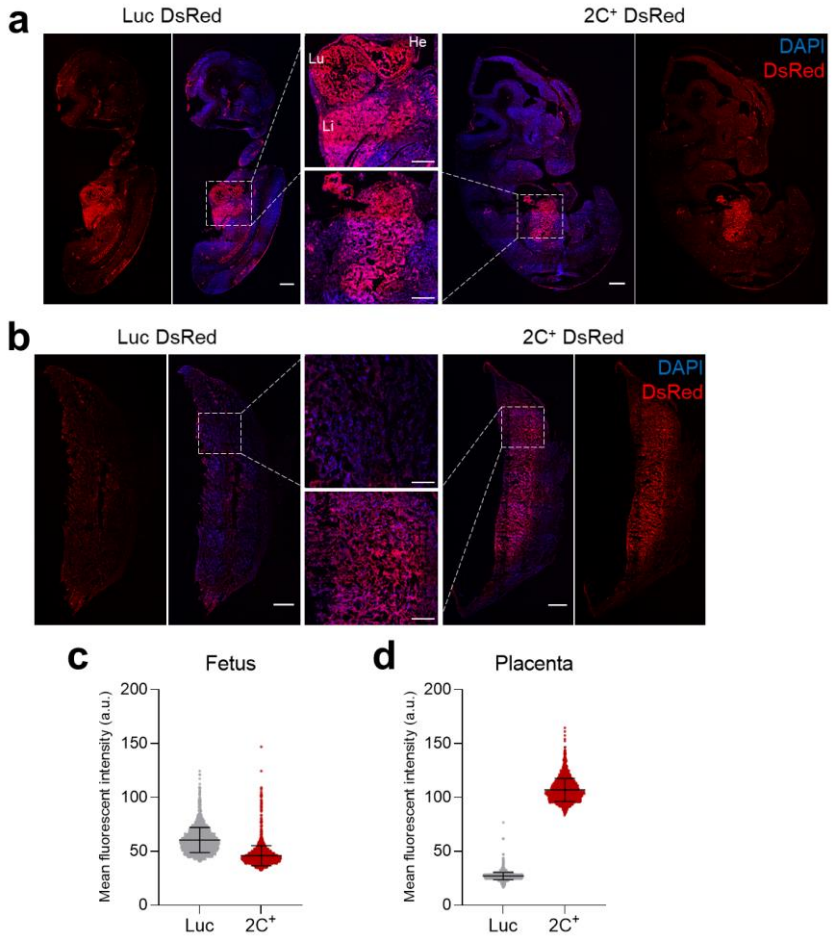

Figure 17. 2C+ DsRed cells engraft in fetal and placental tissues. a, Mosaic immunofluorescence images of DsRed in whole fetuses from Luc DsRed and 2C+ DsRed chimeric E12.5 embryos. Embryonic heart (He), lung (Lu) and liver (Li) were highlighted in the zoomed image. Scale bar, 500 µm. Zoomed images, 200 µm. **b**, Mosaic immunofluorescence images of DsRed in whole placentas from Luc DsRed and 2C+ DsRed chimeric E12.5 embryos. Scale bar, 500 µm. Zoomed images, 200 µm. **c**, Quantification of DsRed mean fluorescent intensity in fetal samples from Luc DsRed and 2C+ DsRed chimeric E12.5 embryos. **d**, Quantification of DsRed mean fluorescent intensity in placental samples from Luc DsRed and 2C+ DsRed chimeric E12.5 embryos.

signal in live conditions. Notably, the progeny of the injected DsRed-tagged cells contributed to the fetal heart, lung and liver, which are highly vascularized organs (Figure 17a). In placental sections, DsRed signal was exclusively detected in $2C^+$ DsRed injected embryos, staining a large proportion of the E12.5 placenta (Figures 17b, 17d).

Given that we observed a large proportion of the placenta stained by DsRed thus suggesting that $2C^+$ DsRed cells had expanded after TE contribution, we decided to further investigate the acquired fate of the $2C^+$ DsRed progeny. We first observed that a fraction of the area stained with DsRed in the placenta corresponded to the trophoblast-specific protein alpha (TPBPA) positive region (Figure 18a). TPBPA is a marker of spongiotrophoblasts, which are placental cells that act as structural support for the developing villous of the labyrinth (Simmons and Cross 2005). Unfortunately, we could not assess co-localization between TPBPA and DsRed due to antibody species incompatibilities. We then checked the expression pattern of proliferin, which is a placental marker of trophoblast giant cells. Trophoblast giant cells are terminally differentiated, polyploidy cells that mediate many vital processes in the maternal-fetal crosstalk (Simmons and Cross 2005). $2C^+$ DsRed cells co-expressed DsRed and proliferin, suggesting that some of the progeny of the $2C^+$ DsRed injected cells not only expanded *in vivo* in the placenta but also acquired functional markers of tissue resident populations (Figures 18a, 18b).

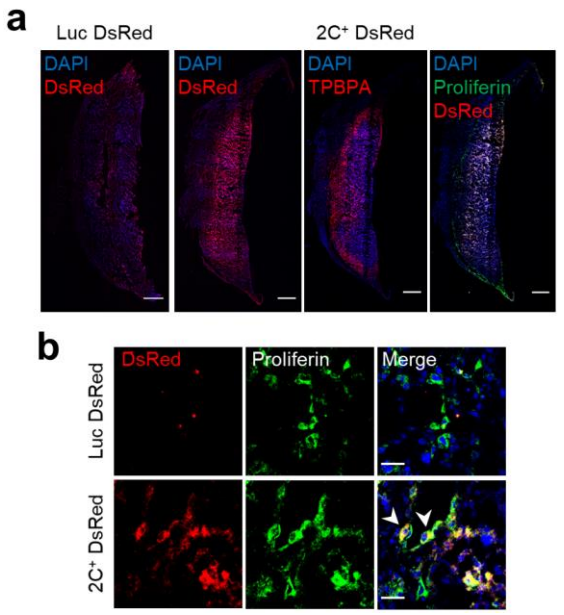

Figure 18. 2C⁺ cells express placental markers. a, Whole chimeric placentas from 2C⁺ DsRed-injected cells stained with trophoblast-specific protein alpha (TPBPA) and co-stained with DsRed and proliferin. Luc DsRed placenta was included as control for DsRed signal. Scale bar, 500 µm. **b**, Chimeric placentas from Luc DsRed or 2C⁺ DsRed-injected cells co-stained with DsRed and proliferin. Arrowheads indicate cells co-expressing DsRed and proliferin. Scale bar, 50 µm.

These results indicate that, 2C⁺ cells can differentiate toward fetal and placental fates *in vivo*, suggesting the acquisition of totipotent-like features (Abad et al. 2013; Yang et al. 2017a; Yang et al. 2017b), although with low frequencies. Noteworthy, 2C⁺ cells showed a high tendency to colonize mostly the placenta, indicating they acquire prevalently a trophoblast stem cell-like fate and a bias toward extraembryonic linages. The latter is especially relevant since 2C⁺ DsRed progeny expressed markers of more differentiated cell types resident in the placenta. It is nevertheless noteworthy that

spongiotrophoblasts and trophoblast giant cells might share some common progenitor, which derived from extraembryonic ectoderm (Simmons and Cross 2005). Yet, some additional experiments must be performed in order to fully comprehend 2C-like cell contribution to extraembryonic lineage differentiation.

Chromatin-bound proteome profiling allows the identification of dynamic chromatome changes during 2C-like cell reprogramming

Next, after having investigated the *in vivo* potency of the Dux-derived 2C-like cells, we aimed to identify potential chromatin-associated factors involved in *de novo* establishment of heterochromatin. We performed DNA-mediated chromatin pull-down (Dm-ChP) coupled to tandem mass spectrometry (MS/MS) (Aranda et al. 2019; Aranda et al. 2020). We captured the whole genome labelled with 5-ethynyl-2'-deoxyuridine (EdU) and identified candidate proteins differentially enriched in the 2C-like chromatin-bound (chromatome) fraction (Figure 19a). Due to the high reprogramming efficiency of the Dux mediated system, we were able to FACS sort approximately 10 million of cells per replicate and condition (Figure 19b).

After performing mass spectrometry, we first analyzed the chromatome of $2C^+$, $2C^-$ and luciferase (Luc) populations to characterize the chromatin-bound proteome profile of these

distinct states. We identified a comparable number of proteins in each independent replicate among the $2C^+$, $2C^-$ and luciferase (Luc) populations, with the only exception of the replicate 2 in the $2C^+$ condition where fewer proteins were

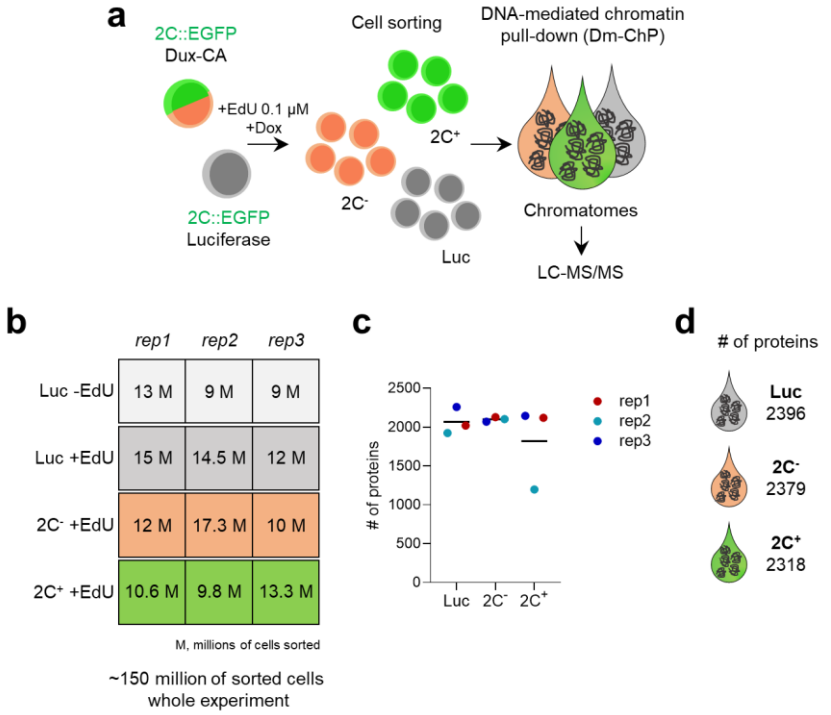

Figure 19. Chromatin-bound proteome profiling of the $2C^+$ reprogramming. a, Schematic representation of the samples collected to perform DNA-mediated chromatin pull-down (Dm-ChP) followed by mass spectrometry workflow. EdU, 5-ethynyl-2'-deoxyuridine. LC-MS/MS, liquid chromatography-tandem mass spectrometry. **b**, Summary of the number of cells sorted per replicate and condition used as inputs for the DNA-mediated chromatin pull-down protocol. Values indicate millions (M) of cells sorted. **c**, Quantification of the number of proteins detected per replicate and condition in mass spectrometry. The number of proteins identified was selected on the basis of a minimum of fold change (FC) > 1 compared with the -EdU condition in protein abundance. **d**, Representation of the average number of proteins identified per condition.

detected, as compared with the –EdU sample (Figure 19c). Overall, the average number of proteins identified was alike for

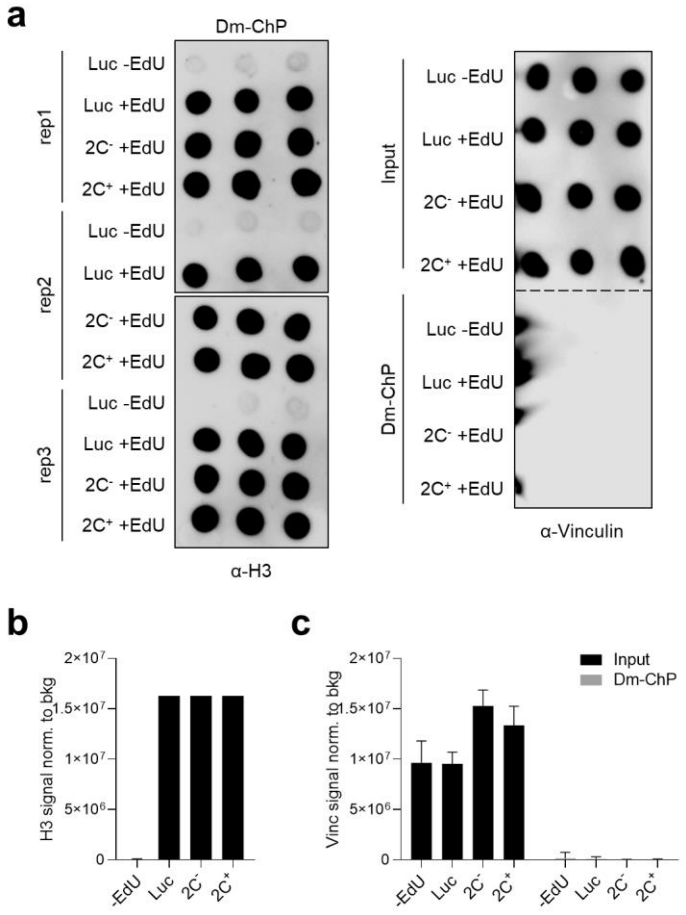

Figure 20. Evaluation of 2C⁺ Dm-ChP samples. a, Independent DNA-mediated chromatin pull-down (Dm-ChP) eluates from sorted luciferase, 2C⁻ and 2C⁺ replicates in the absence or presence of EdU (±EdU) were analyzed by dot blot with an anti-H3 antibody (left). Input and eluates from equivalent preparations were incubated with an anti-vinculin antibody (right). Each condition was spotted in triplicates. **b**, Quantification of histone H3 signal detected by dot blot in the absence or presence of EdU. Data are presented as mean ± SD of H3 signal normalized to the background. **c**, Quantification of vinculin (Vinc) signal detected by dot blot in input or Dm-ChP samples. Data are presented as mean ± SD of vinculin signal normalized to the background.

$2C^+$ (n = 2318), $2C^-$ (n = 2379) and Luc (n = 2396) conditions (Figure 19d).

In fact, we confirmed that we could enrich the Dm-ChP preparations for chromatin proteins, such as histone H3 (Figure 20a, 20b). Importantly, we showed that we could devoid them as well of cytoplasmic markers, such as vinculin, demonstrating that we precisely captured the chromatome fraction (Figure 20a, 20c).

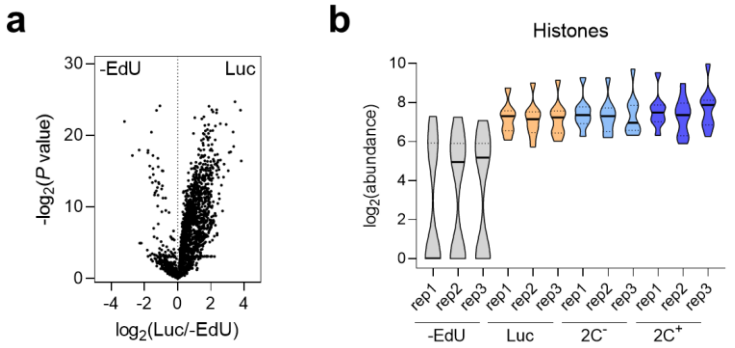

Figure 21. Dm-ChP enrich for chromatin-bound proteins. a, Volcano plot of proteins identified by mass spectrometry after DNA-mediated chromatin pull-down in Luc and -EdU conditions. **b**, Abundance of histones in the individual replicates from -EdU, Luc, $2C^-$ and $2C^+$ conditions. The following histones were included in the analysis: core histones H2A, H2B and H4, macro-H2A.1, macro-H2A.2, H2A.V, H2A.X, H3.3 and CENP-A. Violin plot shows median with a solid line and quartiles with dashed lines.

To better characterize and measure the specificity of the chromatome dataset, we compared the samples containing EdU, where the chromatome was supposed to be captured, with the blank control condition (-EdU). We identified 2396 proteins in the Luc condition as compared with the –EdU control, suggesting efficient pull-down of chromatin-associated

factors (Figure 21a). Particularly, chromatin-resident proteins, such as core histones and histone variants, were comparably enriched in all +EdU replicates but less represented in the –EdU replicates (Figure 21b).

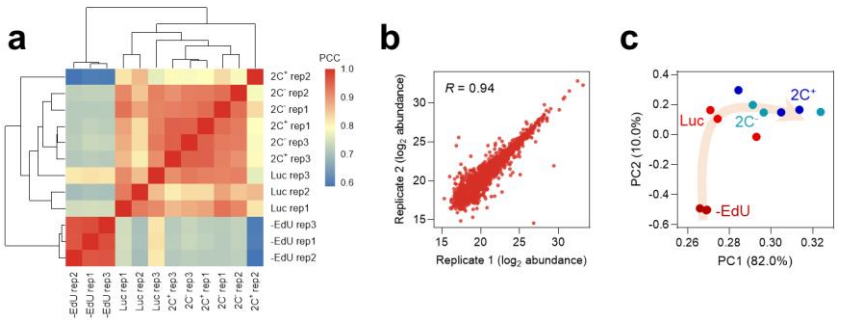

Figure 22. Reproducibility among replicates of the 2C⁺ Dm-ChP dataset. **a**, Correlation matrix showing reproducibility among independent replicates of -EdU, Luc, 2C⁻ and 2C⁺ protein abundances. Hierarchical clustering analysis of the individual replicates is also shown. Cells in the correlation matrix are color-coded based on PCC values. PCC, Pearson's correlation coefficient. **b**, Reproducibility between replicates of independent chromatome preparations. Correlation between replicate 1 and replicate 2 from the 2C⁻ condition is shown. *R* indicates Pearson's R. **c**, Principal component analysis (PCA) of the chromatome dataset. Each point corresponds to a single replicate, which is colored according to the assigned condition. PC1 separates -EdU, Luc, 2C⁻ and 2C⁺ chromatomes, explaining 82 % of the variance. Beige arrow indicates the 2C-like reprogramming trajectory.

To further strengthen the characterization of the Dm-ChP dataset, we computed Pearson's correlation coefficients (PCC) and principal component analysis (PCA) of independent replicates of 2C⁺, 2C⁻ and Luc samples. Both PCC and PCA showed consistent results regarding the abundance of the proteins detected (Figures 22a-c). Interestingly, Luc replicates clustered separately from 2C⁺ and 2C⁻ conditions, indicating significant changes in the chromatomes of these fractions

(Figure 22a). The degree of reproducibility was appreciable within each condition, as exemplified with the replicates 1 and 2 from the 2C$^-$ condition (Figure 22b). Moreover, PC1 component allowed tracing the 2C-like reprogramming trajectory from bulk chromatome abundances (Figure 22c). Altogether, these data illustrate that the Dm-ChP dataset we generated profiling the chromatome of the 2C$^+$, 2C$^-$ and Luc populations is robust and holds potential to discover chromatin-bound factors controlling the remodeling of chromocenters.

We then ranked the identified chromatin-associated factors to interrogate the differences in protein-chromatin interactions in the 2C$^+$, 2C$^-$ and Luc chromatomes (Figures 23a-c). Since we profiled the chromatome of all the cellular states induced along the 2C-like reprogramming including the intermediate 2C$^-$ cells, we first analyzed the differences in a stepwise fashion (Figures 23a, 23b). Members of the ZSCAN4 (Zinc finger and SCAN domain containing 4) family of proteins, that are well-characterized markers of the 2C stage (Falco et al. 2007; Ishiuchi et al. 2015; De Iaco et al. 2017; Hendrickson et al. 2017), were found among the top enriched factors already in the 2C$^-$ chromatome (Figure 23a), supporting previous findings (Rodriguez-Terrones et al. 2018). Then, we identified negative regulators of 2C-like cells such as the DNA hydroxylase TET1, the non-canonical Polycomb (PcG) Repressor Complex 1 (PRC1) member PCGF6, and the TGF-β regulator SMAD7, together with the pluripotency transcription factor SOX2, and

the heterochromatic H4K20me3 methyltransferase SUV420H2 depleted from the $2C^+$ chromatome when compared to the $2C^-$

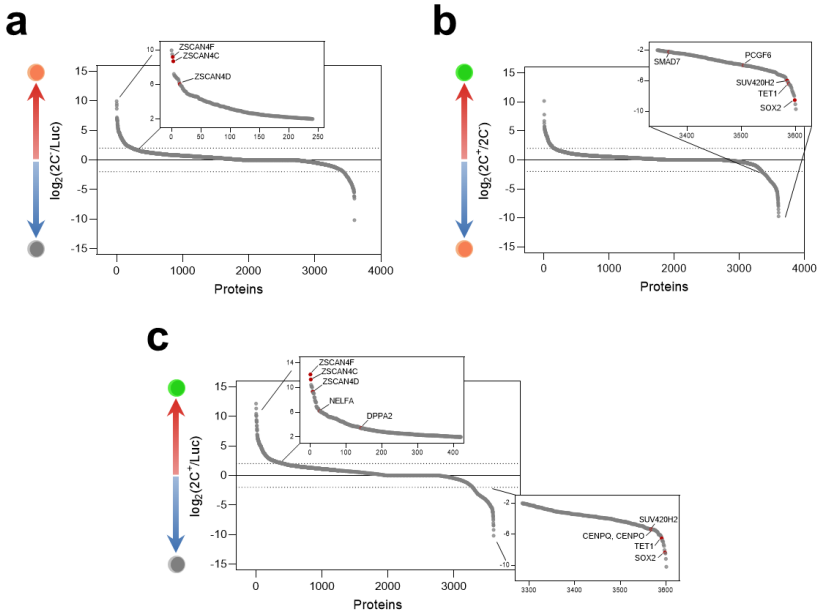

Figure 23. Pairwise enrichment comparisons between states of the $2C^+$ reprogramming. a, Protein enrichment analysis of the comparison between $2C^-$ and Luc chromatomes. The red dots indicate known regulators of the 2C-like state, which were found enriched in the $2C^-$ chromatome (ZSCAN4F: log_2FC = 9.2; ZSCAN4C: log_2FC = 8.7; ZSCAN4D: log_2FC = 6.0). Dashed lines indicate log_2 FC ± 2. **b**, Protein enrichment analysis of the comparison between $2C^+$ and $2C^-$ chromatomes. The red dots indicate novel factors which were found depleted from the $2C^+$ chromatome (SOX2: log_2FC = -8.5; TET1: log_2FC = -6.4; SUV420H2: log_2FC = -6.0; PCGF6: log_2FC = -4.0; SMAD7: log_2FC = -2.2). Dashed lines indicate log_2 FC ± 2. **c**, Protein enrichment analysis of the comparison between $2C^+$ and Luc chromatomes. In the upper panel, the red dots indicate known regulators of the 2C-like state, which were found enriched in the $2C^+$ chromatome (ZSCAN4F: log_2FC = 12.2; ZSCAN4C: log_2FC = 11.3; ZSCAN4D: log_2FC = 9.3; NELFA: log_2FC = 6.2; DPPA2: log_2FC = 3.4). In the lower panel, the red dots indicate novel factors which were found depleted from the $2C^+$ chromatome (SOX2: log_2FC = -8.4; TET1: log_2FC = -6.5; CENPQ: log_2FC = -5.4; SUV420H2: log_2FC = -5.4; CENPO: log_2FC = -5.3). Dashed lines indicate log_2 FC ± 2.

condition (Aloia et al. 2013; Lu et al. 2014; Fu et al. 2019) (Figure 23b).

We also compared the chromatome differences between the two endpoints of the 2C-like reprogramming process, Luc and $2C^+$ cells (Figure 23c). It was evident that ZSCAN4 family members, such as ZSCAN4F and ZSCAN4C, were identified among the top enriched factors in the $2C^+$ chromatome (Figures 23c, 24b). Remarkably, we also identified recently described drivers of the zygotic transcriptional program and, therefore, of the 2C-like state, such as the maternal negative elongation factor A (NELFA) (Hu et al. 2020) and the developmental pluripotency-associated factor 2 (DPPA2) (De Iaco et al. 2019; Eckersley-Maslin et al. 2019; Yan et al. 2019), associated with the $2C^+$ chromatome (Figures 23c, 24b). On the contrary, SOX2, TET1, SUV420H2 and the centromeric proteins CENPQ and CENPO were identified among the top enriched factors in the Luc chromatome (Figure 23c).

Next, we selected a number of markers to expand the benchmarking of the Dm-ChP dataset. The pluripotency transcription factors NANOG, OCT4, STAT3, SOX2 and ESRRB were, as expected, exclusively enriched in the $2C^-$ and Luc chromatomes (Figure 24a). 2-cell-specific factors, including members of the previously mentioned ZSCAN4 cluster, NELFA and DPPA2/4, were enriched in the $2C^-$ chromatome and, to a greater extent, in the $2C^+$ chromatome (Figure 24b).

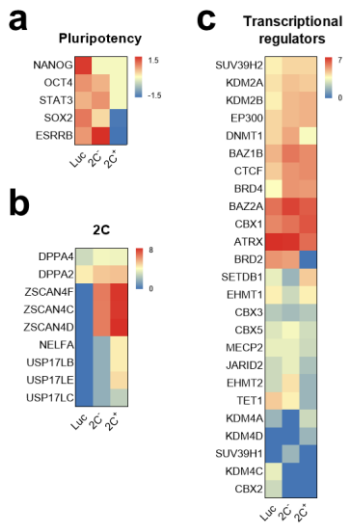

Figure 24. Benchmarking of the 2C⁺ Dm-ChP dataset. a, Heat map representation of the chromatin-bound abundance of pluripotency transcription factors in Luc, 2C⁻ and 2C⁺ cells. Shown are $\log_2$ FC of PSM values to -EdU. Chromatin factors were ordered according to hierarchical clustering. PSM, peptide spectrum match. **b**, Heat map representation of the chromatin-bound abundance of 2-cell-specific factors in Luc, 2C⁻ and 2C⁺ cells. **c**, Heat map representation of the chromatin-bound abundance of transcriptional regulators and epigenetic factors in Luc, 2C⁻ and 2C⁺ cells.

We also identified several epigenetic enzymes differentially enriched in the 2C⁺, 2C⁻ and Luc chromatomes (Figure 24c). For instance, we identified a gradual recruitment to chromatin of the transcriptional regulators EP300, BRD4 and CTCF, among others, when progressing toward 2C⁺ cells (Figure 24c). Conversely, the PRC1 component CBX2, the lysine demethylase KDM4C, TET1, the lysine methyltransferase EHMT2 (also referred to as G9A) and the PRC2 co-factor JARID2 were gradually decreasing their abundance on chromatin (Figure 24c). Altogether, these data indicate that

ESC reprogramming toward 2C-like state correlates with a major reorganization of the chromatin-bound proteome.

We used the Significance Analysis of INTeractome (SAINT) algorithm (Choi et al. 2011) to further interrogate protein-chromatin interactions in the Dm-ChP datasets. To identify molecular drivers of chromocenter reorganization, we compared the enriched proteins in the $2C^+$, $2C^-$ and Luc chromatomes (Figure 25a). We identified 397 proteins shared by the $2C^-$ and Luc chromatomes that were not enriched in the $2C^+$ chromatome (Figure 25a). We decided to focus on analyzing this cluster since chromocenter structure is still established in $2C^-$ and Luc cells. Of note, the largest degree of overlap between chromatomes was observed in similar conditions, that is Luc/$2C^-$ and $2C^-$/$2C^+$ (Figure 25a). No common protein was detected among the mentioned $2C^+$, $2C^-$ and Luc chromatomes (Figure 25a). To identify putative factors responsible for chromocenter reorganization, we ranked the commonly identified proteins included in $2C^-$ and Luc chromatomes according to their fold change (Figure 25b). Notably, this protein cluster included known transcriptional regulators such as the DNA methyltransferase DNMT3L, the bromodomain-containing protein BRD2, the core pluripotency factor OCT4 and the DNA topoisomerase 2-binding protein 1, TOPBP1 (Figure 25b). We focused our attention on TOPBP1, since topoisomerases control genome structure and folding (Wang 2002). These results motivated us to further investigate TOPBP1 network. TOPBP1 has been shown to interact with

chromatin remodelers such as the SWI/SNF-like remodeler SMARCAD1 in yeast and human cells (Liu et al. 2004; Bantele et al. 2017) (Figure 25c).

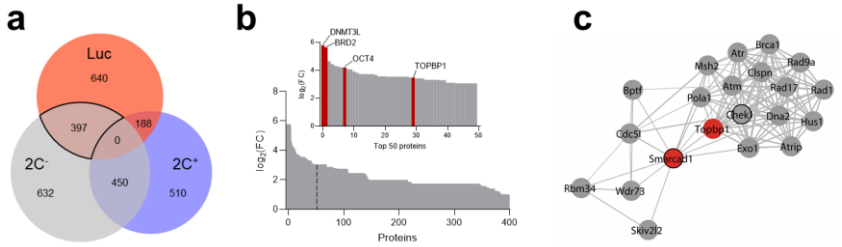

Figure 25. TOPBP1 and SMARCAD1 are leading candidates for controlling chromocenter remodeling. a, Venn diagram indicating the overlap between the identified proteins enriched in Luc, $2C^-$ and $2C^+$ chromatomes after Significance Analysis of INTeractome (SAINT) analysis. Proteins were selected on the basis of a minimum of FC ≥ 2. Solid line highlights the 397 proteins enriched in the Luc and $2C^-$ chromatomes that were not enriched in the $2C^+$ chromatome. **b**, Distribution of the 397 commonly enriched chromatin-bound proteins identified in the Luc and $2C^-$ chromatomes ranked by $\log_2$ FC. Dashed line indicates top 50 protein boundary. In the zoomed panel, several chromatin factors were colored in red including DNMT3L ($\log_2$FC = 5.8), BRD2 ($\log_2$FC = 5.6), OCT4 ($\log_2$FC = 4.2) and TOPBP1 ($\log_2$FC = 3.5). **c**, Functional protein network of TOPBP1 interactors. TOPBP1 and SMARCAD1 nodes are colored in red. Black node border indicates chromatin remodeling function. Network edges indicate the degree of confidence prediction of the interaction. Protein interaction data were retrieved from the STRING database. The network was visualized with Cytoscape.

The referred protein cluster included gene ontology (GO) terms associated with active remodeling activity (e.g. ATPase activity, ATP binding), cell cycle and DNA replication (e.g. replication fork, mitotic cell cycle, DNA repair, DNA replication, cell cycle checkpoint), heterochromatin (e.g. pericentromeric heterochromatin, gene silencing, regulation of H3K9

methylation) and loss of stem cell identity (response to LIF, blastocyst growth) (Figure 26a-c). These terms suggested that there could be an association between heterochromatin regulation and cell cycle progression through an active remodeling process. Indeed, these data reinforced our interest in further exploring the role of the chromatin remodeler SMARCAD1, which is indeed an interactor of TOPBP1, in the 2C-like system (Figure 25c). Besides, these data also prompted us to investigate the putative relationship between cell cycle regulation of the establishment of the 2C-like state.

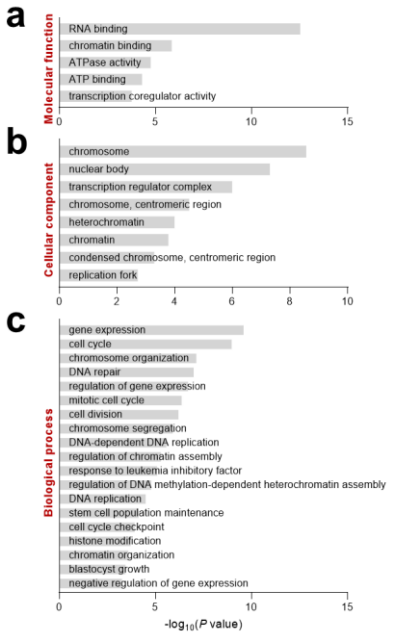

Figure 26. Gene ontology analysis of the Luc and 2C$^-$ chromatomes. a, Gene ontology (GO) of molecular functions of the commonly enriched chromatin-bound proteins in the Luc and 2C$^-$ chromatomes. *P* values are plotted in $-\log_{10}$. **b**, GO of cellular components. **c**, GO of biological processes.

To summarize, we profiled the chromatin-bound proteome during the $2C^+$ reprogramming showing how it is dynamically changing along the 2C-like cell fate conversion. Ultimately, we identified TOPBP1 and SMARCAD1 as potential candidate factors controlling the remodeling of chromocenters.

Entry in the 2C-like state is characterized by the remodeling of H3K9me3 heterochromatic regions

Having investigated the chromatome landscape changes during the $2C^+$ reprogramming, we next asked about the reorganization of heterochromatic regions upon reprogramming of ESCs into $2C^+$ cells. H3K9me3 is a well-known pericentric heterochromatin histone modification that prominently associates with constitutive heterochromatin (Rea et al. 2000; Nakayama et al. 2001; Peters et al. 2001; Peters et al. 2003). H3K9me3 can therefore be used as an accurate marker for chromocenters. H3K9me3 foci in $2C^+$ cells were morphologically distinct from those of $2C^-$ cells (Figure 27a). They were 2.3-fold fewer (3.89 ± 0.19 foci/nucleus) (Figures 27a, 27b), and occupied 2.4-fold larger area (4.76 ± 0.33 μm^2) in $2C^+$ as compared with both ESCs (8.88 ± 0.30 foci/nucleus; 1.99 ± 0.07 μm^2) and $2C^-$ cells (8.72 ± 0.25 foci/nucleus; 1.93 ± 0.08 μm^2) (Figures 27a, 27c). These results suggest that H3K9me3 heterochromatin undergoes massive spatial reorganization, during the reprogramming of ESCs into 2C-like

state. The increased size of the H3K9me3 foci and the reduction in the number of H3K9me3 foci per nucleus might be due to the decompaction or fusion of several chromocenters.

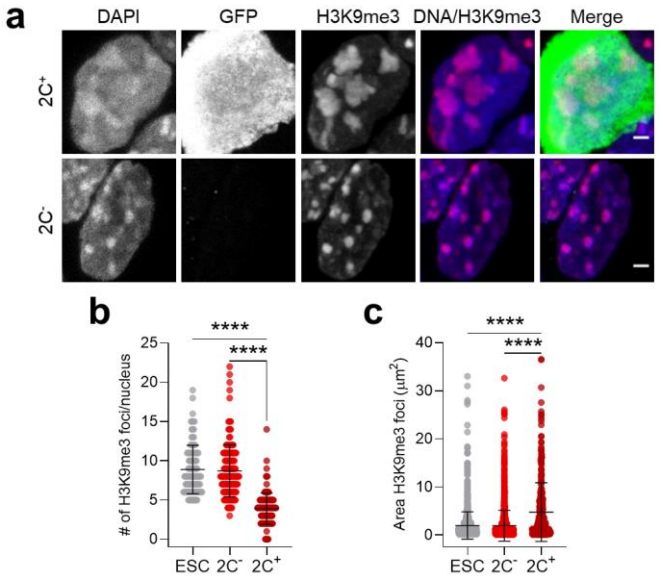

Figure 27. 2C$^+$ cells remodel H3K9me3 heterochromatin. a, Representative immunofluorescence images of the 2C::EGFP reporter and H3K9me3 in 2C$^-$ and 2C$^+$ cells. Scale bar, 2 µm. **b**, Quantification of the number of H3K9me3 foci in ESCs, 2C$^-$ and 2C$^+$ cells. Shown are scatter dot plots with line at mean ± SD (n > 3 independent cultures, ESCs = 103 cells, 2C$^-$ = 170 cells and 2C$^+$ = 119 cells). *P* < 0.0001**** by one-way ANOVA (Tukey's multiple comparisons test). **c**, Quantification of H3K9me3 foci area in ESCs, 2C$^-$ and 2C$^+$ cells. Shown are scatter dot plots with line at mean ± SD (n > 3 independent cultures, ESCs = 1712 foci, 2C$^-$ = 1445 foci and 2C$^+$ = 340 foci). *P* < 0.0001**** by one-way ANOVA (Tukey's multiple comparisons test).

We then imaged global DNA organization with Stochastic Optical Reconstruction super-resolution Microscopy (STORM). DNA was labeled using the nucleotide analog 5-ethynyl-2'-deoxycytidine (EdC) (Zessin et al. 2012; Otterstrom et al. 2019) (Figure 28a). EdC is incorporated into newly replicated DNA.

Then, through the copper-catalyzed Click chemistry reaction, the fluorophore AF647 is covalently bound to the EdC nucleotide, which allows STORM imaging. Super-resolution DNA images were later quantified by Voronoi tessellation analysis (Levet et al. 2015; Andronov et al. 2016) which can precisely determine the DNA density based on the number of localizations in each Voronoi tessel (Figure 28a). In brief, Voronoi tessellation allows partitioning of a given biological structure (e.g. the nucleus) based on its geometrical organization. Voronoi tessellation requires the so-called seeds to distribute the space, making it attractive for super-resolution microscopy applications since their output represents the coordinates of the single molecules that were detected in the sample. For this purpose, Voronoi tessels would serve as a proxy of the nuclear compaction status, where the larger the tessels are the lower the local density will be. To identify the proper labeling conditions and EdC concentration, we pulsed ESCs with increasing EdC concentrations. After performing the Click chemistry reaction, EdC-AF647 fluorescent signal was exclusively detected inside the nucleus (Figure 28b). The EdC-AF647 fluorescent intensity increased in a concentration dependent manner, suggesting that the detected signal was specific (Figures 28b, 28c). Of note, 5 μM of EdC reached almost saturation levels thereby we decided to proceed with 2.5 μM of EdC for further experiments (Figures 28b, 28c). We next pulsed ESCs for different periods of time to assess the appropriate labelling conditions. We imaged them using

STORM microscopy without observing appreciable differences between the time-points analyzed (Figure 28d).

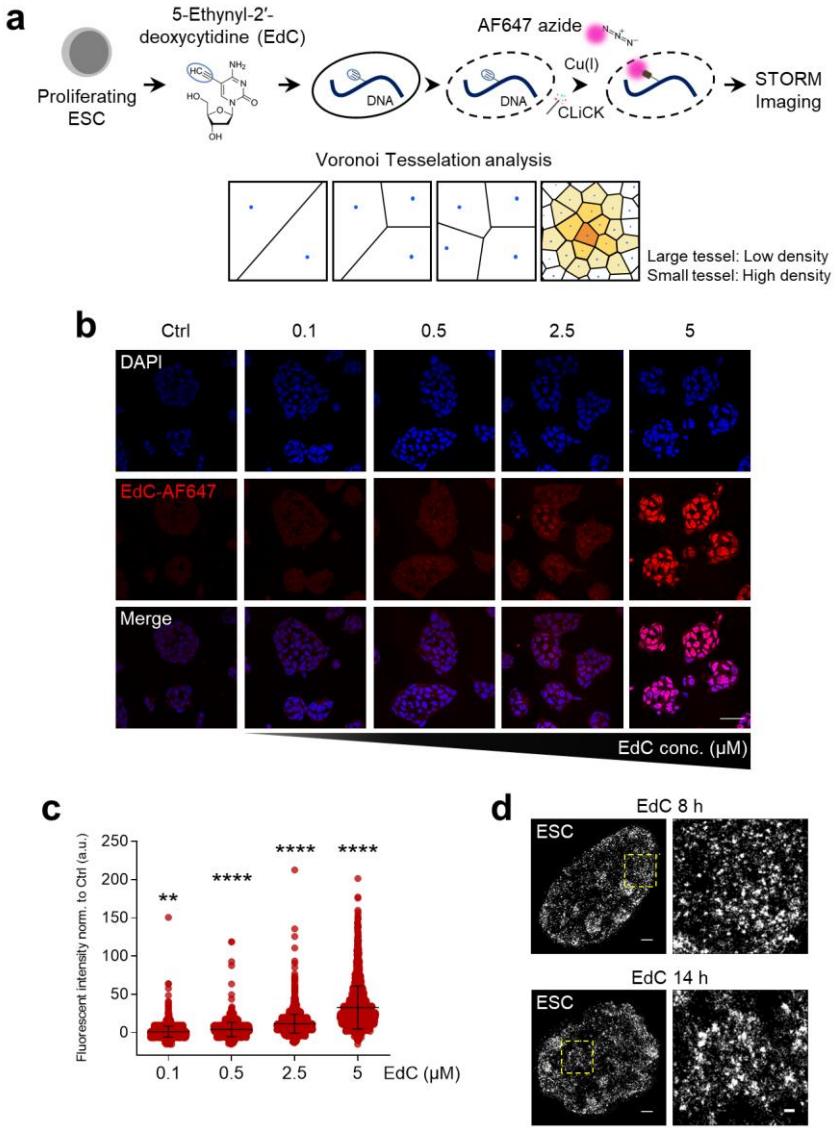

Figure 28. Super-resolution imaging of DNA in ESCs. a, Schematic representation of the experimental procedure for direct DNA labeling via EdC incorporation and Click chemistry reaction. EdC, 5-ethynyl-2'-deoxycytidine. Cu(I), copper(I). AF647, Alexa Fluor 647 dye. Simplistic view of the principles behind Voronoi Tesselation analysis. Blue dots represent single-molecule localizations.

Figure 28 (cont.). b, Representative images of direct DNA labeling via EdC incorporation and Click chemistry reaction in ESCs incubated with increasing EdC concentrations. ESCs grown in the absence of EdC were included as control for EdC-AF647 signal. EdC concentration values are in µM range. Scale bar, 50 µm. **c**, Quantification of EdC-AF647 fluorescent intensity in ESC samples incubated with increasing EdC concentrations. Fluorescent intensity values were normalized to the average control (0 µM) condition. Shown are scatter dot plots with line at mean ± SD (0.1 µM = 3705 cells, 0.5 µM = 1929 cells, 2.5 µM = 3327 cells and 5 µM = 1875 cells). $P = 0.0081$**, $P < 0.0001$**** by one-way ANOVA (Dunnett's multiple comparisons test). **d**, Representative STORM images of direct DNA labeling in ESCs with 8 hours (upper) and 14 hours (lower) EdC incubation time. Dashed square indicates zoomed region shown on the right. Scale bar, 1 µm. Zoomed images, 200 nm.

Next, we proceeded to apply the direct DNA labeling via EdC incorporation to image nuclear compaction in $2C^+$ and $2C^-$ cells. Voronoi analysis showed a marked decrease in localization density of the chromatin in $2C^+$ cells (Figure 29a). Furthermore, Voronoi analysis confirmed the decreased DNA density as a function of the GFP intensity in $2C^+$ cells (Figure 29b). Interestingly, $2C^-$ cells were heterogeneous with respect to DNA density, with the majority of them showing low DNA density as compared with $2C^+$ cells, suggesting that DNA might undergo decompaction prior to GFP activation (Figure 29b).

Overall, the DNA decompaction of the chromatin fibers in $2C^+$ cells is in agreement with the chromatin landscape of early/late 2-cell embryos, which was reported to be more accessible by Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) (Wu et al. 2016a; Hendrickson et al. 2017).

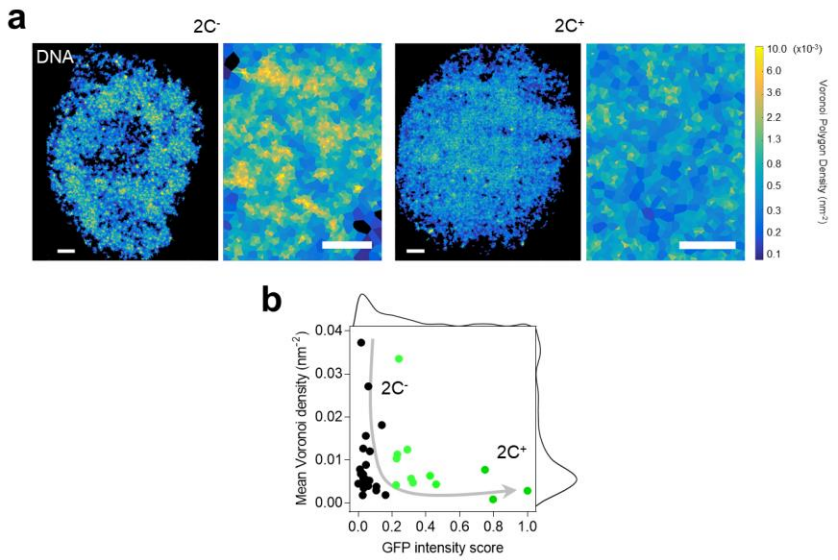

Figure 29. Super-resolution imaging reveals DNA decompaction in 2C⁺ cells. a, Voronoi tessellation rendering of super-resolution images of DNA in 2C⁻ and 2C⁺ cells. Full nuclei (left; scale bar, 1 µm) and zoomed images (right; scale bar, 400 nm) are shown. Voronoi polygons are color-coded based on their area, from yellow for the smallest polygons (density > 0.01 nm^{-2}) to blue for larger polygons (density < 538 0.0001 nm^{-2}). The largest 0.5 % of polygons were set to black. **b**, Biaxial density plot showing mean Voronoi density of DNA (inverse of the polygon area) as a measure of chromatin compaction and GFP intensity score in 2C⁻ and 2C⁺ cells. Cells with a GFP intensity score > 0.2 are colored in green. Black dots indicate 2C⁻ cells and green dots indicate 2C⁺ cells. Each dot represents a single-cell (2C⁻ = 23 cells and 2C⁺ = 12 cells). Histograms represent the distribution of all the cells along the corresponding feature (mean Voronoi density or GFP intensity score). Grey arrow indicates the 2C-like reprogramming trajectory.

To obtain a more dynamic view of the changes in chromatin compaction as ESCs process to the 2C-like state, we performed long-term live cell imaging. The Dux-CA line was engineered to transiently express a TALE-mClover reporter targeting mouse major satellites (TALE_MajSat) (Figure 30a). We observed that those ESCs that were successfully

reprogrammed to 2C$^+$ cells started to decompact their chromocenters as early as 6 hours after Dox induction (Figure 30b). Remarkably, this DNA decompaction event occurred right after a cell division round, suggesting again a relationship between cell cycle progression and induction of 2C-like cells (Figure 30b). It could also be that the hypothesized relationship links a failure in DNA replication, which translates in defective heterochromatin establishment, or vice versa.

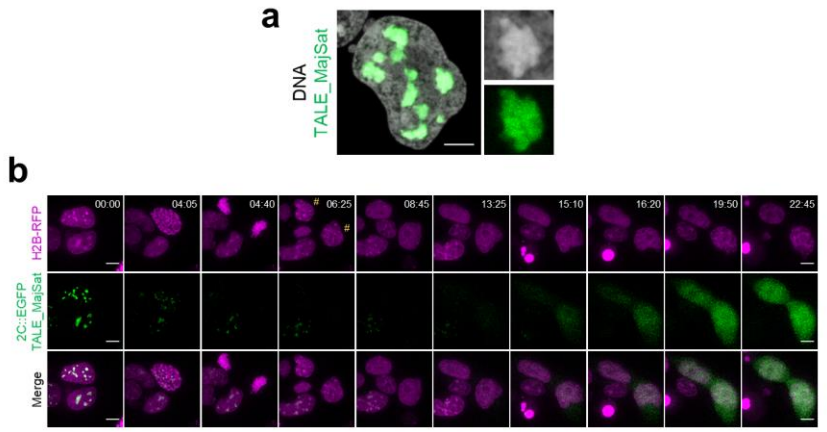

Figure 30. Visualization of chromocenter decompaction during 2C$^+$ reprograming by long-term live cell imaging. a, Representative live cell image of an ESC transiently expressing a TALE-mClover reporter targeting mouse major satellite sequences (TALE_MajSat). Hoechst 33342 was used to stain DNA. A DNA-dense heterochromatic foci displaying major satellite signal is shown on the right. Scale bar, 5 µm. **b**, Snapshots of a 24 hours time-lapse movie of ESCs reprogramming to 2C$^+$ cells. ESCs were modified to constitutively express H2B tagged with monomeric RFP (H2B-RFP, magenta). ESCs were transiently expressing a TALE-mClover reporter targeting mouse major satellite sequences (TALE_MajSat, green). Time stamp indicates hh:mm (hours:minutes). From 13:25 onwards, 2C::EGFP reporter expression was detected in the green channel. Hash symbols indicate cells that fail to compact chromocenters after a mitotic division and activate the 2C::EGFP reporter. Scale bar, 50 µm.

H3K9me3 heterochromatin becomes rapidly formed following exit from the 2C-like state

We then asked whether $2C^+$ cells could undergo the reverse transition, exiting the 2C-like state and subsequently re-entering pluripotency, thereby becoming ESC-like cells. To answer this question about the kinetics of the reverse transition, we followed the expression of EGFP in FACS-sorted $2C^+$ cells 24 hours, 48 hours, 72 hours and one week after sorting (Figure 31a). Strikingly, over 60 % of the $2C^+$ cells in culture lost the expression of the MERVL reporter 24 h after sorting. Moreover, 48 h after sorting, only 6 % of the cells still expressed the reporter, suggesting rapid repression of the 2C program, and quick re-establishment of the pluripotency network (Figure 31b). 72 hours and 7 days (7 d) after sorting, EGFP expression levels were comparable to those derived from the endogenous fluctuation ("*steady state*") of ESC cultures (Figure 31b).

The decay in EGFP levels was accompanied by a downregulation of MERVL and endogenous Dux gene expression (Figure 31c). Both MERVL and Dux had a peak of expression at 24 hours after $2C^+$ sorting and reached basal expression levels after 7 days (Figure 31c). In fact, GFP negative ESC colonies were observed 72 hours after sorting, which turned in rounded shape morphology by day 5 (Figure 31d). These results indicate that 2C-like cells could revert to

an ESC-like phenotype after Dux overexpression, and that such transition occurs rapidly, as early as 24 h after sorting.

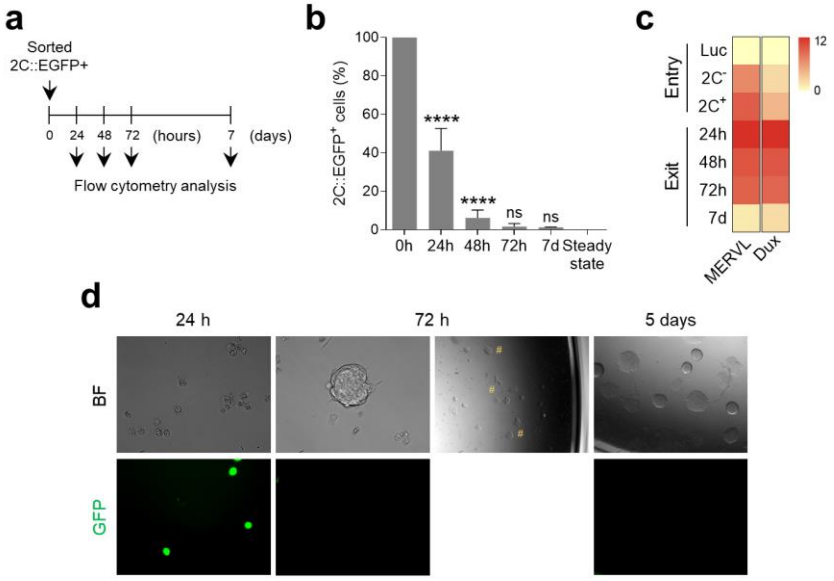

Figure 31. 2C$^+$ cells can revert to an ESC-like state. a, Scheme of the 2C$^+$ exit experiment in which GFP$^+$ 2C-like cells were purified after 24 hours of Dox treatment by flow cytometry sorting and re-cultured. At the indicated time points, cells were re-analyzed by flow cytometry. **b**, Quantification of the percentage of 2C-like cells 24 hours, 48 hours, 72 hours and 7 days after 2C$^+$ cell sorting. Endogenous 2C-like fluctuation was used as the steady state condition. Shown are the mean ± SD (n = 3 independent experiments). P = 0.7656^{ns}, P < 0.0001**** by one-way ANOVA (Tukey's multiple comparisons test). **c**, Heat map representation of MERVL and *Dux* expression in luciferase (Luc), 2C$^-$ and 2C$^+$ sorted cells (entry) and in ESC-like cells at 24 hours, 48 hours, 72 hours and 7 days (7d) after 2C$^+$ sorting. Expression values are presented as $\log_2$ fold change (FC) to luciferase detected by qRT-PCR. **d**, Representative images showing bright-field (BF) and GFP channels at 24 hours, 72 hours and 5 days after 2C$^+$ flow cytometry sorting. Hash symbols indicate GFP$^-$ ESC colonies appearing after 72 hours of 2C$^+$ cells re-culturing.

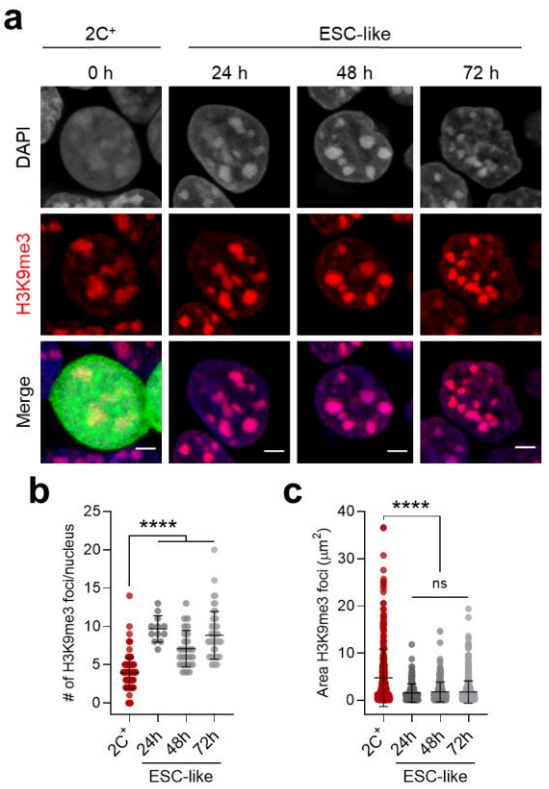

Figure 32. H3K9me3 heterochromatin becomes rapidly formed after exiting the 2C-like state. a, Representative immunofluorescence images of H3K9me3 at 0 hours ($2C^+$ before exit), 24 hours, 48 hours, 72 hours after 2C-like state exit. Scale bar, 3 µm. **b**, Quantification of the number of H3K9me3 foci in $2C^+$ cells and at 24 hours, 48 hours, 72 hours after 2C-like state exit. Shown are scatter dot plots with line at mean ± SD (n = 2 independent cultures, $2C^+$ = 119 cells, same dataset plotted in Fig. 27b; ESC-like 24 hours = 12 cells; ESC-like 48 hours = 27 cells; ESC-like 72 hours = 49 cells). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). **c**, Quantification of H3K9me3 foci area in $2C^+$ cells and at 24 hours, 48 hours, 72 hours after 2C-like state exit. Shown are scatter dot plots with line at mean ± SD (n = 2 independent cultures, $2C^+$ = 340 foci, same dataset plotted in Fig. 27c; ESC-like 24 hours = 168 foci; ESC-like 48 hours = 238 foci; ESC-like 72 hours = 605 foci). $P > 0.05^{ns}$, $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test).

We then quantified the number and area of H3K9me3 foci during the $2C^+$-to-ESC-like transition (Figures 32a-c). Our results indicate that chromocenters underwent rapid re-formation and increased in number (24 h: 9.67 ± 0.50; 48 h: 7.07 ± 0.46; 72 h: 8.82 ± 0.44 foci/nucleus) as compared to $2C^+$ cells (3.89 ± 0.19 foci/nucleus), concomitantly to the loss of EGFP expression and to the exit from the 2C-like state (Figures 32a, 32b). The areas of chromocenters in ESC-like cells were similar across the different time-points analyzed (24 h: 1.54 ± 0.15; 48 h: 1.77 ± 0.14; 72 h: 1.75 ± 0.10 μm^2) and smaller of those of $2C^+$ cells (4.76 ± 0.33 μm^2) (Figures 32a, 32c). These results suggest that the *in vitro* transition of the $2C^+$ cells toward ESC-like state can be used as model system to study chromocenter formation and chromatin reorganization occurring during early development.

Cell cycle arrest in G2/M and S increases the conversion of ESCs toward 2C-like

Several prominent GO terms associated with cell cycle and DNA replication appeared enriched in the $2C^-$ and Luc chromatomes. Long-term live cell imaging also provided some hint about a presumed connection between cell cycle and 2C-like cell induction. Thus, we aimed to investigate the role of the cell cycle in the 2C-like conversion. First, we looked at cell cycle progression in the heterogeneous population of cells

generated after Dux overexpression since it has been previously shown that spontaneous 2C-like cells have an altered cell cycle (Eckersley-Maslin et al. 2016).

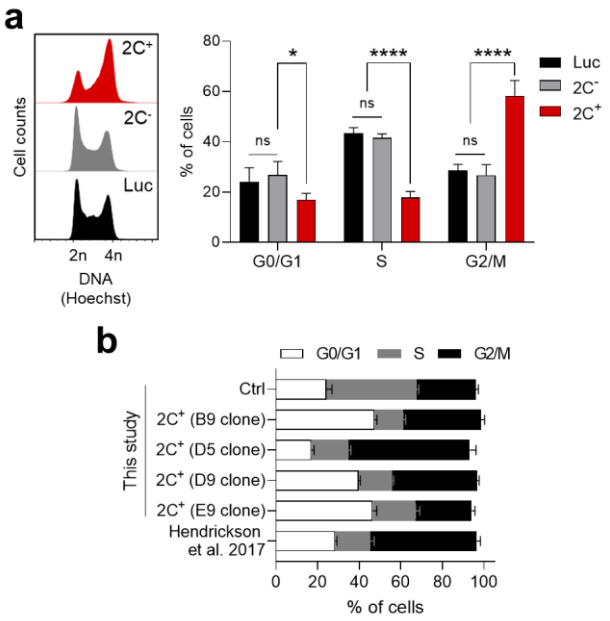

Figure 33. 2C⁺ cells have a prolonged G2/M and a short G1/S cell cycle phase. a, Cell cycle profile of non-induced luciferase (Luc) control ESCs, 2C⁻ and 2C⁺ cells (left). Quantification of the percentage of ESCs, 2C⁻ and 2C⁺ cells in different phases of the cell cycle (right). Shown are the mean ± SD (n > 3 independent cultures). $P > 0.05^{ns}$, $P = 0.0289$*, $P < 0.0001$**** by two-way ANOVA (Dunnett's multiple comparisons test). **b**, Quantification of the percentage of control ESCs (Ctrl) and 2C⁺ cells in different phases of the cell cycle. 2C⁺ cells induced from several Dux overexpressing clonal lines generated in our laboratory were analyzed. An independent Dux-CA clonal line was included for comparison, ref. Hendrickson et al., 2017. Data are presented as mean ± SD (n ≥ 3, technical replicates).

2C⁻ cells displayed a cell cycle profile comparable to that of luciferase cells, whereas 2C⁺ cells accumulated in the G2/M cell cycle phase (Figure 33a). The observation that 2C⁺ cells

showed very reduced S phase and accumulated in G2/M was consistent in several clonal lines (Figure 33b), suggesting a possible G2/M cell cycle arrest.

Histone deacetylases (HDAC) inhibition has been reported to increase the number of 2C-like cells (Macfarlan et al. 2012). On the other hand, the HDAC inhibitor trichostatin A (TSA) can induce G2/M cell cycle arrest in various cancer cell lines (Qiu et al. 2000; Yamashita et al. 2003; Meng et al. 2012). Thus, we used TSA to validate whether a G2/M cell cycle arrest was sufficient to induce ESC-to-2C$^+$ transition.

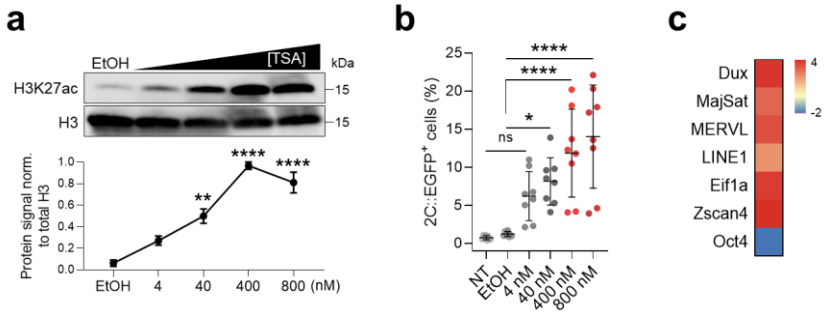

Figure 34. HDAC inhibition induces the 2C-like state. a, Western blot showing endogenous protein levels of histone H3 lysine 27 acetylation (H3K27ac) in EtOH and TSA-treated ESCs (top). Quantification of the normalized H3K27ac protein signal (bottom). Data are presented as mean ± SE (n = 3 independent experiments and blots). P = 0.0013**, P < 0.0001**** by one-way ANOVA (Dunnett's multiple comparisons test). **b**, Effect of increasing concentrations of TSA treatment on the percentage of 2C$^+$ cells. Data are presented as scatter dot plots with line at mean ± SD (n = 3 independent experiments). $P > 0.05^{ns}$, P = 0.0176*, P < 0.0001**** by one-way ANOVA (Tukey's multiple comparisons test). **c**, Heat map representation of the expression values of 2C-related genes and the pluripotency marker Oct4 in sorted 2C$^+$ cells after 24 hours of treatment with 40 nM of TSA. Expression values are presented as $\log_2$ fold change (FC) to 2C$^-$ detected by qRT-PCR.

TSA-treated ESCs displayed, as expected, higher levels of histone acetylation (Figure 34a). Additionally, TSA-treated ESCs displayed a dose dependent activation of the 2C::EGFP reporter (Figure 34b). Accordingly, we found upregulation of several 2C-specific genes, such as Dux, MERVL, LINE1 and Zscan4, and downregulation of the pluripotency transcription factor Oct4 (Figure 34c).

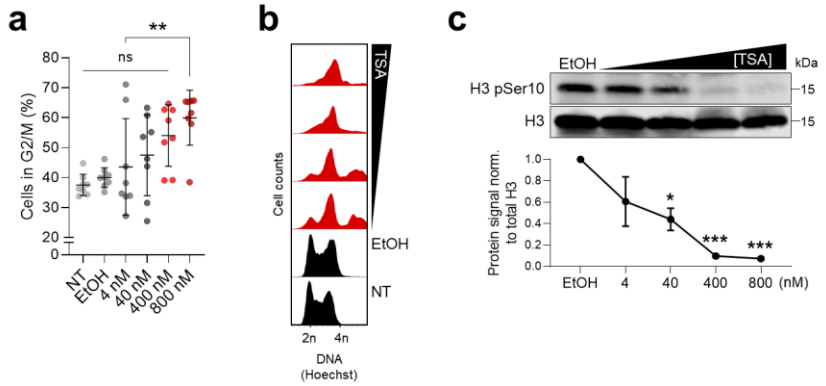

Figure 35. HDAC inhibition generates a G2/M cell cycle arrest. a, Effect of increasing concentrations of TSA treatment on the percentage of cells in the G2/M cell cycle phase. Data are presented as scatter dot plots with line at mean ± SD (n = 3 independent experiments). $P > 0.05^{ns}$, $P = 0.0055^{**}$ by one-way ANOVA (Tukey's multiple comparisons test). **b**, Cell cycle profile of non-treated (NT), EtOH and TSA-treated ESCs. Increasing concentrations of TSA were administered. **c**, Western blot showing endogenous protein levels of phospho-histone H3 Ser10 (H3 pSer10) in EtOH and TSA-treated ESCs (top). Quantification of the normalized H3 pSer10 protein signal (bottom). Data are presented as mean ± SE (n = 3 independent experiments and blots). $P = 0.0187^{*}$, $P = 0.0008^{***}$, $P = 0.0006^{***}$ by one-way ANOVA (Dunnett's multiple comparisons test).

We asked whether TSA treatment induces G2/M arrest in $2C^{+}$ cells since we already observed that pharmacological inhibition of HDACs was sufficient to increase the fraction of 2C-like

cells. TSA-treated ESCs showed increased numbers of $2C^+$ cells blocked in G2/M in a dose dependent manner (Figures 35a, 35b). Then, we wondered if G2/M-arrested $2C^+$ cells were blocked after DNA replication (G2) or during mitosis (M), thus we investigated if increasing TSA concentrations altered the levels of histone H3 phosphorylation, which is a hallmark of mitotic chromosome condensation. Histone H3 phosphorylated at serine 10 (pSer10) was barely detectable at TSA concentrations higher than 40 nM (Figure 35c). Given that phosphorylated histone H3, a key signal to initiate mitotic chromosome condensation, was lost upon induction of the 2C-like state and that $2C^+$ cells showed an interphase-like nuclear morphology, we concluded that $2C^+$ cells are arrested in G2 phase upon TSA treatment.

Progressive HDAC inhibition showed a linear correlation with both the proportion of induced $2C^+$ cells ($R^2 = 0.95$) and with the fraction of cells blocked in G2/M phase ($R^2 = 0.93$) (Figures 36a, 36b). H3 acetylation, however, reached saturation, suggesting an exponential plateau model for non-linear regression ($R^2 = 0.90$) (Figure 36c). These results further support the correlation between G2/M cell cycle arrest and 2C-like cell induction.

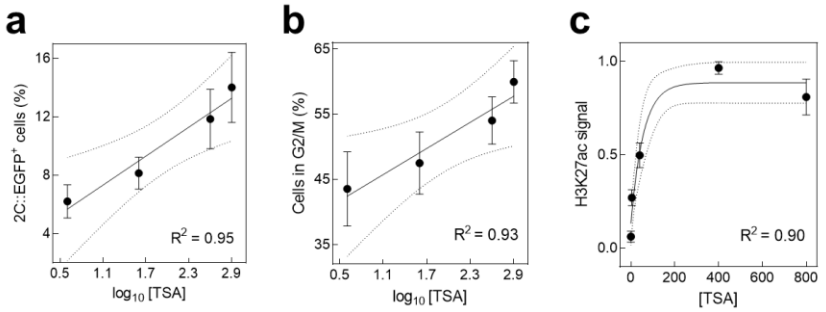

Figure 36. Correlations between HDAC inhibition and 2C$^+$ induction, G2/M arrest and histone acetylation. a, Correlative plot between TSA concentration and the percentage of 2C$^+$ cells. Data are presented as mean ± SE. Solid line represents fitting to linear regression model. Dotted lines show 95 % of confidence interval. **b**, Correlative plot between TSA concentration and the fraction of cells in G2/M cycle phase. Data are presented as mean ± SE. Solid line represents fitting to linear regression model. **c**, Correlative plot between TSA concentration and H3K27ac signal. Data are presented as mean ± SE. Solid line represents fitting to exponential plateau model for non-linear regression.

Having found that TOPBP1 was specifically enriched in the 2C$^-$ and Luc chromatomes and not in the chromatome of 2C$^+$ cells in the chromatin proteomics analysis (Figure 25b), we asked if the lack of topoisomerase activity could promote 2C$^+$ cell induction. Thus, we investigated the effects of DNA topoisomerases inhibition on such transition. We treated ESCs with camptothecin (CPT) and ICRF-193, which are inhibitors of DNA topoisomerases I and II, respectively (Downes et al. 1994; Pommier 2006). Inhibition of topoisomerase II alone increased the number of 2C$^+$ cells 1.5-fold (Figure 37a) and triggered a prominent cell cycle arrest in the G2/M phase (Downes et al. 1994; Robinson et al. 2007) (Figure 37b). Simultaneous inhibition of DNA topoisomerases I and II

resulted into an enhanced effect, leading to a 2.4-fold increase in the fraction of 2C$^+$ cells (Figures 37a, 37c). In summary, we have gained evidence that induction of cell cycle arrest in G2/M, here mediated by topoisomerase inhibition, can trigger the 2C-like state.

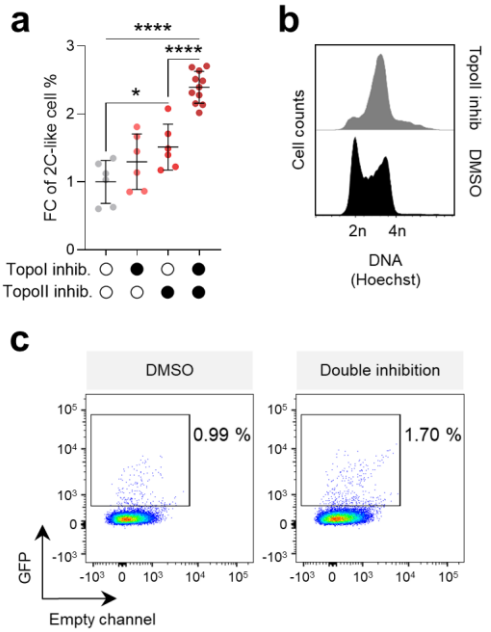

Figure 37. Topoisomerase inhibition increases 2C$^+$ cell conversion. **a**, Quantification of the percentage of 2C$^+$ cells after inhibition of DNA topoisomerase I (TopoI inhib.), DNA topoisomerase II (TopoII inhib.) or its combined inhibition. Data are presented as scatter dot plots with line at mean ± SD (n ≥ 3 independent experiments). *P* = 0.0421*, *P* < 0.0001**** by one-way ANOVA (Tukey's multiple comparisons test). **b**, Cell cycle profile of DMSO and DNA topoisomerase II inhibited ESCs. **c**, Representative FACS plots showing GFP$^+$ cells in DMSO and double DNA topoisomerase inhibition conditions.

Cell cycle is tightly regulated in pluripotent stem cells (Liu et al. 2019). Having shown an accumulation of 2C$^+$ cells in G2 phase after Dux overexpression and upon TSA treatment, we further

investigated the effect of different cell cycle arrests in the initiation of the 2C-like cell fate conversion. We treated ESCs with nocodazole, an inhibitor of cell cycle progression (Yiangou et al. 2019). This cell cycle arrest occurs without affecting ESC pluripotency or their global gene expression pattern (Yiangou et al. 2019). Treatment of ESCs with nocodazole led to a progressive and time-dependent increase in the number of $2C^+$ cells, reaching a 4.3 fold-increase after 16 hours treatment (Figure 38a). Indeed, it was evident that nocodazole synchronizes ESCs in prometaphase (Figure 38b). This result confirmed that also the arrest in the M phase is a strong inductor of 2C-like state emergence.

Although G2/M-specific cell cycle checkpoint activation stood out as the most prominent cell cycle phases to test 2C-like induction, we wondered what the contribution of DNA replication in the process was. We thus tested whether modulation of the replication fork speed could induce $2C^+$ cells. To do so, we treated cells with hydroxyurea (HU), a commonly used DNA replication fork-stalling agent (Saintigny et al. 2001). HU treatment induced generation of $2C^+$ cells, reaching a 2.2 fold-increase at 100 µM and 3.0 fold-increase at 200 µM (Figure 38c). These results suggest that reducing the speed of the replication fork also has an impact in inducing 2C-like cells. Nonetheless, such impact is milder when compared to the direct blockade of the cells in G2/M.

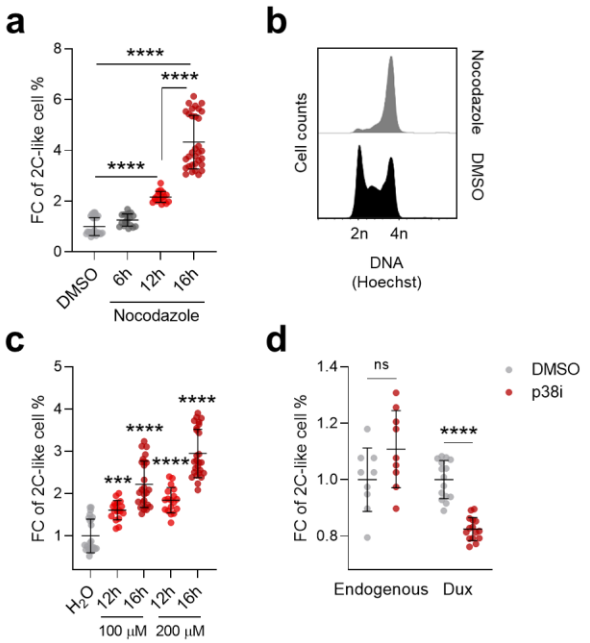

Figure 38. 2C⁺ cells increase upon perturbations in cell cycle progression. a, Effect of nocodazole treatment on the percentage of 2C⁺ cells. Data are presented as scatter dot plots with line at mean ± SD (n = 3 independent experiments). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). **b**, Cell cycle profile of DMSO and nocodazole treated ESCs. **c**, Effect of hydroxyurea (HU) on the percentage of 2C⁺ cells. Data are presented as scatter dot plots with line at mean ± SD (n = 3 independent experiments). $P = 0.0001$***, $P < 0.0001$**** by one-way ANOVA (Dunnett's multiple comparisons test). **d**, Effect of p38 inhibition (p38i) on the percentage of 2C⁺ cells in endogenous fluctuation and Dux overexpressing conditions. Data are presented as scatter dot plots with line at mean ± SD (n = 3 independent experiments). $P = 0.0853^{ns}$, $P < 0.0001$**** by unpaired two-tailed Student's *t*-test.

Finally, we tested the effect of inhibiting p38 mitogen-activated protein kinase (MAPK) pathway, which acts at multiple cell cycle phases, on 2C-like state induction. p38 MAPK regulates the G2/M as well as the G1/S cell cycle checkpoints upon different cellular stresses, such as upon induction of DNA

damage (Mikhailov et al. 2004; Cuenda and Rousseau 2007; Thornton and Rincon 2009). In addition, cell cycle arrest mediated by topoisomerase inactivation can only be triggered if the p38 MAPK pathway is active (Mikhailov et al. 2004). Thus, we reasoned that inhibition of p38 MAPK should lead to a reduction in the number of $2C^+$ cells. To test our hypothesis, we cultured ESCs in the presence of SB203580, a potent inhibitor of p38 MAPK (Bain et al. 2007). We found that, upon Dux overexpression, the percentage of $2C^+$ cells was significantly reduced (Figure 38d). Notably, this effect was only observed after challenging ESCs with Dux overexpression since no impact was detected in the endogenous fluctuation of 2C-like cells (Figure 38d). These results further confirmed that any type of cell cycle arrest, specifically G2/M and S, allows the transition of ESCs into the 2C-like state. The promiscuity observed with regard to the effects of different cell cycle perturbations in boosting $2C^+$ reprogramming might reflect a more intricate underlying mechanism controlling 2C-like state repression in pluripotent cultures.

CRISPR-Cas9 screening identifies the SWI/SNF-like remodeler SMARCAD1 as a contributing factor in the $2C^+$ to ESC-like transition

The results of the Dm-ChP suggested exploring the role of SMARCAD1 as well as a subset of other potential candidate

factors controlling the remodeling of chromocenters that we selected by literature mining. We therefore decided to investigate these potential regulators in $2C^+$ cells undergoing transition to ESC-like cells where chromocenters are formed *de novo*. We designed a small-content CRISPR-Cas9 loss-of-function screening to evaluate the relevance of some of the candidate factors in influencing the exit kinetics from the 2C-like state (Figures 39a, 39b). We cloned single guide RNAs (sgRNAs) targeting each of the genes of interest in an expression vector containing the SpCas9 (Cas9 protein from *Streptococcus pyogenes*) linked to mCherry protein via a T2A peptide (Figure 39a). We included a sgRNA targeting the luciferase sequence as control. Two independent sgRNA sequences targeting the same gene of interest were cloned in separate vectors and batches (Batch 1 and Batch 2) (Figure 39a).

We proceed to deliver the sgRNA-containing vector to the Dux-CA line where the gene knockout was performed (Figure 39b). Next, we purified $2C^+$ cells containing the gene knockout, tracked their $2C^+$ exit kinetics at different time points and compared them with the luciferase control condition (Figure 39b). It is important to highlight that the 2C-like state is metastable and, to date, it cannot be preserved in culture. Therefore, this dynamic CRISPR screening represents the most acceptable exploratory design we conceived to experimentally test a small subset of candidate factors.

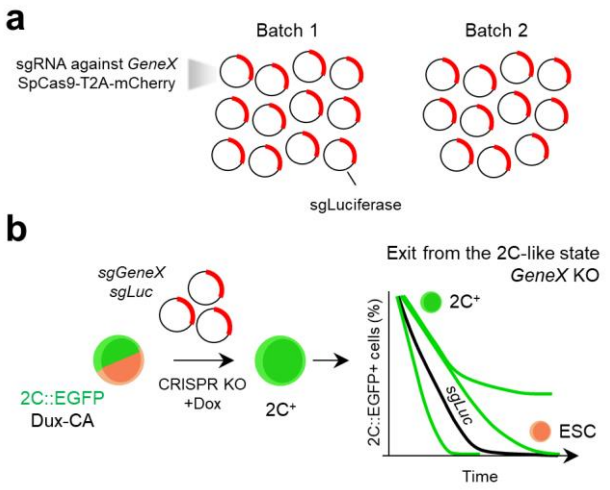

Figure 39. CRISPR-Cas9 loss-of-function screening to interrogate candidate factors during the 2C$^+$ exit. a, Scheme of the plasmids used in the CRISPR-Cas9 loss-of-function screening. sgRNAs targeting each individual gene of interest (GeneX) were cloned in an SpCas9-T2A-mCherry vector. **b**, Schematic representation of the CRISPR-Cas9 loss-of-function (CRISPR KO) screening workflow. sgGeneX refers to the gene X that is targeted by the sgRNA. sgLuc refers to an sgRNA targeting the luciferase sequence as control.

This subset of candidate factors was predicted to be interacting among each other (Figure 40a). Indeed, some factors were previously associated with heterochromatin, such as the helicase HELLS, CBX5 (also referred to as heterochromatin protein 1, HP1) and SMARCAD1, while others function as chromatin remodelers like the histone-binding proteins RBBP4/7, HELLS and SMARCAD1 (Figure 40a). We then interrogated the effect of the individual downregulation of each of these factors, monitoring the expression of the 2C::EGFP reporter 24 hours, 48 hours and 72 hours after the 2C$^+$ exit (Figure 40b). We classified the candidate genes into two

distinct sets, namely Set 1 and Set 2, based on the alterations of the $2C^+$ exit kinetics (Figure 40b). Knocking-out (KO) genes from Set 1 resulted in faster repression of the MERVL elements, as compared to control (sgLuc), especially at 24 hours and 48 hours from the exit (Figure 40b). Knocking-out genes from Set 2, instead, extended the time-window of MERVL expression (until 72 hours after the initial induction for most of the cases) (Figure 40b). Interestingly, Set 2 mainly included genes coding for DNA replication-related factors. Of note, deletion of Set 2 genes compromised cell viability to a larger extent when compared with targeting Set 1 genes, which was more tolerated.

This further confirms the importance of DNA replication for a successful exit from the 2C-like state, and reinforces the notion that cell cycle regulation, particularly DNA replication and G2/M phase arrest, is interconnected with 2C-like state induction. Among all candidates, we observed that Smarcad1 KO (sgSmarcad1) led to the strongest increase in $2C^+$ exit kinetics (Figure 40b). We then asked whether this observation was the result of an altered number of $2C^+$ cells at initial time points upon SMARCAD1 depletion. SMARCAD1 depletion resulted in no major impact in the $2C^+$ conversion either in the endogenous fluctuation or in Dux-induced cells (Figure 40c), thus prompting to suggest a specific role of SMARCAD1 during the $2C^+$ exit.

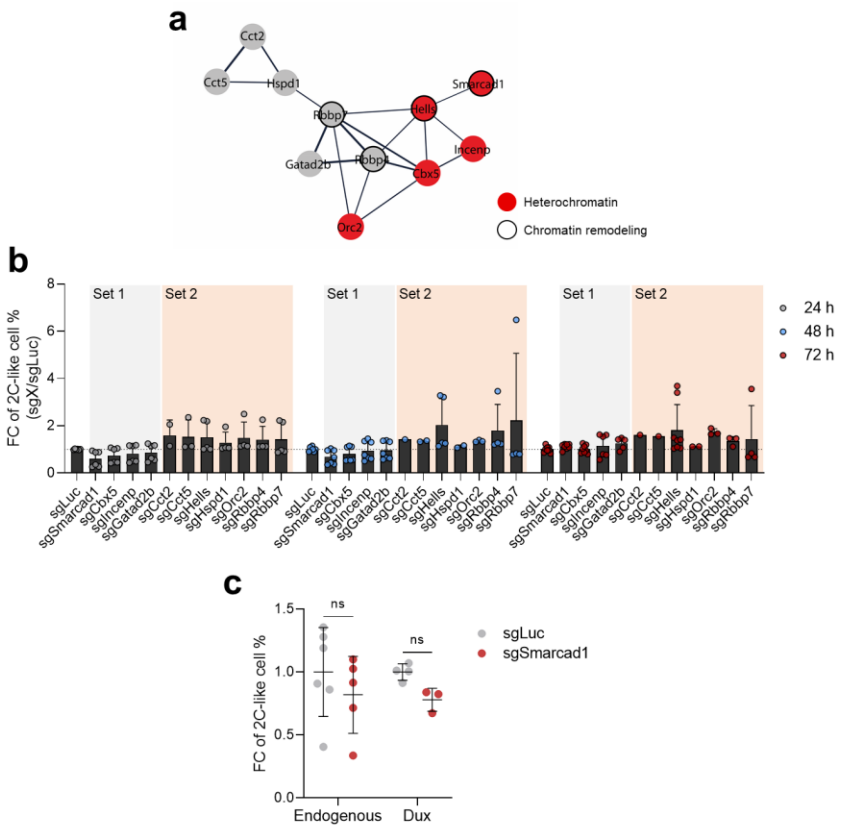

Figure 40. CRISPR-Cas9 loss-of-function screening during the 2C$^+$ to ESC-like transition. a, Functional protein network of the candidate factors involved in chromocenter reorganization. Nodes are colored in red to highlight association with heterochromatin. Black node border indicates chromatin remodeling function. Network edges indicate the degree of confidence prediction of the interaction. Protein interaction data were retrieved from the STRING database. The network was visualized with Cytoscape. **b**, Impact of the individual candidate factor gene knockout (KO) on 2C-like cell percentage during the 2C$^+$ exit (24 hours, 48 hours and 72 hours after 2C$^+$ cell sorting). Data are presented as mean ± SD of n = 2 independent CRISPR KO screenings along the 2C$^+$ exit with independent sgRNAs targeting the same target gene. Scatter dot plots indicate individual scores within each independent CRISPR KO screening. Dotted line indicates the control KO cells (sgLuc) kinetics during the 2C$^+$ exit. sgX refers to the gene X targeted by the sgRNA.

Figure 40 (cont.). b, sgX refers to the gene X targeted by the sgRNA. Grey panel indicates genes classified within Set 1 (Smarcad1, Cbx5, Incenp, Gatad2b). Beige panel indicates genes classified within Set 2 (Cct2, Cct5, Hells, Hspd1, Orc2, Rbbp4, Rbbp7). **c**, Impact of targeting Smarcad1 (sgSmarcad1) on the endogenous fluctuation and the Dux-induced 2C-like conversion. Data are presented as scatter dot plots with line at mean SD (n≥3 independent CRISPR-Cas9 KO rounds). $P = 0.4286^{ns}$, $P = 0.0571^{ns}$ by Mann-Whitney test.

SMARCAD1, a SWI/SNF-like chromatin remodeler, is known to promote heterochromatin maintenance during DNA replication in terminally differentiated cells and silencing of endogenous retroviruses in ESCs (Rowbotham et al. 2011; Sachs et al. 2019). Nonetheless, it is not known whether SMARCAD1 plays a role in 2C-like fate transition and early embryo development. Remarkably, and as mentioned previously, SMARCAD1 was reported to interact with TOPBP1 (Bantele et al. 2017), which is cell cycle-dependent and one of the top factors enriched in the $2C^-$ and Luc chromatomes (Figure 25b).

It should be mentioned that control KO cells targeting luciferase (sgLuc) followed comparable exit kinetics to non-transfected (NT) ESCs (Figure 41a). Just a minor deviation from the NT kinetics was observed at 48 hours after $2C^+$ exit (Figure 41a). We then verified that the changes reported in the CRISPR loss-of-function screening were actually due to a drastic reduction in the expression of the genes of interest (Figure 41b). We detected a downregulation in the expression levels of all targeted genes, suggesting that the observed

variation in 2C$^+$ exit kinetics could be attributed to the loss of the stated genes (Figure 41b).

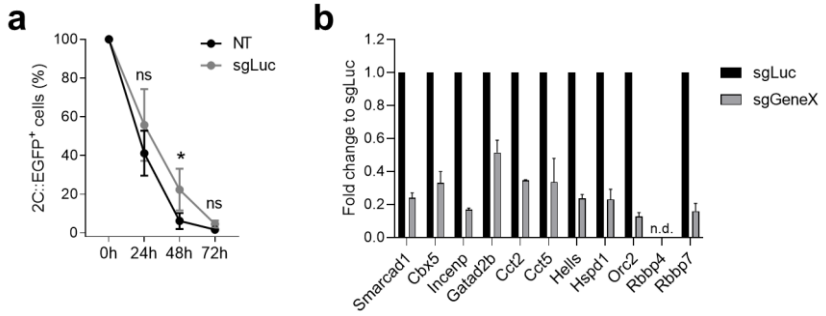

Figure 41. Characterization of the CRISPR screening. a, Quantification of the percentage of 2C-like cells 24 hours, 48 hours and 72 hours after 2C$^+$ cell sorting in non-transfected (NT) and luciferase-targeted (sgLuc) ESCs. Data are presented as mean ± SD. $P > 0.05^{ns}$, $P = 0.0463$* by two-way ANOVA (Sidak's multiple comparisons test). **b**, qRT-PCR of CRISPR-Cas9 sgRNA target genes in control luciferase-targeted (sgLuc) or candidate factor-targeted (sgGeneX) mCherry$^+$ sorted ESCs. Data are presented as mean ± SD from two replicates transfected in independent rounds of sgRNA delivery. Independent sgRNAs targeting the same target gene were used in each round. n.d. indicates that values were not detectable in the qRT-PCR run.

We decided to further explore the effect of targeting candidate genes at the transcriptome level by performing RNA sequencing (RNA-seq). First, we confirmed that the genes of interest were largely downregulated in each of the independent samples compared with sgLuc control (Figure 42a). To elucidate whether the downregulation of these genes was associated with a derepression or silencing of zygotic genome activation (ZGA) genes and/or pluripotency-associated genes, we interrogated the expression of the differentially expressed genes (DEGs) after the CRISPR perturbation during the stages of preimplantation development (Figures 42b, 42c). Genes up-

regulated after targeting Set 1 were mainly expressed from 4-cell stage to blastocyst, peaking at the 8-cell to 16-cell stage, whereas down-regulated genes showed a preference to be

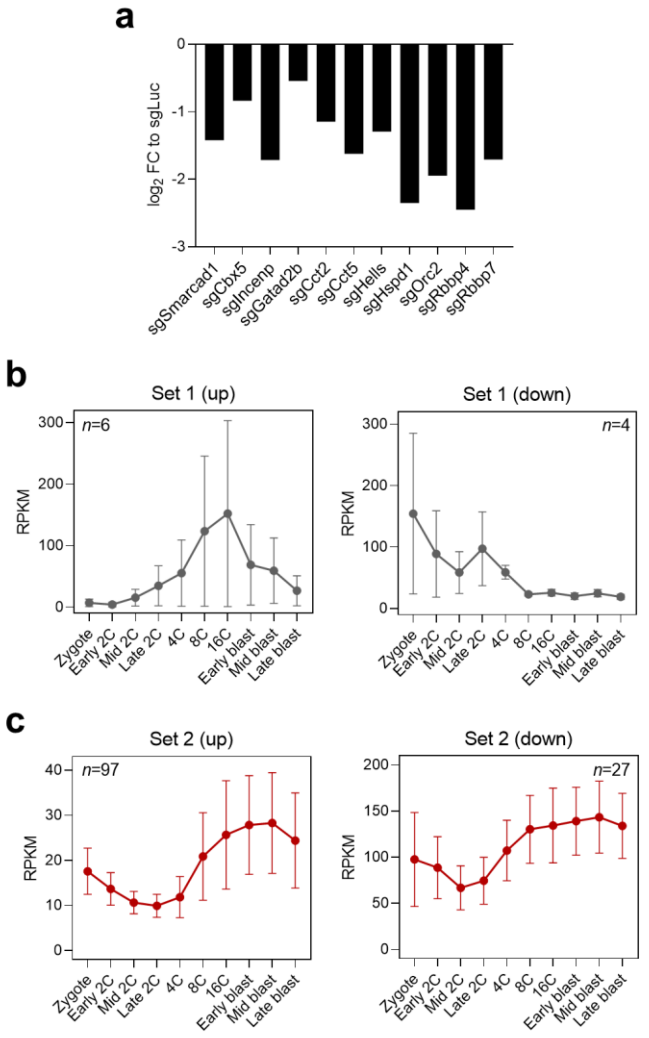

Figure 42. Transcriptome changes after targeting Set 1 and Set 2 genes by RNAseq. a, Expression values of CRISPR KO target genes in mCherry$^+$ sorted ESCs after 48 h of sgRNA delivery. Expression values are presented as $\log_2$ fold change (FC) to sgLuc detected by RNA-seq from two replicates transfected in independent rounds of sgRNA delivery.

Figure 42 (cont.). b, Single-cell RNA-seq (scRNA-seq) expression values of differentially up-regulated (left; n = 6) and down-regulated (right; n = 4) genes of the Set 1 cluster during pre-implantation development (from zygote to the late blastocyst stage). Data are presented as mean ± SE from all single-cells profiled in each developmental stage. Differentially expressed genes were retrieved after applying an adjusted *p*-value < 0.01 in our dataset. scRNA-seq data was obtained from ref. Deng et al., 2014. Blast, blastocyst. RPKM, reads per kilobase of transcript per million mapped reads. **c**, Single-cell RNA-seq (scRNA-seq) expression values of differentially up-regulated (left; n = 97) and down-regulated (right; n = 27) genes of the Set 2 cluster during pre-implantation development (from zygote to the late blastocyst stage). Data are presented as mean ± SE from all single-cells profiled in each developmental stage. Differentially expressed genes were retrieved after applying an adjusted *p*-value < 0.01 in our dataset.

expressed in early stages, starting at zygote to 2-cell stage (Figure 42b).

It should be noted that Set 1 perturbations did not resulted in a high number of DEGs although their specific target genes were prominently downregulated (Figure 42b). For instance, sgSmarcad1 merely showed Smarcad1 gene affected (log_2FC = -1.42; adjusted *p*-value = 1.14e-51). Genes both up- and down-regulated after targeting Set 2 were mainly expressed from 4-cell/8-cell stage to blastocyst with a quite uniform expression profile (Figure 42c). In fact, they showed an expression profile typical of those genes transcribed after the major ZGA.

SMARCAD1 associates with H3K9me3 in ESCs and its nuclear localization is lost in the 2C-like state

The results of the Dm-ChP revealed TOPBP1 as potential regulator of chromocenter reorganization. SMARCAD1 has been shown to interact with TOPBP1 in yeast and human cells (Bantele et al. 2017) (Figure 25b), and provided the strongest increase in $2C^+$ exit kinetics in the CRISPR screening (Figure 40b). As alluded previously, the SWI/SNF-like chromatin remodeler SMARCAD is known to promote heterochromatin maintenance during DNA replication in terminally differentiated cells and silencing of endogenous retroviruses in ESCs (Rowbotham et al. 2011; Sachs et al. 2019). However, it is still mysterious whether SMARCAD1 plays a role in 2C-like fate transition and early embryo development.

We therefore decided to investigate SMARCAD1 and TOPBP1 in ESCs and $2C^+$ cells undergoing transition to ESC-like cells where chromocenters are formed *de novo*. First, we decided to evaluate whether SMARCAD1 was associated with TOPBP1 and/or H3K9me3 in ESCs, since TOPBP1 enriched in the $2C^-$ and Luc chromatomes (Figure 25b). Indeed, we observed that SMARCAD1 was associated with both H3K9me3 and TOPBP1 (Figure 43). However, we could not detect any evidence of association between TOPBP1 and H3K9me3 (Figure 43), therefore suggesting SMARCAD1 as a direct potential regulator of H3K9me3 heterochromatin in ESCs and $2C^+$ cells.

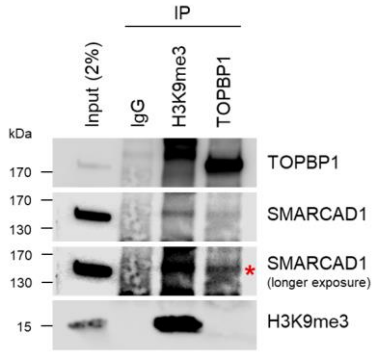

Figure 43. SMARCAD1 interacts with H3K9me3 and TOPBP1 in ESCs. Co-immunoprecipitation (co-IP) experiment in ESCs using antibodies against H3K9me3 and TOPBP1. ESC nuclei were lysed in the presence of benzonase nuclease. Western blots of H3K9me3, SMARCAD1 and TOPBP1 are shown. Rabbit IgGs were used as isotype controls. Asterisk indicates band detected after exposing the SMARCAD1 blot for longer time.

We then asked whether SMARCAD1 localized at chromocenters in ESCs. We found that SMARCAD1 co-localized with H3K9me3 in heterochromatin foci of chromocenters in both ESCs and $2C^-$ cells (Figures 44a-c). In contrast, the expression of SMARCAD1 was lost in $2C^+$ cells, where foci were much reduced in number (Figures 44b, 44c).

In agreement with these observations, using published single-cell RNA-seq (scRNA-seq) data (Deng et al. 2014), we found that Smarcad1 is expressed during preimplantation development (Figure 45). Smarcad1 expression starts at the 2-cell stage, but increases at the 4-cell stage embryo, which is the time when chromocenters compact during mouse embryo development (Figure 45).

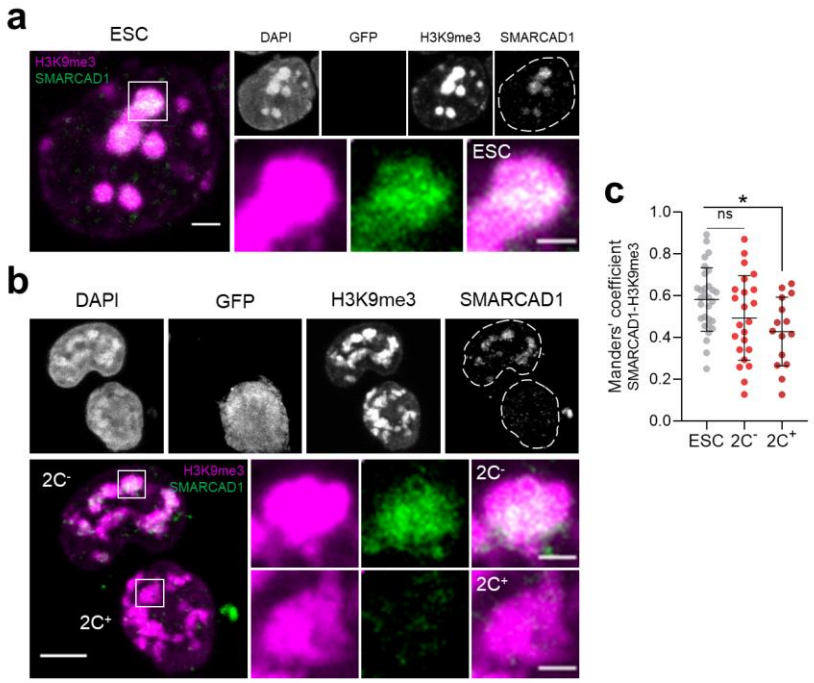

Figure 44. SMARCAD1 associates with H3K9me3 in ESCs and 2C$^-$ cells but not in 2C$^+$ cells. a, Representative immunofluorescence images of H3K9me3 and SMARCAD1 in ESCs. Dashed lines indicate nuclei contour. Scale bar, 2 µm. Zoomed images of H3K9me3 and SMARCAD1 foci are shown for comparisons. Scale bar, 1 µm. **b**, Representative immunofluorescence images of H3K9me3 and SMARCAD1 in 2C$^-$ and 2C$^+$ cells. Dashed lines indicate nuclei contour. Scale bar, 5 µm. Zoomed images of H3K9me3 and SMARCAD1 foci are shown for comparisons. Scale bar, 1 µm. **c**, Co-localization analysis showing Manders' coefficient between SMARCAD1 and H3K9me3 in ESCs, 2C$^-$ and 2C$^+$ cells. Data are presented as scatter dot plots with line at mean ± SD from ESC (n = 30), 2C$^-$ (n = 23), 2C$^+$ (n = 15) SMARCAD1-H3K9me3 foci. $P > 0.05^{ns}$, $P = 0.0124^*$ by one-way ANOVA (Dunnett's multiple comparisons test).

Notably, Topbp1 showed a similar expression profile during preimplantation development, peaking at the 4-cell stage embryo (Figure 46).

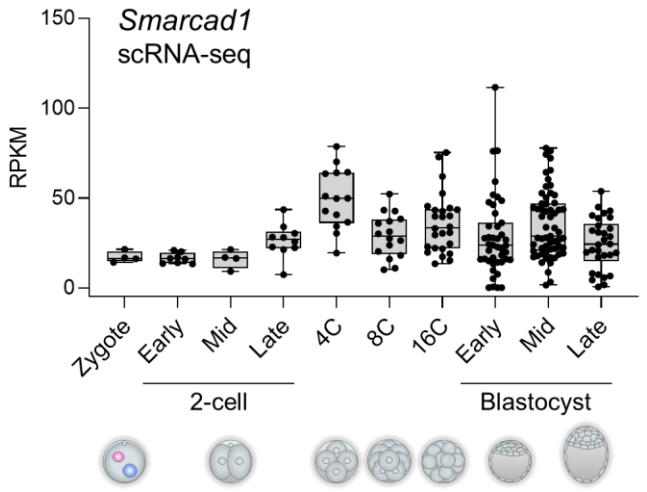

Figure 45. Smarcad1 expression profile in preimplantation mouse embryos. Single-cell RNA-seq (scRNA-seq) expression profile of Smarcad1 in pre-implantation mouse embryos. Data are presented as min-max boxplots with line at median. Each dot represents a single-cell. scRNA-seq data was obtained from ref. Deng et al., 2014. RPKM, reads per kilobase of transcript per million mapped reads.

Given that scRNA-seq pointed toward a role of SMARCAD1 in early embryo development, we proceed to investigate the nuclear distribution of SMARCAD1 during exit from the 2C-like state. We first observed a reduction in SMARCAD1 signal as ESCs started to express the MERVL reporter, attaining nearly total loss of SMARCAD1 in $2C^+$ at the 24 hour time point (Figures 47a, 47b). SMARCAD1 nuclear signal was then gradually recovered in the heterochromatin foci as $2C^+$ cells were converted in ESC-like cells up to the 72 hours from the exit, indicating reversibility of foci formation (Figures 47a, 47b).

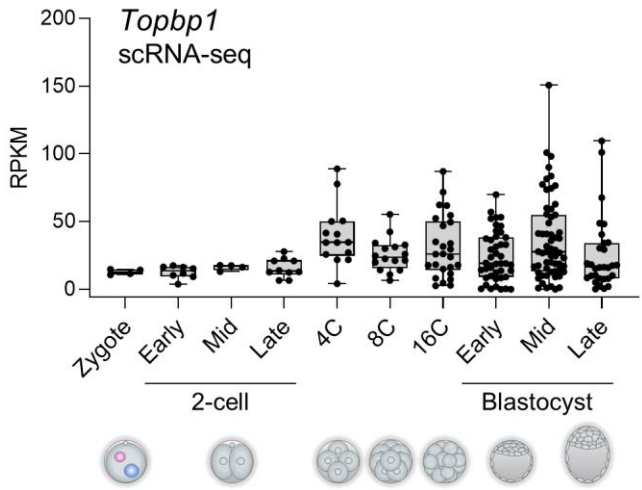

Figure 46. Topbp1 expression profile in preimplantation mouse embryos. Single-cell RNA-seq (scRNA-seq) expression profile of Topbp1 in pre-implantation mouse embryos. Data are presented as min-max boxplots with line at median. Each dot represents a single-cell. scRNA-seq data was obtained from ref. Deng et al., 2014. RPKM, reads per kilobase of transcript per million mapped reads.

Surprisingly, the fraction of cells that repressed retroelements within 24 hours from the $2C^+$ exit (ESC-like at 24 hours) already showed SMARCAD1 enriched foci (Figure 47a). These results suggest that SMARCAD1 was lost from chromatin as ESCs progress to the 2C-like state and, later, SMARCAD1 nuclear distribution was reverted during the $2C^+$ exit.

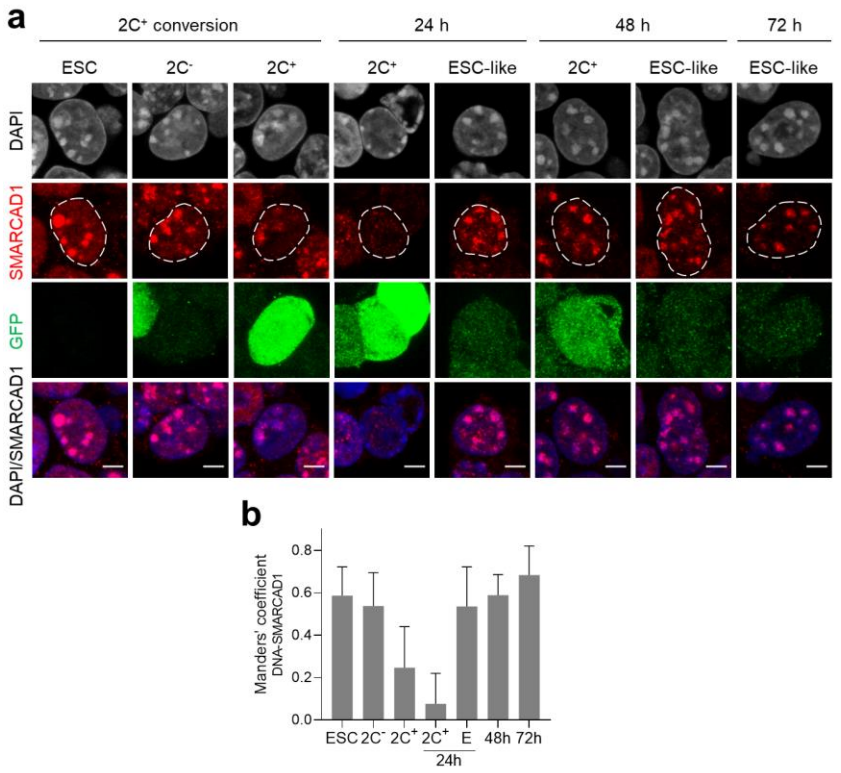

Figure 47. SMARCAD1 is transiently lost from chromatin as ESCs progress to the 2C-like state. a, Representative immunofluorescence images of SMARCAD1 and the 2C::EGFP reporter along the ESCs to 2C$^+$ reprogramming and during the 2C$^+$ exit (24 hours, 48 hours and 72 hours). Dashed lines indicate nuclei contour. Scale bar, 4 µm. **b**, Co-localization analysis showing Manders' coefficient between DNA and SMARCAD1 along the conversion of ESCs into 2C$^+$ cells and during the 2C$^+$ exit (24 hours, 48 hours and 72 hours). E, ESC-like.

SMARCAD1 downregulation impairs mouse embryo development and it is associated with a partial H3K9me3 reduction

Collectively, our findings suggested that SMARCAD1 could be a potential regulator of H3K9me3 heterochromatin in the $2C^{+}$ transition. With this in mind, we aimed at investigating SMARCAD1 function in preimplantation embryos. We injected zygote-stage (E0.5) embryos with morpholino antisense oligos (MO) targeting Rbbp7, Smarcad1, along with a scrambled control morpholino (Ctrl MO) (Figure 48).

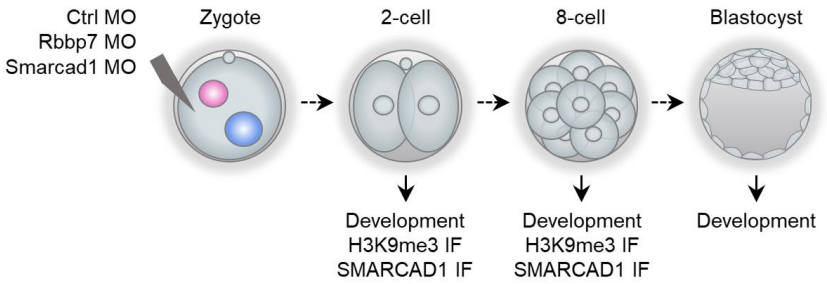

Figure 48. Role of SMARCAD1 in mouse embryos. Schematic representation of the experimental design to assess SMARCAD1 function in early mouse embryo development. Morpholino antisense oligos (MO) targeting Rbbp7, Smarcad1 and a scrambled control (Ctrl) sequence were microinjected into the cytoplasm of zygotes (E0.5 embryos). Embryo development was monitored daily from 2-cell stage (E1.5) until late blastocyst stage (E5.5). H3K9me3 and SMARCAD1 nuclear intensity and distribution were quantified by immunofluorescence (IF) in 2-cell (E1.5) and 8-cell stage (E2.5) embryos.

As representative gene from Set 2 genes, we included Rbbp7 for which we found that the KO prolonged kinetics of 2C exit (Figure 40b). We monitored embryo development daily from 2-

cell stage (E1.5) until late blastocyst stage (E5.5). H3K9me3 and SMARCAD1 nuclear intensity and distribution were quantified by immunofluorescence (IF) in 2-cell (E1.5) and 8-cell stage (E2.5) embryos (Figure 48).

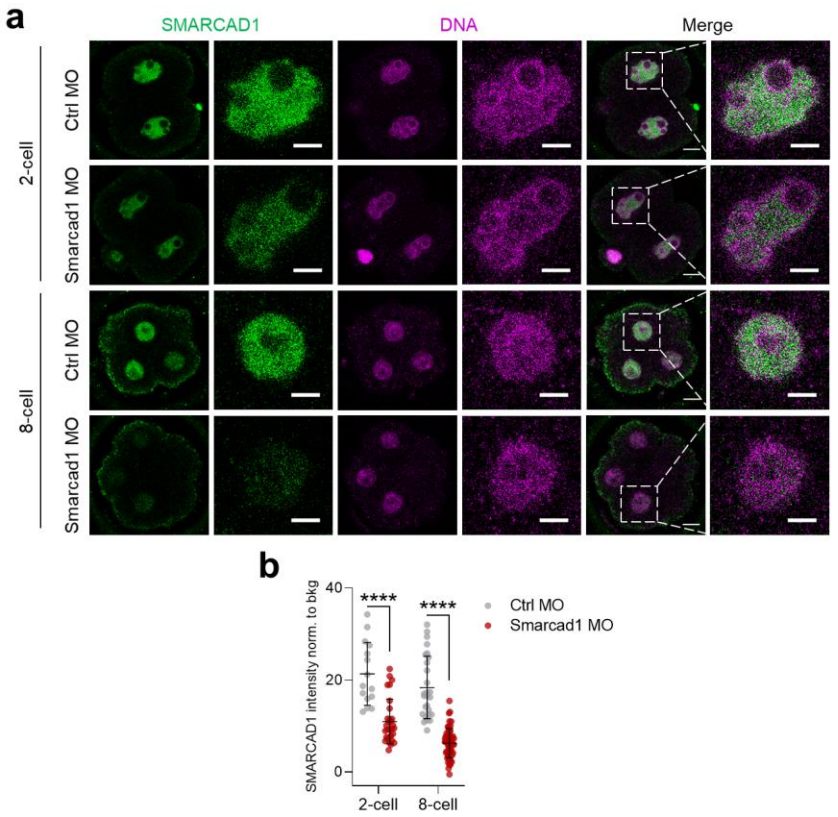

Figure 49. Smarcad1 morpholino efficiently degrades SMARCAD1 in mouse embryos. a, Representative immunofluorescence images of SMARCAD1 in Ctrl MO and Smarcad1 MO embryos at 2-cell (E1.5) and 8-cell stage (E2.5) embryos. Representative blastomere nuclei are shown. Scale bar, 10 µm. Zoomed images, 5 µm. **b**, Quantification of SMARCAD1 mean fluorescence intensity in control (Ctrl) and Smarcad1 morpholino-injected (MO) embryos at 2-cell (E1.5) and 8-cell stage (E2.5). Data are presented as scatter dot plots with line at mean ± SD (2-cell: Ctrl MO = 8 embryos, Smarcad1 MO = 15 embryos; 8-cell: Ctrl MO = 11 embryos, Smarcad1 MO = 15 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test.

As expected from MO, which acts by blocking translation,

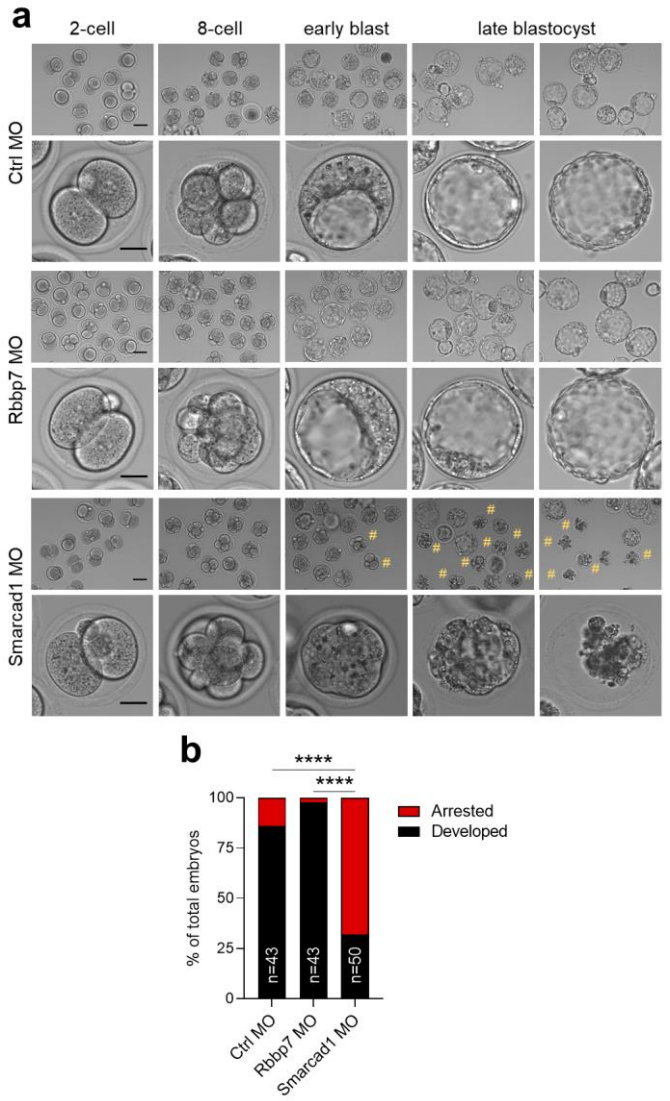

Figure 50. SMARCAD1 is necessary for early embryo development. **a**, Representative embryos from control (Ctrl), Rbbp7 and Smarcad1 morpholino-injected (MO) groups from 2-cell (E1.5) to late blastocyst stage (E5.5). Scale bar, 50 µm. Zoomed images, 20 µm. Hash symbols indicate dead or arrested embryos. **b**, Quantification of the percentage of arrested or fully developed embryos at late blastocyst stage (E4.5). *P* < 0.0001**** by Fisher's exact test (Ctrl MO = 43 embryos, Rbbp7 MO = 43 embryos, Smarcad1 MO = 50 embryos).

SMARCAD1 was degraded from 2-cell stage, reaching almost undetectable levels at 8-cell stage, in Smarcad1 MO injected embryos (Figures 49a, 49b). It is noteworthy that SMARCAD1 localizes exclusively in the nucleus of preimplantation embryos (Figure 49a).

We observed that embryos developed slower than normal when Smarcad1 was silenced (Figure 50a). Indeed, they did not show formation nor expansion of a blastocoel cavity at early blastocyst stage, indicating a severe developmental delay (Figure 50a). Notably, 68 % of the embryos deficient for Smarcad1 arrested and did not develop until late blastocyst stage (Figure 50b). In contrast, Rbbp7 knockdown completed development until blastocyst stage normally with no observed phenotypical defects (Figure 50a).

Since we observed that SMARCAD1 was necessary for embryo developmental progression, we decided next to image H3K9me3 upon depletion of SMARCAD1 (Figures 48, 51a). H3K9me3 signal was appreciably reduced in the embryos injected with Smarcad1 MO already at 8-cell stage (E2.5), almost one day earlier than early blastocyst (E3.5), when the developmental delay was morphologically visible (Figures 51a, 51b). The reduction in H3K9me3 signal at 8-cell stage was independent of changes in the nuclear size of the corresponding blastomeres (Figure 51c). These results show that SMARCAD1 knockdown impairs mouse embryo development and that the developmental arrest might be

associated with the partial loss of H3K9me3 observed already at 8-cell stage embryos. Overall, our results suggest that SMARCAD1 contributes to proper early embryo development.

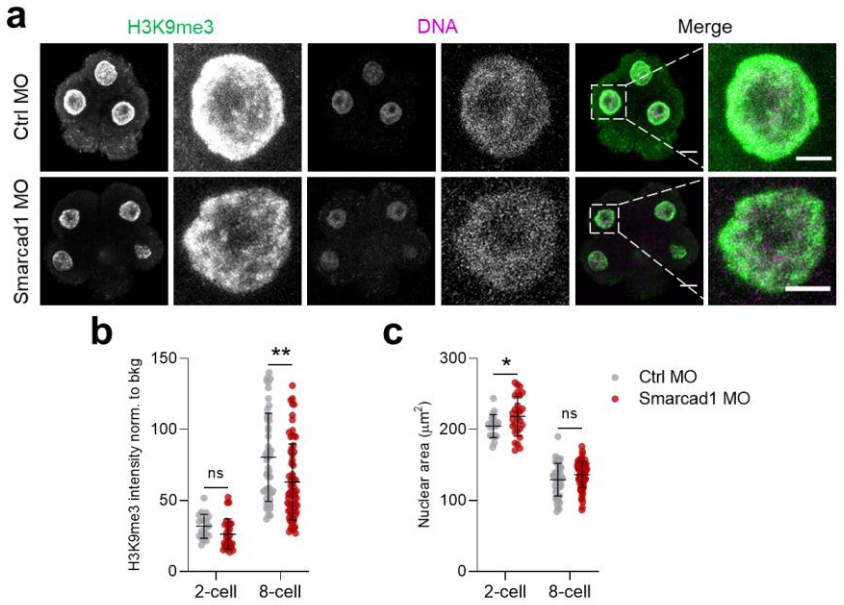

Figure 51. Depletion of SMARCAD1 leads to a reduction of H3K9me3 in preimplantation embryos. a, Representative immunofluorescence images of H3K9me3 in Ctrl and Smarcad1 MO embryos at 8-cell stage (E2.5) embryos. Representative blastomere nuclei are shown. Scale bar, 10 µm. Zoomed images, 5 µm. **b**, Quantification of H3K9me3 mean fluorescence intensity in control (Ctrl) and Smarcad1 morpholino-injected (MO) embryos at 2-cell (E1.5) and 8-cell stage (E2.5). Data are presented as scatter dot plots with line at mean ± SD (2-cell: Ctrl MO = 12 embryos, Smarcad1 MO = 15 embryos; 8-cell: Ctrl MO = 16 embryos, Smarcad1 MO = 20 embryos). H3K9me3 signal was normalized to the average background signal. $P = 0.0618^{ns}$, $P = 0.0016^{**}$ by unpaired two-tailed Student's *t*-test. **c**, Quantification of the nuclear area in control (Ctrl) and Smarcad1 morpholino-injected (MO) embryos at 2-cell (E1.5) and 8-cell stage (E2.5). Data are presented as scatter dot plots with line at mean ± SD (2-cell: Ctrl MO = 12 embryos, Smarcad1 MO = 15 embryos; 8-cell: Ctrl MO = 16 embryos, Smarcad1 MO = 20 embryos). $P = 0.0682^{ns}$, $P = 0.0404^{*}$ by unpaired two-tailed Student's *t*-test.

PART III - DISCUSSION and CONCLUSIONS

DISCUSSION

Heterochromatin formation during early embryogenesis is a fundamental aspect of development (Probst and Almouzni 2011). During the first cleavage stages, constitutive heterochromatin reorganizes in the nucleus to form highly compacted chromocenters (Jones 1970; Probst and Almouzni 2011; Burton et al. 2020). Although chromocenters are firstly detected in the 2-cell to 4-cell-stage transition, a systematic identification of the underlying factors involved in *de novo* heterochromatin establishment and, thus, in chromocenter compaction, is still lacking mostly because of the minuscule amount of material available during embryogenesis.

In this Thesis, using the Dux-dependent reprogramming system, we have reported that the transition from the 2C-like to the pluripotent state is a robust *in vitro* model system to investigate *de novo* establishment of heterochromatin foci and their dynamic reorganization in early embryo development (Figure 52). Previously, several studies have reported the transcriptional (Rodriguez-Terrones et al. 2018; Fu et al. 2019; Fu et al. 2020), epigenetic (Ishiuchi et al. 2015; Hendrickson et al. 2017; De Iaco et al. 2019; Eckersley-Maslin et al. 2019; Yan et al. 2019; Zhang et al. 2021) and metabolic (Rodriguez-Terrones et al. 2020) changes and properties associated with 2C-like cells. However, there was still a lack of knowledge with respect to the proteome, and more specifically the chromatin-bound proteome or chromatome, in the 2C-like state.

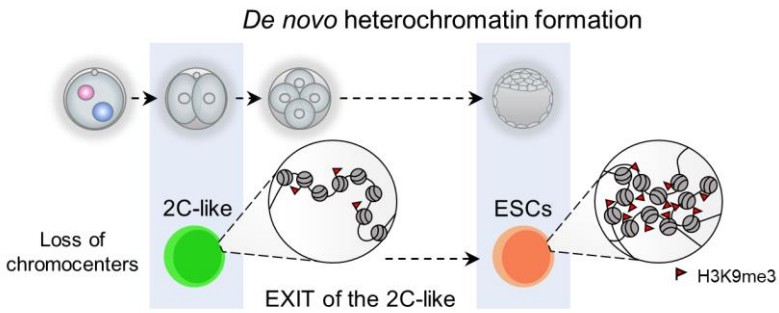

Figure 52. Deciphering the rules of *de novo* heterochromatin formation exploiting the 2C-like cell reprogramming system. 2C-like cells, which mimic the 2-cell stage of mouse embryos, are a unique cellular model to investigate *de novo* heterochromatin formation. Particularly, the exit of the 2C-like state toward pluripotency (ESCs) offers an unprecedented opportunity to follow heterochromatin (H3K9me3) dynamics since 2C-like cells have been shown to be devoid of chromocenters.

Almost the vast majority of the abovementioned studies focused on endogenous 2C-like cells. Endogenous 2C-like cells, which arise spontaneously in ESC and iPSC cultures (Macfarlan et al. 2012), appeared in very low numbers (~ 1 %), thus limiting the range of downstream experimental techniques that could be applied. Nonetheless, a series of groundbreaking studies in the field reported that the cleavage-specific transcription factor Dux was necessary and sufficient to orchestrate the reprogramming of ESCs toward 2C-like cells in a highly efficient and coordinated manner (De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017).

We then exploited the Dux-dependent 2C-like system to profile the chromatome of the different cellular states appearing during the $2C^{+}$ reprogramming. It is important to highlight that

we isolated millions of luciferase, $2C^-$ and $2C^+$ cells to perform chromatin proteomics, which would have been inconceivable in the absence of Dux activation. Ultimately, we used these data to uncover regulators controlling chromocenter remodeling, leading to the identification of the chromatin remodeler SMARCAD1, which in fact influenced H3K9me3 levels in mouse embryos.

Assessing totipotency or extended potential of early embryo-derived stem cell types is not a trivial task (Torres-Padilla 2020; Posfai et al. 2021). Indeed, there are several attempts ongoing in the stem cell field to define criteria to evaluate these totipotent or extended potential features since there are growing embryo-derived stem cell systems exhibiting many of these characteristics (Posfai et al. 2021). Some of the experimental approaches that have been performed to investigate potential totipotency are mouse chimera assays (Macfarlan et al. 2012), somatic cell nuclear transfer (Ishiuchi et al. 2015) and cell-based aggregations to form blastocyst-like structures (Baker and Pera 2018; Rivron et al. 2018), among others. Along these lines, endogenous 2C-like cells were associated with expanded potency in chimeric embryos toward extraembryonic lineages (Macfarlan et al. 2012). Similarly, 2C-like cells generated after downregulation of CAF-1 exhibited higher potential to be reprogrammed with respect to ESCs upon nuclear transfer (Ishiuchi et al. 2015). Even *in vivo* reprogrammed iPS cells showed totipotency features since they showed an enhanced ability to differentiate to trophoblast

stem cells *in vitro*, contributed to the trophectoderm lineage *in vivo*, and generated embryo-like structures that expressed embryonic and extraembryonic markers when injected *in vivo* (Abad et al. 2013). However, to date, there has not been any *in vivo* evidence supporting totipotency or extended potential features in Dux-derived 2C-like cells.

We took up the challenge of assessing the *in vivo* potency of Dux-derived 2C-like cells. We identified that 2C-like cells generated via Dux overexpression predominantly contributed to extraembryonic tissues *in vivo* in chimeric embryos. It should be stated that we did not observe an equal contribution toward ICM and TE (or fetus and placenta for post-implantation embryos) in the $2C^+$ DsRed injected embryos, as one would have expected for totipotent cells. What we observed was a drastic tendency toward contributing to extraembryonic TE or placentas. Nevertheless, in a minority of the cases, $2C^+$ cells differentiated toward both fetal and placental fates *in vivo*, suggesting the acquisition of totipotent-like or expanded potential features (Abad et al. 2013; Yang et al. 2017a; Yang et al. 2017b). Another important aspect that prompted us to analyze the chimeric embryos in post-implantation stages, meaning in a more stringent scenario, was the mild and somewhat fragmented DsRed signal localized at the TE in $2C^+$ DsRed injected embryos. We were concern of the viability of those DsRed expressing cells located in the outer epithelium of the blastocysts since, recently, the laboratory of Verena Ruprecht reported that the surface epithelium of zebrafish as

well as mouse embryos performs efficient phagocytic clearance of apoptotic cells (Hoijman et al. 2021). However, we robustly demonstrated that the progeny of the injected $2C^+$ DsRed cells engrafted and contributed to the tissue, as confirmed by immunostaining against the DsRed protein.

Interestingly, we demonstrated that $2C^+$ cells not only showed a high tendency to colonize mostly the placenta but also expressed placental markers of spongiotrophoblasts and trophoblast giant cells, which share the extraembryonic ectoderm as common progenitor (Simmons and Cross 2005). These results could indicate that $2C^+$ cells acquire prevalently a trophoblast stem cell-like fate, which is in accordance with the observed bias toward extraembryonic lineages. We propose that Dux-derived 2C-like cells dedifferentiate, since they did not behave as luciferase cells (pluripotent stem cells) *in vivo*, although deviating to trophoblast fate. Very infrequently, $2C^+$ cells achieve some degree of totipotency. In any case, additional molecular and *in vivo* experiments should be performed in the near future to fully comprehend the mechanistic details behind the extraembryonic bias and totipotent capabilities observed in $2C^+$ cells. Additionally, it would be really interesting to explore the reported features of 2C-like cells in self-organizing embryo-like structures (Harrison et al. 2017; Rivron et al. 2018; Sozen et al. 2018), and evaluate whether these cells can form *in vitro* blastoids such as those described by Nicolas Rivron and colleagues. These data could also serve as a call of caution and reflection before deriving

claims using this, or similar, embryo-derived stem cell systems without an in-depth characterization of the *in vitro* model. At the same time, we should agree that these systems are the best tools we currently have to dissect complex processes restricted to early developmental phases, despite their drawbacks and incomplete fidelity to the *in vivo* scenario.

Using DNA-mediated chromatin proteomics, we profiled the dynamic changes occurring in the chromatin-bound proteome (chromatome) during 2C-like cell reprogramming and identified factors potentially involved in chromocenter reorganization. In particular, the technology we deployed in this work, DNA-mediated chromatin pull-down, has been recently set up and used by the laboratory of Luciano Di Croce to reveal a new regulatory axis between the metabolic state and proliferation of pluripotent stem cells (Aranda et al. 2019), showing the potential of exploring the chromatome. Unlike previous reports that focused exclusively on transcriptional changes (Rodriguez-Terrones et al. 2018; Fu et al. 2019; Fu et al. 2020), our study exploited chromatin proteomics by genome capture to unravel an additional layer of information and complexity in the 2C-like system. Thus, we provided a detailed characterization of the stepwise chromatome dynamics occurring during the 2C-like state transition. In brief, we captured with high confidence many of the drivers of the zygotic transcriptional program and, therefore, of the 2C-like state, enriched in the $2C^+$ chromatome, such as members of the ZSCAN4 (Zinc finger and SCAN domain containing 4)

family of proteins (Falco et al. 2007; Ishiuchi et al. 2015; De Iaco et al. 2017; Hendrickson et al. 2017), the maternal negative elongation factor A (NELFA) (Hu et al. 2020), and the developmental pluripotency-associated factor 2 (DPPA2) (De Iaco et al. 2019; Eckersley-Maslin et al. 2019; Yan et al. 2019). More importantly, we identified novel potential candidate factors controlling the remodeling of chromocenters like the DNA topoisomerase 2-binding protein 1 (TOPBP1), which was shared by the $2C^-$ and Luc chromatomes but not enriched in the $2C^+$ chromatome.

After profiling the chromatome landscape along the $2C^+$ reprograming and identifying a large and dynamic redistribution in terms of chromatin association of transcriptional regulators and epigenetic factors, we decided to focus on dissecting H3K9me3 reorganization. We demonstrated that H3K9me3-marked heterochromatin foci in 2C-like cells generated via Dux overexpression became larger and decreased in number during the reprogramming of ESCs to 2C-like cells. Interestingly, it has been previously reported that transient bursts of Zscan4 expression in ESCs involved derepression of heterochromatin regions (Akiyama et al. 2015), similar to what we found occurring in $2C^+$ cells. In fact, this report postulated that the rapid heterochromatin derepression observed after Zscan4 expression, which mimics quite faithfully the H3K9me3 remodeling event we reported, might be the basis for maintaining the extraordinary degree of genome stability seen in ESCs. The authors attributed this to a

resetting of the heterochromatin status, which was also somehow speculated by Azim Surani and Julia Tischler in 2012 with respect the spontaneous and fluctuating nature of endogenous 2C-like cells (Surani and Tischler 2012). In reality, they went further proposing that, perhaps, the transition of ESCs toward the 2C-like state might be crucial for resetting the whole epigenome, for the repair and maintenance of telomeres, and for refreshing the core genetic network underlying pluripotency (Surani and Tischler 2012).

Remarkably, we also reported that $2C^+$ cells could undergo the reverse transition, exiting the 2C-like state and re-entering pluripotency. In this particular window, we demonstrated that the chromocenters re-formed upon transition of 2C-like cells into ESC-like cells. We here illustrated how 2C-like cells retained the ability to re-establish chromocenters reassembling ESCs and how, in particular, monitoring the $2C^+$ exit could serve as an excellent platform to address the question of whether the epigenome is reset and/or maintained upon several rounds of fluctuation in the near future. We further envision that single-cell multi-omics, low-cell number and low-input epigenomic technologies might be of utmost importance to analyze these changes since we will encounter an asynchronous, scarce sub-population achieving full ESCs conversion after the $2C^+$ exit as we preliminary demonstrated (Clark et al. 2016; Kelsey et al. 2017; Chappell et al. 2018; Argelaguet et al. 2019).

Since chromatin proteomics captured terms associated with cell cycle regulation, we identified that 2C-like cells exhibited an abnormal cell cycle profile with an accumulation in the G2/M phase. We also identified that that cell cycle arrest in G2/M and S boosted the reprogramming efficiency from ESCs to 2C-like cells. The prolonged G2/M phase of the cell cycle observed in $2C^+$ cells resembles 2-cell embryos that have an exceptionally long G2 phase (10-12 h), which coincides with the major zygotic genome activation (ZGA) and the initiation of chromocenter compaction (Ciemerych and Sicinski 2005; Probst and Almouzni 2011). In the *in vivo* scenario, this prolonged G2 phase might be necessary to rewire specific epigenetic modifications in the 2-cell blastomeres to allow heterochromatin formation. This is a key step before the blastomeres can embark into the correct developmental process, as proposed for early Drosophila embryos (Seller et al. 2019).

Topoisomerases are topological enzymes which disentangle sister chromosomes after DNA replication (Wang 2002). The chromatin of $2C^+$ cells was depleted of the DNA topoisomerase 2-binding protein 1, TOPBP1, suggesting that either these cells are prone to chromosome instability or they are arrested in a quiescent state where the chromatin can be remodeled and heterochromatin can be established simultaneously to the association of topoisomerases to the chromosomes. In yeast, topoisomerases likely cooperate with the chromatin remodeling factor SMARCAD1 (Bantele et al. 2017), which we

show here essential for mouse embryo development. We also showed that topoisomerase inhibition led to an increase in the fraction of 2C-like cells. Thus, it is also tempting to speculate that cell cycle progression, especially since we observed that $2C^+$ cells might be arrested in the G2/M phase, has a role in regulating SMARCAD1 recruitment and/or function on chromatin during the $2C^+$ exit. Besides, the yeast and human SMARCAD1 homologs (Fun30/SMARCAD1) were reported to be involved in the DNA double-strand break response (Chen et al. 2012; Costelloe et al. 2012). High DNA damage is caused by replication-induced stress during early development (Zeman and Cimprich 2014), and it has been postulated that activation of the ATR-mediated response to replication stress generates ESCs with expanded fate potential (Atashpaz et al. 2020). Thus, it is tempting to speculate that DNA repair mediated by SMARCAD1 has a role in promoting exit of $2C^+$ state, while cells are arrested in the G2/M phase. To further confirm or refute this possibility, additional data might be needed to elucidate the role of SMARCAD1 in the intersection between DNA repair, cell cycle arrest and early embryogenesis.

Finally, we identified the chromatin remodeler SMARCAD1 to be associated with H3K9me3 in heterochromatin foci in ESCs. This was not the case of TOPBP1, even if we identified TOPBP1 associated with SMARCAD1, probably suggesting a regulatory role over the chromatin remodeler. Additionally, in the small-content loss-of-function CRISPR-Cas9 screening,

we reported that Smarcad1 depletion resulted in a pronounced alteration of the $2C^+$ exit kinetics, indicating an impairment to differentiate to ESC-like cells. In fact, this also suggest that it is during the $2C^+$ to ESC transition, which reflects the developmental trajectory toward pluripotent, the moment where SMARCAD1 might exert its function. Supporting the specificity of SMARCAD1 during the $2C^+$ exit, we could not observe major defects either on the spontaneous 2C-like fluctuation or in the Dux-induced $2C^+$ cells upon SMARCAD1 reduction. Precisely, the association between SMARCAD1 and H3K9me3 was lost upon entry of ESCs in the 2C-like state, although SMARCAD1 nuclear localization was recovered after 2C-like state exit.

Finally, we tested whether our results in the $2C^+$ to ESC transition were reproducible *in vivo* and, thus, if SMARCAD1 held any role in early mouse development. Depletion of SMARCAD1 induced mouse embryo developmental arrest and was accompanied by a partial loss of H3K9me3 in the embryos, suggesting a contribution role of SMARCAD1 activity in *de novo* heterochromatin formation during early development. Remarkably, we identified the chromatin remodeler factor SMARCAD1 to be essential for embryo development. Depletion of SMARCAD1 in preimplantation embryos led to a substantial reduction of H3K9me3 levels. These findings have important implications because the establishment of heterochromatin foci during embryo development is a key step to instruct the embryonic totipotent

program of the 2-cell stage toward pluripotency, ultimately leading the embryo to undergo lineage differentiation (Zernicka-Goetz et al. 2009; Burton and Torres-Padilla 2014).

Endogenous retroviruses (ERVs) are transposable elements flanked by long terminal direct repeats (LTRs) (Friedli and Trono 2015; Rodriguez-Terrones and Torres-Padilla 2018). Tight control of ERVs and their transposable activity is essential for genome integrity and play an important role in early development and pluripotency (Friedli and Trono 2015; Rodriguez-Terrones and Torres-Padilla 2018). H3K9me3 has been associated with retrotransposons through the KRAB-associated protein 1, KAP1 (Rowe et al. 2010). KAP1 led to the silencing of ERVs in ESCs by inducing H3K9me3 heterochromatin formation via the recruitment of the H3K9 histone methyltransferase SETDB1 (Wolf and Goff 2007; Matsui et al. 2010; Rowe et al. 2010). SMARCAD1 was recently discovered to directly interact with KAP1 by the laboratory of Jacqueline Mermoud and, therefore, be an important regulator of the KAP1-SETDB1 silencing complex in ESCs (Sachs et al. 2019). SMARCAD1 is a SWI/SNF-like chromatin remodeler and a key factor for ERV silencing in ESCs (Sachs et al. 2019). Additionally, the laboratory of Deborah Bourc'his has recently identified SMARCAD1 as one of their top hits in a CRISPR screen for IAP suppressors in ESCs (Chelmicki et al. 2021). Our observation that SMARCAD1 enriches in H3K9me3 heterochromatin foci during the transition from the 2C-like state to pluripotency and that it

is essential for early mouse embryo development is aligned with the observations previously reported in ESCs. It will be interesting in the future to study whether SMARCAD1 can tether the KAP1-SETDB1 to directly induce the formation of H3K9me3 heterochromatin foci at the exit of the 2-cell stage in the embryos.

Recently, the H3K9 histone methyltransferase SUV39H2 has been reported to catalyze *de novo* H3K9me3 in the paternal pronucleus after fertilization (Burton et al. 2020). Yet, Suv39h2 downregulation in zygote-stage embryos did not translate on appreciable changes in H3K9me3 levels on the maternal chromatin. This opens up the possibility that different methyltransferases, and their regulators like SMARCAD1, could be responsible for H3K9me3 acquisition in this early developmental stage. It will be interesting in the future to study whether SMARCAD1 can tether the KAP1-SETDB1 to directly induce the formation of H3K9me3 heterochromatin foci at the exit of the 2-cell stage in the embryos. Interestingly, the BRG1-associated factor (BAF) chromatin remodeler complex has been reported to interact with topoisomerase II alpha and to facilitate DNA decatenation (Dykhuizen et al. 2013). SMARCAD1 might be the remodeler factor that, by either interacting with topoisomerases or mediating the response to DNA double-strand breaks, can facilitate DNA decatenation or DNA repair at the exit of the totipotent 2-cell stage when heterochromatin is established *de novo*.

In conclusion, by using chromatin proteomics, we have provided additional data that will help to elucidate the molecular intricacies of the 2C-like state and early mammalian development. In the current Thesis, we focused on heterochromatin establishment and we identified SMARCAD1, which might be the remodeler factor that, by regulating methyltransferases, can facilitate H3K9me3 deposition at the exit of the totipotent 2-cell stage when heterochromatin is established *de novo* (Figure 53).

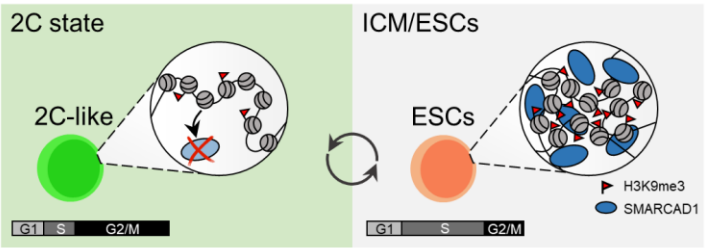

Figure 53. SMARCAD1 contributes to heterochromatin establishment at the transition from the 2C-like to the pluripotent state. 2C-like cells generated via Dux overexpression lack SMARCAD1, which we have shown to be necessary for mouse embryo development. Indeed, SMARCAD1 downregulation in embryos is associated with a partial H3K9me3 reduction. Remarkably, the 2C-like conversion is characterized by the remodeling of H3K9me3 heterochromatic regions. 2C-like cells have a rewired cell cycle profile. Finally, different cell cycle perturbations boost the 2C-like reprogramming.

CONCLUSIONS

Conclusions

The following conclusions can be drawn from the results presented in this PhD Thesis:

1. 2C-like cells generated via Dux overexpression predominantly contribute to extraembryonic tissues *in vivo* and express placental markers in chimeric post-implantation embryos.
2. 2C-like cells generated via Dux overexpression exhibit totipotent-like features, although with very low frequencies.
3. Chromatin-bound proteome profiling captures dynamic chromatome changes during 2C-like cell reprogramming, identifying known as well as novel regulators of 2C-like cells and early embryonic development.
4. Chromatome profiling of the 2C-like cell reprogramming identifies TOPBP1 and SMARCAD1 as candidate factors controlling the remodeling of chromocenters.
5. Entry in the 2C-like state is characterized by the remodeling of H3K9me3 heterochromatic regions, as identified by reduction in the number of H3K9me3 foci and enlargement of their respective heterochromatic areas.
6. Super-resolution imaging shows DNA decompaction of the chromatin fiber of 2C-like cells.
7. 2C-like cells undergo the reverse transition, exiting the 2C-like state and re-entering the pluripotent state, as exemplified by the efficient downregulation of MERVL elements and Dux gene.

8. H3K9me3 heterochromatin becomes rapidly formed following exit from the 2C-like state, as identified by the resetting of the number and areas of H3K9me3 foci.
9. 2C-like cells are a suitable model to investigate *de novo* chromocenter reorganization, among other key features of early embryogenesis.
10. 2C-like cells generated via Dux overexpression exhibit a rewired cell cycle profile with an accumulation in the G2/M phase and a reduction in the S phase, although the intermediate $2C^-$ cells have a comparable cell cycle profile to that of pluripotent stem cells.
11. Cell cycle arrest in G2/M and S phases mediated by different chemical perturbations boosts the reprogramming efficiency from ESCs to 2C-like cells.
12. Small-content CRISPR-Cas9 loss-of-function screening identifies the SWI/SNF-like remodeler SMARCAD1 as a contributing factor in the $2C^+$ to ESC-like transition.
13. SMARCAD1 depletion results in no major impact in the $2C^+$ conversion in either the endogenous fluctuation or the Dux-mediated system.
14. SMARCAD1 associates with H3K9me3 in ESCs and $2C^-$ cells and with TOPBP1 in ESCs, whereas TOPBP1 does not directly associate with H3K9me3 in ESCs.
15. SMARCAD1 and TOPBP1 are expressed during preimplantation development, with a peak of expression at the 4-celll stage.
16. SMARCAD1 nuclear localization is lost in the 2C-like state.

17. SMARCAD1 is necessary for early embryo development, whereas RBBP7 is dispensable for developmental progression.
18. SMARCAD1 localizes exclusively in the nucleus of mouse preimplantation embryos.
19. SMARCAD1 downregulation in mouse embryos is associated with a partial H3K9me3 reduction.

PART IV - MATERIALS and METHODS

Cell culture

E14Tg2a mouse ESCs were cultured in gelatinized plates in high glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 15% fetal bovine serum (FBS) (Sigma), 1x GlutaMAX, 1x sodium pyruvate, 1x minimum essential medium non-essential amino acids (MEM NEAA), 1x penicillin/streptomycin, 100 µM 2-mercaptoethanol, 1000 U/ml mouse leukemia inhibitory factor (mLIF) (ESG1107, Millipore), 1 µM Mek inhibitor (PD0325901, Selleck Chemicals) and 3 µM GSK3 inhibitor (CHIR99021, Selleck Chemicals).

After viral infection, ESCs were selected and maintained with ES medium containing the appropriate combination of selection drugs (250 µg/ml Geneticin (G418; 10131035, Life Technologies), 0.5 µg/ml Puromycin (A1113803, Life Technologies)). ESCs were treated with 2 µg/ml doxycycline (D9891, Sigma) for 24 hours to induce Dux expression.

E14 ESCs containing the 2C::EGFP reporter were a gift from M.-E. Torres-Padilla. Independent E14 ESC clone containing the Dux-CA overexpressing cassette were a gift from B.R. Cairns.

To inhibit HDACs, ESCs were treated with Trichostatin A (TSA) (T8552, Sigma) for 24 hours at the concentration indicated in each figure. For synchronization into the G2/M cell cycle phase, ESCs were treated with 100 ng/ml of nocodazole (M1404, Sigma) for 6, 12 or 16 hours. For modulation of the

DNA replication fork speed, ESCs were treated with 100 µM or 200 µM of hydroxyurea (H8627, Sigma) for 12 or 16 hours. To inhibit DNA topoisomerases, ESCs were treated with 500 nM of the topoisomerase I inhibitor camptothecin (CPT; ab120115, Abcam) and/or with 5 µM of the topoisomerase II inhibitor ICRF-193 (I4659, Sigma) for 12 hours. For effective inhibition of p38 MAPK, ESCs were cultured in medium supplemented with 2 µM of SB203580 (hydrochloride) (72222, STEMCELL Technologies) for more than 4 passages before performing the experiment.

Lentivirus production and ESC infection

Lentiviral particles were produced following the RNA interference Consortium (TRC) instructions for viral production and cell infection (http://www.broadinstitute.org/rnai/public/).

At day 0.8×10^6 HEK293T cells were plated in 150 mm plate. The day after plating, HEK293T cells were co-transfected with 30 µg of the lentiviral plasmid of interest (pCW57.1-Luciferase or pCW57.1-mDux-CA), 19.5 µg pCMV-ΔR8.9, and 10.5 µg pCMV-VSV-G, using the calcium phosphate transfection kit (631312, Clontech). pCW57.1-Luciferase and pCW57.1-mDux-CA were a gift from Stephen Tapscott (Addgene plasmids #99283 and #99284).

The day after transfection, the HEK293T culture medium was substituted with the previously mentioned ESC culture medium

this time supplemented with 30% FBS. Then, 3×10^6 ESCs were plated onto gelatin-coated 100 mm plates the day before transduction. The lentiviral-containing medium was harvested from HEK293T cells at 48 and 72 hours after transfection, filtered, and used for ESC infection. Two days after the last round of infection, ESCs were selected with the indicated concentration of selection drug (see Cell culture).

Fluorescence-activated cell sorting (FACS)

Quantification of GFP positive cells, and cell cycle analysis was performed with a LSR II Analyzer (BD Biosciences). For cell sorting, an Influx Cell Sorter (BD Biosciences) was used to sort the specified populations in each experiment.

Flow cytometry analysis

For flow cytometry analysis, cultured ESCs were detached with trypsin-EDTA solution (T4174, Sigma) and collected by centrifugation at 300 rcf for 5 minutes. They were resuspended in ES medium or PBS with DAPI (5 mg/ml).

Cell cycle analysis by flow cytometry

For cell cycle analysis of live cells, 5×10^4 ESCs were plated per well in gelatin-coated 6-well plates one day before starting the experiment. At the moment of the assay, ESCs were trypsinized with trypsin-EDTA solution (T4174, Sigma) and collected by centrifugation at 300 rcf for 5 minutes. Cells were

washed with PBS before incubation with ES medium supplemented with 10 µg/ml Hoechst 33342 (H1399, Thermo Fisher) for 30 minutes at 37 ºC. Propidium iodide (PI) (1 µg/ml, P4864, Sigma) was added to stain dead cells.

All flow cytometry data were processed and analyzed with FlowJo (v10).

Immunostaining, image processing and quantification

Immunofluorescence (IF) staining of ESCs for confocal and STORM imaging

ESCs were plated at a concentration of 56.000 cells/cm^2 in gelatin-coated borosilicate glass bottom Nunc Lab-Tek (155411) or µ-Slide Ibidi (80827) 8-well chambers. Cells were washed with PBS, fixed with 4% paraformaldehyde (PFA) for 10 minutes, permeabilized and blocked (10% goat serum (GS), 2.5% bovine serum albumin (BSA), 0.4% Triton X-100) for 30 minutes at room temperature. Incubation with the corresponding primary antibodies lasted 3 hours at 37 ºC. Cells were then washed with PBS and incubated with Alexa Fluor (Molecular Probes, Invitrogen) secondary antibodies for 1 hour at room temperature. Finally, cells were washed three times with PBS containing DAPI (5 mg/ml) for nuclear counterstain.

Images were acquired on a Leica TCS SP5 confocal microscope equipped with a 63x oil objective.

The following antibodies were used: chicken anti-GFP (1:500; ab13970, Abcam), mouse anti-Oct-3/4 (1:200; sc-5279, Santa Cruz), rabbit anti-histone H3K9me3 (1:500; ab8898, Abcam), mouse anti-SMARCAD1 (1:500; ab67548, Abcam), anti-chicken Alexa Fluor 488, anti-mouse Alexa Fluor 568, anti-rabbit Alexa Fluor 568, anti-mouse Alexa Fluor 647. All secondary antibodies were provided by Molecular Probes (Invitrogen) and used 1:1,000 in PBS. Full details of each antibody used for IF can be found in Table 1.

Table 1 – List of antibodies

Primary Antibody	Host	Source	Catalog no.	Use	Dil.
Oct-3/4	Ms	Santa Cruz	sc-5279	IF	1:200
Proliferin	Ms	Santa Cruz	sc-271891	IF	1:200
SMARCAD1	Ms	Abcam	ab67548	IF WB coIP	1:500 1:500 4 µg
Vinculin	Ms	Merck	V-9131	DB	1:1000
IgG	Ms	Jackson Immunoresearch	015-000-003	coIP	4 µg
Cell cycle cocktail	Rb	Abcam	ab136810	WB	1:250
DsRed	Rb	Clontech	632496	IF	1:200
Histone H3	Rb	Abcam	ab1791	DB WB	1:2000
H3K9me3	Rb	Abcam	ab8898	IF WB coIP	1:500 1:500 4 µg
H3K27ac	Rb	Abcam	ab177178	WB	1:5000
TOPBP1	Rb	Abcam	ab2402	WB coIP	1:1000 4 µg
TPBPA	Rb	Abcam	ab104401	IF	1:200
IgG	Rb	Jackson Immunoresearch	011-000-003	coIP	4 µg

GFP	Ck	Abcam	ab13970	IF	1:500
Secondary Antibody					
Anti-Chicken AF488	Gt	Molecular Probes	A11039	IF	1:1000
Anti-Mouse AF568	Gt	Molecular Probes	A11031	IF	1:1000
Anti-Rabbit AF568	Gt	Molecular Probes	A11036	IF	1:1000
Anti-Mouse AF647	Gt	Molecular Probes	A21235	IF	1:1000
Alexa Fluor 647 Azide		Molecular Probes	A10277	CC	10 mM
Anti-Mouse IgG HRP	Sh	GE Healthcare	NA931	WB	1:1000
Anti-Rabbit IgG HRP	Dy	GE Healthcare	NA934	WB	1:2000

Dy, Donkey; Ck, Chicken; Gt, Goat; Ms, Mouse; Rb, Rabbit; Sh, Sheep.

DB, Dot blot; CC, Click chemistry; coIP, Co-immunoprecipitation; HRP, Horseradish peroxidase; IF, Immunofluorescence; WB, Western blot.

EdC incorporation and DNA labelling

To label DNA, a 14 hours incorporation pulse of 5-ethynil-2'-deoxycytidine (EdC; T511307, Sigma) at 2.5 µM was performed in ESCs, in parallel to doxycycline treatment. Cells were plated in gelatin-coated borosilicate glass bottom chambers at a concentration of 56.000 cells/cm^2 in ES medium supplemented with EdC for 14 hours. At the end of EdC incorporation, ESCs were fixed with PFA 4 % (43368, Thermo Fisher Alfa Aesar) for 10 minutes at room temperature and rinsed with PBS three times for 5 minutes each. Cells were permeabilized with 0.4 % Triton X-100 in PBS for 15 minutes and rinsed with PBS three times for 5 minutes each.

Click chemistry reaction was performed by incubating cells for 30 minutes at room temperature in Click chemistry buffer (100 mM Hepes pH 8.2, 50 mM Amino Guanidine (396494, Sigma), 25 mM Ascorbic Acid (A92902, Sigma), 1 mM $CuSO_4$, 2 % Glucose (G8270, Sigma), 0.1 % Glox solution (0.5 mg/ml glucose oxidase, 40 mg/ml catalase (G2133 and C100, Sigma)) and 10 mM Alexa Fluor 647 Azide (A10277, Thermo Fisher)) (Zessin et al. 2012; Raulf et al. 2014; Otterstrom et al. 2019). After washing the samples three times with PBS, we directly proceeded to perform STORM imaging.

STORM imaging

Stochastic Optical Reconstruction Microscopy (STORM) imaging was performed on a N-STORM 4.0 microscope (Nikon) equipped with a CFI HP Apochromat TIRF 100x 1.49 oil objective and a iXon Ultra 897 camera (Andor) with a pixel size of 16 µm. This objective/camera combination provides an effective pixel size of 160 nm. STORM images were acquired with 10 msec exposure time for 60000 frames using highly inclined (HILO) illumination. An activator/reporter pair strategy was used with AF405 and AF647 fluorophores, respectively.

Continuous imaging acquisition was performed with simultaneous 405 nm and 647 nm illumination. 647 nm laser was used at constant ~2 kW/cm^2 power density. 405 nm laser was used at low laser power and gradually increased during the imaging to enhance fluorophore reactivation and to maintain the density of localizations per frame constant. Before

STORM imaging, we acquired conventional fluorescence images of GFP for each nucleus to discriminate between $2C^-$ and $2C^+$ cells. Imaging buffer composition for STORM imaging was 100 mM Cysteamine MEA (30070, Sigma), 1 % Glox Solution and 5 % Glucose (G8270, Sigma) in PBS.

STORM images were analyzed and rendered in Insight3 as previously described (Rust et al. 2006; Bates et al. 2007). Localizations were identified based on an intensity threshold and the intensity distribution of their corresponding Point Spread Functions (PSFs) fit with a 2D Gaussian to determine the x-y positions of their centers with high accuracy (~20 nm).

Voronoi Tesselation analysis

For Voronoi Tesselation analysis, we used the list of localization from STORM (Levet et al. 2015; Andronov et al. 2016) and then we used a previously developed custom-made Matlab script (Otterstrom et al. 2019). XY coordinates of the localizations were used to generate the Voronoi polygons. Local densities were defined as the inverse value of the area of each Voronoi polygon. For visualization, we color-coded each Voronoi polygon based on their area, from yellow for the smallest polygons (density > 0.01 nm^{-2}) to blue for larger polygons (density < 0.0001 nm^{-2}). Finally, the largest 0.5 % of polygons were set to black. For each nucleus, we computed the mean Voronoi density (nm^{-2}) as a measure of global DNA compaction.

For the GFP intensity score, we quantified the GFP conventional images (488 nm channel) with lower intensities in order to assign a GFP intensity score to each nucleus. We summed the fluorescence intensity ADU counts inside each nucleus and divided it by the total number of pixels to obtain the average GFP intensity. Then, we used the distribution of GFP intensities from the different nuclei to normalize the values, obtaining a GFP intensity score ranging from 0 (less bright) to 1 (most bright). We then performed a cell-by-cell analysis of the relation between GFP intensity score and global chromatin compaction obtained from Voronoi Tesselation analysis.

Fixing, sectioning and immunofluorescence of tissue cryosections

Intact fetuses and placentas were fixed by immersion in 4 % PFA for 45 minutes at 4 °C. After 30 minutes wash with PBS, samples were subjected to a sucrose gradient following 15% sucrose for 30 minutes, 20% sucrose for 1 hour and 30% sucrose overnight. Optimal cutting temperature (OCT) compound was used to embed fetal and placental samples before frozen sectioning. Serial transversal sections of 5 µm of thickness were prepared and processed for immunofluorescence.

For the immunofluorescence, sections were thawed and then placed in a plastic rack with permeabilization buffer containing 0.3 % Triton X-100 and 10 mM sodium citrate in PBS for 1 hour

at room temperature. Antigen retrieval was performed by boiling the slides for 10 minutes. After a wash with cold water, sections were then blocked for 45 minutes (3 % bovine serum albumin (BSA), 300 µM glycine, 0.1 % Tween-20 in PBS). They were incubated with the corresponding primary antibodies diluted in PBS, 1.5 % BSA for two consecutive overnights at 4 ºC. On the following day, slides were washed with PBS and incubated with secondary antibodies for 2 hours at room temperature. Cryosections were mounted with Vectashield with DAPI (Vector Laboratories) and imaged using a Leica TCS SP5 confocal microscope equipped with a 10x air objective.

The following antibodies were used: rabbit anti-DsRed (1:200; 632496, Clontech), rabbit anti-TPBPA (1:200; ab104401, Abcam), mouse anti-Proliferin (1:200; sc-271891, Santa Cruz), anti-rabbit Alexa Fluor 568 and anti-mouse Alexa Fluor 647. All secondary antibodies were provided by Molecular Probes (Invitrogen) and used 1:1,000 in PBS.

Immunofluorescence of pre-implantation embryos

Preimplantation embryos at 2-cell (E1.5) and 8-cell stages (E2.5) were fixed with 2 % PFA for 10 minutes at room temperature, permeabilized (0.25 % Triton X-100) for 10 minutes, and then blocked (3 % BSA) for 1 hour at 37 ºC. Incubation with the corresponding primary antibodies at the indicated dilutions in 1 % BSA lasted one overnight at 4 ºC. After washing, embryos were incubated with Alexa Fluor (Molecular Probes, Invitrogen) secondary antibodies diluted in

1 % BSA for 1 hour at 37 ºC. Finally, embryos were washed and transferred to imaging buffer containing DRAQ5 (1:500; 62251, Thermo Fisher) for DNA staining. Images were acquired on a Leica TCS SP8 STED3X confocal microscope equipped with a 63x oil objective.

The following antibodies were used: rabbit anti-histone H3K9me3 (1:500; ab8898, Abcam), mouse anti-SMARCAD1 (1:250; ab67548, Abcam), goat anti-rabbit Alexa Fluor 488 and goat anti-mouse Alexa Fluor 488. All secondary antibodies were provided by Molecular Probes (Invitrogen).

Image processing and quantification

Immunofluorescence images were processed and analyzed with the ImageJ software (https://imagej.net/download/). All immunofluorescence images were acquired with z-stacks. Z-stacks were projected using the *maximum intensity* z-projection type.

For DsRed-proliferin and SMARCAD1-H3K9me3 co-immunofluorescence images, a Gaussian blur filtering ($\sigma = 0.5$) was applied to the DsRed, proliferin and SMARCAD1 channels. For DsRed-proliferin images, background was subtracted according to mean background intensity. H3K9me3 foci areas were analyzed using the 3D Object Counter function (https://imagej.net/3D_Objects_Counter, ImageJ). Co-localization analysis was done using the JACoP plugin (https://imagej.net/JaCoP, ImageJ). Manders' coefficient was

calculated with the JACoP plugin. Manders' coefficient was used as co-localization indicator because of its independence of the intensity of the overlapping pixels.

For the quantification of H3K9me3 and SMARCAD1 fluorescence intensities in preimplantation embryos, manual selection of the nuclear area was performed for each blastomere. Fluorescent signals were measured and then normalized by the average cytoplasmic signal (background) in each condition. For the normalization step, the fluorescence intensity of a squared shape of equal size was taken for each individual blastomere.

Time-lapse microscopy

For time-lapse live imaging, 30,000 cells were plated on a pre-gelatinized 35-mm glass-bottom dish (MatTek Corporation) and were left to attach for at least 4 hours before imaging. Cells were imaged on an Andor Revolution XD confocal microscope (inverted, Olympus) featuring a spinning disk unit (CSU-X1 Yokogawa) and an Andor iXon 897E Dual-Mode electron multiplying charge-coupled device camera with a 60× oil-immersed objective. Imaging was performed at 37 °C and 5 % CO_2. Differential interference contrast (DIC) images were taken with an exposure time of 80 ms. Cells were excited with 488 nm and 561 nm lasers with the Acousto-Optical Tunable Filter (AOTF) set at 15 % and with a maximum exposure time

of 120 ms. Fourteen z-stacks were obtained for each pre-assigned XY position, corresponding to section thickness of 13 µm. Consecutive acquisitions were taken each 35 minutes for a total time-lapse duration of 24 hours.

All time-lapse images were analyzed using ImageJ. Z-stacks were merged using the *concatenate* function and projected using the *maximum intensity* z-projection type.

RNA extraction and quantitative real-time PCR (qRT-PCR)

RNA was extracted from pelleted or sorted ESCs using the RNA isolation RNeasy Mini kit (74106, Qiagen), according to the manufacturer's protocol. RNA was reverse-transcribed with iScript cDNA Synthesis kit (1708891, Bio-Rad).

qRT-PCR reactions were performed using LightCycler 480 SYBR Green I Master (4887352001, Roche) in a LightCycler 480 (Roche) instrument, according to the manufacturer recommendations. qRT-PCR data was normalized to *Gapdh* or β-*actin* expression. For each sample, we had at least a technical duplicate. The primers used are listed in Table 2.

Table 2 – List of primers used for quantitative real-time PCR (qRT-PCR)

Gene	Forward (5' to 3')	Reverse (5' to 3')
Dux	GGAGAAGAGATACCTGAGC TTCAA	AATCTGAGACCCCCATTCG
MERVL	CTCTACCCACTTGGACCATA TGAC	GAGGCTCCAAACAGCATCT CTA
MajSat	GCACACTGAAGGACCTGGA ATATG	GATTTCGTCATTTTTCAAGT CGTC
LINE1	GGACCAGAAAAGAAATTCCT CCCG	CTCTTCTGGCTTTCATAGTC TCTGG
Eif1a	AACAGGCGCAGAGGTAAAA A	CTTATATGGCACAGCCTCCT
Zscan4	GAGATTCATGGAGAGTCTG ACTGATGAGTG	GCTGTTGTTTCAAAAGCTTG ATGACTTC
Oct4	CGTGGAGACTTTGCAGCCT G	GCTTGGCAAACTGTTCTAGC TCCT
Smarcad1	AAATTCAGCAAAGACACAGT GATT	CAGAAGGAAGGTCATGGGA TT
Cbx5	CAGGCCTTAGCGTGAGTGA T	GCCTGTTGATCCACCTGAA G
Incenp	AGGCCTCTGCTCGAATCAT	CAACATTCTGAGGCACTTCA AC
Gatad2b	GGCTGAAAAACGCTTTCGT	TGCTGCTGTAATCGCTGTTC
Cct2	GGAATCCGTCGGAACCAT	GCTGTCTCGGCCCTCTCT
Cct5	CTGTTTGCACAAGGGCAGT A	CCATCTGGGTGGCAAGAG
Hells	ACCCTTTCAACAACCAAAGC	CCATTTTCCCAAAGCATCC
Hspd1	CAGAGCTGGGTCCCTCACT	CTGTGGGTAGTCGAAGCAT TT
Orc2	TCATGTTGTCATCAACGGCT A	AGGACATCTTCAGTTATAGA GTTCAGG
Rbbp7	ACGCAAGATGGCGAGTAAA G	CAGATTTTATACTCTTCGTT GATGACA
Gapdh	TCAAGAAGGTGGTGAAGCA GG	ACCAGGAAATGAGCTTGAC AAA
β-actin	GCTGTATTCCCCTCCATCGT G	CACGGTTGGCCTTAGGGTT CAG

DNA-mediated chromatin pulldown (Dm-ChP)

ESCs were plated at a concentration of 34.000 cells/cm2 in gelatin-coated 150-mm dishes. ESCs were pulsed for 20 hours with 0.1 µM of the thymidine analog, 5-ethynyl-2'-deoxyuridine (EdU; T511285, Sigma), in parallel to doxycycline treatment.

Sorted luciferase (± EdU), $2C^-$ and $2C^+$ cells were fixed with 1 % PFA for 10 minutes on a rotating wheel at room temperature and quenched with 0.125 mM glycine (pH 7) for 5 minutes on a rotating wheel at room temperature. Cells were harvested and pelleted by centrifugation (720 g, 10 minutes at 4 °C) immediately after cell sorting. Of note, ~10 $\times 10^6$ cells were sorted per replicate and condition.

ESCs were then lysed for 30 minutes at 4 °C in lysis buffer A (10 mM Hepes (pH 7.9), 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10 % (v/v) glycerol, 1 mM DTT, 10 mM β-glycerol phosphate, 1 mM sodium orthovanadate, Protease Inhibitor Cocktail (PIC; 5056489001, Roche), and 0.1 % (v/v) Triton X-100). Nuclei were pelleted by centrifugation (1300 g, 4 minutes at 4 °C), washed with PBS + PIC, and subjected to Click reaction for 30 minutes at room temperature with 0.2 mM biotin-azide (B10184, Thermo Fisher).

In Click, an organic azide reacts with a terminal acetylene, and the nucleotide-exposed ethynyl residue of EdU is derivatized by a copper-catalyzed cycloaddition reaction, to form a covalent bond between EdU and biotin. Nuclei were repelleted by centrifugation (1300 g, 4 minutes at 4 °C), washed with PBS + PIC, and suspended in shearing buffer (20 mM tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 0.1 % (w/v) SDS, 0.5 % (w/v) sodium deoxycolate, 1 % (v/v) Triton X-100, 10 mM β-glycerol phosphate, 1 mM sodium orthovanadate and PIC). Nuclei suspension was extensively sonicated (Bioruptor, Diagenode)

for four to six cycles of 10 minutes at high intensity (30 s ON, 30 s OFF pulses). Lysates were centrifuged (20,800 g, 20 minutes at 4 °C), and supernatant was collected as input for further analysis.

To analyze shearing efficiency, 5 % of input material was reverse cross-linked by incubation overnight at 65 °C with 250 mM NaCl and then digested with proteinase K (0.1 mg/ml) for 1 h at 55 °C. DNA was purified using the PCR purification kit from Qiagen, following the manufacturer's instruction (QIAquick PCR Purification Kit; 28104, Qiagen).

For chromatin capture, input material was diluted 1:4 with blocking buffer (1 % Triton X-100, 2 mM EDTA (pH 8), 150 mM NaCl, 20 mM tris-HCl (pH 8), 20 mM β-glycerol phosphate, 2 mM sodium orthovanadate, PIC, and salmon sperm DNA (10 mg/ml; 31149, Thermo Fisher)) and then incubated with pre-blocked Dynabead M-280 streptavidin (11205D, Invitrogen) for 30 minutes at 4 °C. Beads were washed twice with blocking buffer (without salmon sperm DNA), twice with high-salt blocking buffer (containing 500 mM NaCl) and once with tris-EDTA buffer (pH 8). Beads were suspended in modified Laemmli buffer (2 % (v/v) SDS, 0.06 M tris-HCl (pH 6.5) and 0.1 M DTT) for 20 minutes at 95 °C, and the supernatant was collected for analysis by dot blot or MS.

For MS analysis of Dm-ChP samples, 9×10^6 to 17×10^6 ESCs were lysed in 5 ml of lysis buffer A, incubated in 1 ml of Click reaction buffer, sheared in 4 ml of shearing buffer, incubated

with 0.5 ml of Dynabeads M-280 streptavidin and resuspended in 150 ml of modified Laemmli buffer.

Mass spectrometry analysis

Sample preparation

Eluted proteins were reduced with 100 mM of dithiothreitol (60 minutes at 37 ºC) and alkylated in the dark with 5 µmol of iodoacetamide (20 minutes at 25 ºC). The resulting protein extract was digested in 2 M urea, 50 mM ammonium bicarbonate with endoproteinase LysC (1:10 w:w, overnight at 37 ºC; 12902541, Wako) and then trypsin (1:10 w:w, 8 hours at 37 ºC; V5113, Promega) according to filter-aided sample preparation procedure (Wisniewski et al. 2009). After digestion, the peptide mixtures were acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc) prior to LC-MS/MS analysis.

Chromatographic and mass spectrometric analysis

Samples were analyzed using a LTQ-Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EASY-nLC 1000 (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Peptides were loaded directly onto the analytical column and were separated by reversed-phase chromatography using a 50 cm column with an inner

diameter of 75 µm, packed with 2 µm C18 particles spectrometer (Thermo Scientific, San Jose, CA, USA).

Chromatographic gradients started at 95 % buffer A and 5 % buffer B with a flow rate of 300 nl/min for 5 minutes and gradually increased to 22 % buffer B and 78 % A in 79 minutes and then to 35 % buffer B and 65 % A in 11 minutes. After each analysis, the column was washed for 10 minutes with 10 % buffer A and 90 % buffer B. Buffer A was 0.1 % formic acid in water and buffer B was 0.1 % formic acid in acetonitrile.

The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 1.9 kV and source temperature at 275 °C. Ultramark 1621 was used for external calibration of the FT mass analyzer prior the analyses, and an internal calibration was performed using the background polysiloxane ion signal at m/z 445.1200. The acquisition was performed in data-dependent acquisition (DDA) mode and full MS scans with 1 micro scans at resolution of 120,000 were used over a mass range of m/z 350-1500 with detection in the Orbitrap mass analyzer. Auto gain control (AGC) was set to 1E5 and charge state filtering disqualifying singly charged peptides was activated. In each cycle of data-dependent acquisition analysis, following each survey scan, the most intense ions above a threshold ion count of 10000 were selected for fragmentation. The number of selected precursor ions for fragmentation was determined by the *Top Speed* acquisition algorithm and a dynamic exclusion of 60 seconds.

Fragment ion spectra were produced via high-energy collision dissociation (HCD) at normalized collision energy of 28 % and they were acquired in the ion trap mass analyzer. AGC was set to 1E4, and an isolation window of 1.6 m/z and maximum injection time of 200 ms were used. All data were acquired with Xcalibur software.

Digested bovine serum albumin (P8108S, NEB) was analyzed between each sample to avoid sample carryover and to assure stability of the instrument and QCloud has been used to control instrument longitudinal performance during the project (Chiva et al. 2018).

Data analysis

Acquired spectra were analyzed using the Proteome Discoverer software suite (v2.3, Thermo Fisher Scientific) and the Mascot search engine v2.6 (Matrix Science) (Perkins et al. 1999). The data were searched against a Swiss-Prot mouse database (as in October 2019) plus a list of common contaminants and all the corresponding decoy entries (Choi et al. 2011). For peptide identification a precursor ion mass tolerance of 7 ppm was used for MS1 level, trypsin was chosen as enzyme, and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications whereas carbamidomethylation on cysteines was set as a fixed

modification. False discovery rate (FDR) in peptide identification was set to a maximum of 5 %.

The analysis of specific chromatin interactors was carried out with SAINT (Analysis of INTeractome) as previously described (Perkins et al. 1999). Replicate 2 of the $2C^+$ condition was excluded for SAINT analysis. Hierarchical clustering of all the chromatome replicates was computed and visualized using Instant Clue (Nolte et al. 2018) v0.5.2 (https://www.instantclue.uni-koeln.de/). Pearson's correlation coefficients were calculated using the Prism software (v9.0, GraphPad, San Diego, CA). Gene ontology (GO) term enrichment was performed with GO Enrichment Analysis using the PANTHER tool (Ashburner et al. 2000; Mi et al. 2019) (https://geneontology.org/). Protein interaction data were retrieved from the STRING database (Szklarczyk et al. 2017) v11.0 and protein networks were visualized with Cytoscape v3.8.2.

Western blotting (WB) and protein quantitation

Cell lysis and total protein extraction

ESCs were trypsinized (Trypsin-EDTA, Sigma) at 37 °C, pelleted at 300 rcf and washed twice with PBS. During each wash cells were pelleted at 300 rcf for 5 minutes at 4 °C. ESCs were lysed instantaneously with a sodium dodecyl sulfate (SDS) lysis buffer (1% SDS, 10 mM EDTA, 50 mM tris-HCl pH

8.0) containing protease (P8340, Sigma) and phosphatase inhibitors (P5726 and P0044, Sigma). After gently mixing, lysates were sonicated (8 cycles, 30 s ON, 30 s OFF; Bioruptor, Diagenode). Protein concentrations were determined using the Pierce BCA protein assay (23225, Thermo Fisher) according to the manufacturer's protocol.

Western blotting

Total protein extract was mixed with 5x SDS-PAGE sample loading buffer (MB11701, Nzytech) and boiled for 10 minutes at 95 °C. Then, protein extract was first loaded on 4–15% precast protein gel (Mini-PROTEAN TGX; 4561084, Bio-Rad), separated by SDS-PAGE gel electrophoresis and then transferred to immuo-blot polyvinylidene difluoride (PVDF) membranes (162-0177, Bio-Rad).

The membranes were blocked with 5 % fat-free dry milk (70166, Sigma) in TBS Tween-20 (TBST, 0.1 % Tween-20 in TBS) for 1 hour and incubated with primary antibodies (anti-Cell Cycle cocktail (1:250; ab136810, Abcam), mouse anti-SMARCAD1 (1:500; ab67548, Abcam), rabbit anti-histone H3 (1:2000; ab1791, Abcam), rabbit anti-histone H3K9me3 (1:500; ab8898, Abcam), rabbit anti-histone H3 acetyl K27 (H3K27ac) (1:5000; ab177178, Abcam), rabbit anti-TOPBP1 (1:1000; ab2402, Abcam)) overnight at 4 °C. Full details of each antibody used for WB can be found in Table 1.

The membranes were then washed with TBST for 15 minutes, incubated with the peroxidase-conjugated secondary antibodies (sheep anti-mouse IgG HRP-linked (1:1000; NA931, GE Healthcare) and donkey anti-rabbit IgG HRP-linked (1:2000; NA934, GE Healthcare)) in TBST with 5 % fat-free dry milk for 1 hour, and washed again with TBST. HRP-derived signal was detected with Pierce ECL Western Blotting Substrate kit (32106, Thermo Fischer) on an Amersham Imager 600 (29083461, GE Healthcare Life Sciences).

Quantification of WB was performed by Image Studio Lite software (v5.2; LI-COR, Biosciences). For quantification, each protein was normalized to total histone H3.

Dot blot analysis

Samples were spotted in triplicates in 1 µl dots onto a nitrocellulose membrane (0.2 µM, Amersham Protan), air-dried, and detected following standard blotting procedures with the corresponding antibodies (rabbit anti-histone H3 (1:1000; ab1791, Abcam), mouse anti-vinculin (1:1000; V9131, Merck)).

Co-immunoprecipitation (co-IP)

ESCs were washed twice with PBS, trypsinized and pelleted by centrifugation (300 g, 5 minutes, 4 ºC). Freshly collected pellets were used for co-IP experiments. ESCs were lysed for 10 minutes on a rotating wheel at 4 ºC with an hypotonic buffer

(10 mM tris-HCl (pH 7.4), 10 mM KCl, 15 mM $MgCl_2$) in the presence of Protease Inhibitor Cocktail (PIC; 5056489001, Roche). Nuclei were pelleted by centrifugation (700 g, 5 minutes at 4 ºC) and lysed for 1 hour on a rotating wheel at 4 ºC with nuclear lysis buffer (300 mM NaCl, 50 mM Hepes (pH 7.5), 0.5 % NP-40, 2.5 mM $MgCl_2$) in the presence of benzonase endonuclease (100 U/ml; 71205, Novagen). Benzonase removes DNA and RNA, thereby releasing DNA-binding proteins. After nuclear lysis, samples were pelleted by centrifugation (16,000 g, 30 minutes at 4 ºC) and supernatants containing protein extracts were harvested. Protein concentrations were determined using the Pierce BCA protein assay (23225, Thermo Fisher) according to the manufacturer's protocol.

For endogenous IPs, an aliquot of protein extract (0.8-1 mg) diluted in dilution buffer (300 mM NaCl, 50 mM Hepes (pH 7.5), 0.5 % NP-40, 5 mM EDTA + PIC) was incubated with 25 µl of pre-washed Dynabeads Protein G (10004D, Thermo Fisher) and with 4 µg of the antibody of interest on a rotating wheel overnight at 4 ºC. For mock IP, ChromPure Mouse (015-000-003, Jackson Immunoresearch) or Rabbit (011-000-003, Jackson Immunoresearch) IgGs were used. Samples were then washed three times with dilution buffer. Elution was performed by incubating the dried beads with 30 µl of 5x SDS-PAGE sample loading buffer (MB11701, Nzytech) supplemented with 100 mM of 1,4-dithiothreitol (DTT) for 15

minutes at 70 °C. After eluting the samples, we directly proceeded to perform western blotting.

The following antibodies were used: rabbit anti-histone H3K9me3 (ab8898, Abcam), rabbit anti-TOPBP1 (ab2402, Abcam) and mouse anti-SMARCAD1 (ab67548, Abcam). Full details of each antibody used for co-IP can be found in Table 1.

CRISPR-Cas9 plasmid generation and delivery

CRISPR-Cas9 vector construction

Single guide RNAs (sgRNAs) targeting each of the specific target genes were retrieved from the Mouse CRISPR Knockout Pooled Library (Addgene #73632). Two sgRNA sequences were selected per gene of interest (for sgRNAs sequences, see Table 3). The sgRNAs with the highest on-target activity score (Rule Set 2) were selected for assembly into the CRISPR-Cas9 vector. A sgRNA targeting the luciferase sequence was also included as control.

Primers containing sequences for the sgRNAs were annealed in the presence of T4 ligation buffer (Thermo Fisher) and T4 PNK (NEB) in a heat block (30 ºC for 30 minutes, 95 ºC for 5 minutes and slow cool down to room temperature). Annealed primers were then cloned using the *BbsI* restriction enzyme into the pU6-(BbsI)_CBh-Cas9-T2A-mCherry plasmid

following a one-step cloning reaction. pU6-(BbsI)_CBh-Cas9-T2A-mCherry was a gift from Ralf Kuehn (Addgene plasmid #64324).

Plasmid nucleofection

ESCs were nucleofected with 4 µg of the sgRNA-containing plasmid individually following the Amaxa Mouse ES cell Nucleofector kit recommendations (VPH-1001, Lonza).

To generate DsRed-tagged ESCs, cells were co-nucleofected with a piggyBac (pB) transposase-encoding plasmid and a pCAG-DsRed-hygroR plasmid following manufacturer's protocol as described before. In both strategies, ESCs were FACS-sorted 48 hours after nucleofection to enrich for the modified cells.

Table 3 – List of top oligonucleotides used for cloning sgRNAs

sgRNA	Genomic Seq.	Target Seq.	PAM	Exon
sgCbx5 Batch #1	NC_000081.6	TGGACAGGCGCATGGTTAAG	GGG	2
sgCbx5 Batch #2	NC_000081.6	CAAAGCAATGATATCGCTCG	GGG	4
sgSmarcad1 Batch #1	NC_000072.6	AACAGAGCACATTTAAACTG	GGG	11
sgSmarcad1 Batch #2	NC_000072.6	AGTCTGTAAAACAGCCGCGA	GGG	4
sgIncenp Batch #1	NC_000085.6	TTGGGAATATTCGGTCAGTG	CGG	4
sgIncenp Batch #2	NC_000085.6	ATGAAAACAGAGATCCCGTG	AGG	3
sgOrc2 Batch #1	NC_000067.6	GGAAAGTACCCACATGACTG	AGG	13
sgOrc2 Batch #2	NC_000067.6	GGAAACGGCTCTGCTAAAGA	TGG	5
sgRbbp4 Batch #1	NC_000070.6	CATTATGACAGTGAAAAAGG	AGG	3
sgRbbp4 Batch #2	NC_000070.6	TCTTGGAATCCAAATCTCAG	TGG	5
sgCct2 Batch #1	NC_000076.6	GAGAATGGTAGCACCGTCGT	TGG	4
sgCct2 Batch #2	NC_000076.6	TGCGGCTAAGACAGTAACAG	AGG	5
sgCct5 Batch #1	NC_000081.6	GATGTGACTATAACAAACGA	TGG	3
sgCct5 Batch #2	NC_000081.6	CGTGAATGCCGTCCTCACGG	TGG	5
sgHells Batch #1	NC_000085.6	ACGGTCATTAAAACTTACAG	AGG	4
sgHells Batch #2	NC_000085.6	GCTGTATCATGGAACCCGGG	AGG	10
sgHspd1 Batch #1	NC_000067.6	AGGGACAATGGACTGAACAC	TGG	7
sgHspd1 Batch #2	NC_000067.6	TGACTTAGGAAAAGTTGGGG	AGG	9
sgRbbp7 Batch #1	NC_000086.7	TACACCGTTTCTGTATGACC	TGG	2
sgRbbp7 Batch #2	NC_000086.7	GCCCAGCACTAGCCAATGAA	GGG	3
sgGatad2b Batch #1	ENSMUSG00000042390:: Gatad2b	GCAAAACGACTGAAAATGGA	NGG	
sgGatad2b Batch #2	ENSMUSG00000042390:: Gatad2b	GGCAAAACGACTGAAAATGG	NGG	
sgLuciferase		ACAACTTTACCGACCGCGCC		

Mouse embryo injections and chimera analysis

Embryos from a cross between BDF1 mice (Charles River Laboratories) were collected at E2.5 morula stage and were cultured in EmbryoMax Advanced potassium simplex optimization medium (KSOM) medium (MR-101-D, Millipore) at 37 ºC until cell injection. Zona pellucida was partially cut using a microneedle for efficient insertion of the injection pipette (outer diameter of 28 µm to 30 µm). Five to eight luciferase or $2C^+$ DsRed-tagged cells were injected between the cytoplasm and zona pellucida. After injection, embryos were cultured until E4.5 in Advanced KSOM medium in low oxygen conditions (5 % CO_2, 5 % O_2) at 37 ºC. Live embryos were examined with an AMG EVOS microscope.

For chimeric embryo generation, E2.5 BDF1 morula embryos were collected and cell injection was performed as described above. The injected embryos were cultured with Advanced KSOM medium until E3.5. Then, 10 to 15 embryos were transferred into CD1 (Charles River Laboratories) pseudo pregnant females (2.5 dpc with vasectomized males; dpc, days post coitum). To determine chimerism at E12.5, foster female mice were sacrificed and whole embryos were collected from their uterus. Live embryos were examined with a Leica MZ16F Stereomicroscope.

Zygote collection

Embryos were collected at E0.5 from 6 to 10 weeks BDF1 females following 5 IU pregnant mare's serum gonadotrophin (PMSG) and 5 IU human chorionic gonadotropin (hCG) injections at 48 hours intervals. Female mice were mated with BDF1 male mice immediately after hCG injection. Embryos were collected from the oviducts 24 hours post-hCG and were briefly cultured in M2 medium supplemented with 0.2 mg/ml hyaluronidase (H3506, Sigma) to remove cumulus cells. Cumulus-free embryos were washed and cultured with Advanced KSOM medium at 37 ºC until microinjection.

Microinjection of morpholino antisence oligos

Morpholino antisense oligos (MO) for Rbbp7, Smarcad1 and non-targeting control were designed and produced by Gene Tools (Gene Tools, LLC). MOs were microinjected into the cytoplasm of E0.5 embryos using a Narishige micromanipulator system mounted on an Olympus IX71 inverted microscope. Embryos were immobilized using a holding pipette and MOs were then microinjected using a Narishige pneumatic microinjector (IM-300, Narishige).

After microinjection, embryos were cultured in Advanced KSOM medium in low oxygen conditions (5 % CO_2, 5 % O_2) at

37 ºC for 5 days (until E5.5). Preimplantation development was examined every 24 hours using an AMG EVOS microscope.

The following MO sequences were used:

Control MO: TCCAGGTCCCCCGCATCCCGGATCC;

Rbbp7 MO: AACATCTCTTTACTCGCCATCTTGC;

Smarcad1 MO: ATATTGGGAGGAACCACCACCCTGA.

All morpholino sequences are written from 5' to 3' and they are complementary to the translation-blocking target.

All animal experiments were approved and performed in accordance with institutional guidelines [Parc de Recerca Biomèdica de Barcelona (PRBB), Barcelona, Spain] and in accordance with the Ethical Committee for Animal Experimentation (CEEA) number PC-17-0019-PI, approved by La Comissió d'Experimentació Animal, Departament de Territori i Sostenibilitat, Direcció General de Polítiques Ambientals i Medi Natural, Generalitat de Catalunya.

Statistical analysis

As specified in the figure legends, data are presented either as scatter dot plots with line at mean ± SD, bar graphs showing mean ± SD or as min to max boxplots with line at median. All statistical tests and graphs were generated using the Prism 9.0

software (GraphPad, San Diego, CA), unless otherwise indicated.

Depending on the experimental setup, we used unpaired two-tailed Student's *t*-test, Fisher's exact test, one-way ANOVA or two-way ANOVA with the indicated post-comparison test. In all cases, a *P* value $P \leq 0.05$ was considered significant ($P \leq 0.05$*; $P \leq 0.01$**; $P \leq 0.001$***; $P \leq 0.0001$****; $P > 0.05^{ns}$, not significant).

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al. 2019) partner repository with the dataset identifier PXD019703.

(http://proteomecentral.proteomexchange.org)

PART V - REFERENCES

References

Aagaard L, Laible G, Selenko P, Schmid M, Dorn R, Schotta G, Kuhfittig S, Wolf A, Lebersorger A, Singh PB et al. 1999. Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J* **18**: 1923-1938.

Abad M, Mosteiro L, Pantoja C, Canamero M, Rayon T, Ors I, Grana O, Megias D, Dominguez O, Martinez D et al. 2013. Reprogramming in vivo produces teratomas and iPS cells with totipotency features. *Nature* **502**: 340-345.

Adenot PG, Mercier Y, Renard JP, Thompson EM. 1997. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development* **124**: 4615-4625.

Aguirre-Lavin T, Adenot P, Bonnet-Garnier A, Lehmann G, Fleurot R, Boulesteix C, Debey P, Beaujean N. 2012. 3D-FISH analysis of embryonic nuclei in mouse highlights several abrupt changes of nuclear organization during preimplantation development. *BMC Dev Biol* **12**: 30.

Akiyama T, Xin L, Oda M, Sharov AA, Amano M, Piao Y, Cadet JS, Dudekula DB, 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.

Akkouche A, Mugat B, Barckmann B, Varela-Chavez C, Li B, Raffel R, Pelisson A, Chambeyron S. 2017. Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. *Mol Cell* **66**: 411-419 e414.

Allshire RC, Madhani HD. 2018. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**: 229-244.

Aloia L, Di Stefano B, Di Croce L. 2013. Polycomb complexes in stem cells and embryonic development. *Development* **140**: 2525-2534.

Ambrosetti DC, Scholer HR, Dailey L, Basilico C. 2000. Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* **275**: 23387-23397.

Andronov L, Orlov I, Lutz Y, Vonesch JL, Klaholz BP. 2016. ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. *Sci Rep* **6**: 24084.

Aoki F, Worrad 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.

Aranda S, Alcaine-Colet A, Blanco E, Borras E, Caillot C, Sabido E, Di Croce L. 2019. Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci Adv* **5**: eaav2448.

Aranda S, Borras E, Sabido E, Di Croce L. 2020. Chromatin-Bound Proteome Profiling by Genome Capture. *STAR Protoc* **1**: 100014.

Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani CA, Imaz-Rosshandler I, Lohoff T, Xiang Y, Hanna CW et al. 2019. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**: 487-491.

Artus J, Piliszek A, Hadjantonakis AK. 2011. The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. *Dev Biol* **350**: 393-404.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25-29.

Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J et al. 2012. piRNAs can trigger a

multigenerational epigenetic memory in the germline of C. elegans. *Cell* **150**: 88-99.

Atashpaz S, Samadi Shams S, Gonzalez JM, Sebestyen E, Arghavanifard N, Gnocchi A, Albers E, Minardi S, Faga G, Soffientini P et al. 2020. ATR expands embryonic stem cell fate potential in response to replication stress. *Elife* **9**.

Audergon PN, Catania S, Kagansky A, Tong P, Shukla M, Pidoux AL, Allshire RC. 2015. Epigenetics. Restricted epigenetic inheritance of H3K9 methylation. *Science* **348**: 132-135.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**: 126-140.

Aze A, Sannino V, Soffientini P, Bachi A, Costanzo V. 2016. Centromeric DNA replication reconstitution reveals DNA loops and ATR checkpoint suppression. *Nat Cell Biol* **18**: 684-691.

Bain J, Plater L, Elliott M, Shpiro N, Hastie CJ, McLauchlan H, Klevernic I, Arthur JS, Alessi DR, Cohen P. 2007. The selectivity of protein kinase inhibitors: a further update. *Biochem J* **408**: 297-315.

Baker CL, Pera MF. 2018. Capturing Totipotent Stem Cells. *Cell Stem Cell* **22**: 25-34.

Balakier H, Pedersen RA. 1982. Allocation of cells to inner cell mass and trophectoderm lineages in preimplantation mouse embryos. *Dev Biol* **90**: 352-362.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* **21**: 381-395.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**: 120-124.

Bantele SC, Ferreira P, Gritenaite D, Boos D, Pfander B. 2017. Targeting of the Fun30 nucleosome remodeller by the Dpb11 scaffold facilitates cell cycle-regulated DNA end resection. *Elife* **6**.

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.

Bates M, Huang B, Dempsey GT, Zhuang X. 2007. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* **317**: 1749-1753.

Becker JS, Nicetto D, Zaret KS. 2016. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet* **32**: 29-41.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.

Bessonnard S, Coqueran S, Vandormael-Pournin S, Dufour A, Artus J, Cohen-Tannoudji M. 2017. ICM conversion to epiblast by FGF/ERK inhibition is limited in time and requires transcription and protein degradation. *Sci Rep* **7**: 12285.

Bessonnard S, De Mot L, Gonze D, Barriol M, Dennis C, Goldbeter A, Dupont G, Chazaud C. 2014. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**: 3637-3648.

Beuchle D, Struhl G, Muller J. 2001. Polycomb group proteins and heritable silencing of Drosophila Hox genes. *Development* **128**: 993-1004.

Bickmore WA. 2013. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet* **14**: 67-84.

Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**.

Blau HM, Chiu CP, Webster C. 1983. Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell* **32**: 1171-1180.

Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, Nichols J, Smith A, Bertone P. 2015. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev Cell* **35**: 366-382.

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.

Bouniol C, Nguyen E, Debey P. 1995. Endogenous transcription occurs at the 1-cell stage in the mouse embryo. *Exp Cell Res* **218**: 57-62.

Braude P, Bolton V, Moore S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**: 459-461.

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA et al. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**: 191-195.

Brook FA, Gardner RL. 1997. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A* **94**: 5709-5712.

Bultman SJ, Gebuhr TC, Pan H, Svoboda P, Schultz RM, Magnuson T. 2006. Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes Dev* **20**: 1744-1754.

Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de Nobelen S, Shukeir N, Popow J, Gerle B, Opravil S, Pagani M et al. 2012. A transcription factor-based mechanism for mouse heterochromatin formation. *Nat Struct Mol Biol* **19**: 1023-1030.

Burton A, Brochard V, Galan C, Ruiz-Morales ER, Rovira Q, Rodriguez-Terrones D, Kruse K, Le Gras S, Udayakumar VS, Chin HG et al. 2020. Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat Cell Biol* **22**: 767-778.

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.

Campbell KH, McWhir J, Ritchie WA, Wilmut I. 1996. Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**: 64-66.

Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ, Al-Sady B. 2011. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol Cell* **41**: 67-81.

Casanova M, Pasternak M, El Marjou F, Le Baccon P, Probst AV, Almouzni G. 2013. Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. *Cell Rep* **4**: 1156-1167.

Chalker DL, Meyer E, Mochizuki K. 2013. Epigenetics of ciliates. *Cold Spring Harb Perspect Biol* **5**: a017764.

Chappell L, Russell AJC, Voet T. 2018. Single-Cell (Multi)omics Technologies. *Annu Rev Genomics Hum Genet* **19**: 15-41.

Chazaud C, Yamanaka Y. 2016. Lineage specification in the mouse preimplantation embryo. *Development* **143**: 1063-1074.

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

Chelmicki T, Roger E, Teissandier A, Dura M, Bonneville L, Rucli S, Dossin F, Fouassier C, Lameiras S, Bourc'his D. 2021. m(6)A RNA methylation regulates the fate of endogenous retroviruses. *Nature* **591**: 312-316.

Chen X, Cui D, Papusha A, Zhang X, Chu CD, Tang J, Chen K, Pan X, Ira G. 2012. The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. *Nature* **489**: 576-580.

Chen Z, Zhang Y. 2019. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat Genet* **51**: 947-951.

Chiva C, Olivella R, Borras E, Espadas G, Pastor O, Sole A, Sabido E. 2018. QCloud: A cloud-based quality control system for mass spectrometry-based proteomics laboratories. *PLoS One* **13**: e0189209.

Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras AC, Nesvizhskii AI. 2011. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* **8**: 70-73.

Choi YJ, Lin CP, Risso D, Chen S, Kim TA, Tan MH, Li JB, Wu Y, Chen C, Xuan Z et al. 2017. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**.

Ciemerych MA, Sicinski P. 2005. Cell cycle in mouse development. *Oncogene* **24**: 2877-2898.

Clapier CR, Iwasa J, Cairns BR, Peterson CL. 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**: 407-422.

Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, Armour JA, Hewitt JE. 2007. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* **81**: 264-279.

Clark SJ, Lee HJ, Smallwood SA, Kelsey G, Reik W. 2016. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* **17**: 72.

Cloos PA, Christensen J, Agger K, Helin K. 2008. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev* **22**: 1115-1140.

Cohen DE, Melton D. 2011. Turning straw into gold: directing cell fate for regenerative medicine. *Nat Rev Genet* **12**: 243-252.

Collins PL, Kyle KE, Egawa T, Shinkai Y, Oltz EM. 2015. The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc Natl Acad Sci U S A* **112**: 8367-8372.

Condic ML. 2014. Totipotency: what it is and what it is not. *Stem Cells Dev* **23**: 796-812.

Cossec JC, Theurillat I, Chica C, Bua Aguin 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 e748.

Costelloe T, Louge R, Tomimatsu N, Mukherjee B, Martini E, Khadaroo B, Dubois K, Wiegant WW, Thierry A, Burma S et al. 2012. The yeast Fun30 and human SMARCAD1

chromatin remodellers promote DNA end resection. *Nature* **489**: 581-584.

Cremer T, Cremer M. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**: a003889.

Cuenda A, Rousseau S. 2007. p38 MAP-kinases pathway regulation, function and role in human diseases. *Biochim Biophys Acta* **1773**: 1358-1375.

Czermin B, Schotta G, Hulsmann BB, Brehm A, Becker PB, Reuter G, Imhof A. 2001. Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. *EMBO Rep* **2**: 915-919.

Davis RL, Weintraub H, Lassar AB. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987-1000.

De Iaco A, Coudray A, Duc J, Trono D. 2019. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep* **20**.

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

Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* **295**: 1306-1311.

Deng Q, Ramskold D, Reinius B, Sandberg R. 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**: 193-196.

Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**: 1233-1244.

Denker A, de Laat W. 2016. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev* **30**: 1357-1382.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380.

Djupedal I, Portoso M, Spahr H, Bonilla C, Gustafsson CM, Allshire RC, Ekwall K. 2005. RNA Pol II subunit Rpb7

promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev* **19**: 2301-2306.

Do DV, Ueda J, Messerschmidt DM, Lorthongpanich C, Zhou Y, Feng B, Guo G, Lin PJ, Hossain MZ, Zhang W et al. 2013. A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. *Genes Dev* **27**: 1378-1390.

Dodge JE, Kang YK, Beppu H, Lei H, Li E. 2004. Histone H3-K9 methyltransferase ESET is essential for early development. *Mol Cell Biol* **24**: 2478-2486.

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C et al. 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* **16**: 1299-1309.

Downes CS, Clarke DJ, Mullinger AM, Gimenez-Abian JF, Creighton AM, Johnson RT. 1994. A topoisomerase II-dependent G2 cycle checkpoint in mammalian cells. *Nature* **372**: 467-470.

Dykhuizen EC, Hargreaves DC, Miller EL, Cui K, Korshunov A, Kool M, Pfister S, Cho YJ, Zhao K, Crabtree GR. 2013. BAF complexes facilitate decatenation of DNA by topoisomerase IIalpha. *Nature* **497**: 624-627.

Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D. 2016. Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev Cell* **36**: 611-623.

Eckersley-Maslin M, Alda-Catalinas C, Blotenburg M, Kreibich E, Krueger C, Reik W. 2019. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev* **33**: 194-208.

Eckersley-Maslin MA, Alda-Catalinas C, Reik W. 2018. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol* **19**: 436-450.

Eckersley-Maslin MA, Svensson V, Krueger C, Stubbs TM, Giehr P, Krueger F, Miragaia RJ, Kyriakopoulos C, Berrens RV, Milagre I et al. 2016. MERVL/Zscan4

Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep* **17**: 179-192.

Edwards RG, Beard HK. 1997. Oocyte polarity and cell determination in early mammalian embryos. *Mol Hum Reprod* **3**: 863-905.

Elgin SC, Reuter G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* **5**: a017780.

Evans MJ, Kaufman MH. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**: 154-156.

Falco G, Lee SL, Stanghellini I, Bassey UC, Hamatani T, Ko MS. 2007. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev Biol* **307**: 539-550.

Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jonsson ME, Manne C, Turelli P, Jern P, Cammas F, Trono D et al. 2015. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep* **10**: 20-28.

Filipescu D, Muller S, Almouzni G. 2014. Histone H3 variants and their chaperones during development and disease: contributing to epigenetic control. *Annu Rev Cell Dev Biol* **30**: 615-646.

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.

Frankenberg S, Gerbe F, Bessonnard S, Belville C, Pouchin P, Bardot O, Chazaud C. 2011. Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev Cell* **21**: 1005-1013.

Friedli M, Trono D. 2015. The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol* **31**: 429-451.

Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ, 3rd. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev* **10**: 2067-2078.

Frum T, Halbisen MA, Wang C, Amiri H, Robson P, Ralston A. 2013. Oct4 cell-autonomously promotes primitive

endoderm development in the mouse blastocyst. *Dev Cell* **25**: 610-622.

Fu X, Djekidel MN, Zhang Y. 2020. A transcriptional roadmap for 2C-like-to-pluripotent state transition. *Sci Adv* **6**: eaay5181.

Fu X, Wu X, Djekidel MN, Zhang Y. 2019. Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat Cell Biol* **21**: 835-844.

Fujimori T, Kurotaki Y, Miyazaki J, Nabeshima Y. 2003. Analysis of cell lineage in two- and four-cell mouse embryos. *Development* **130**: 5113-5122.

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH et al. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58-64.

Furey TS. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* **13**: 840-852.

Furlong EEM, Levine M. 2018. Developmental enhancers and chromosome topology. *Science* **361**: 1341-1345.

Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, Ellis J, Bazett-Jones DP. 2012. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep* **13**: 992-996.

Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M. 2011. Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* **12**: 36-47.

Genet M, Torres-Padilla ME. 2020. The molecular and cellular features of 2-cell-like cells: a reference guide. *Development* **147**.

Gifford WD, Pfaff SL, Macfarlan TS. 2013. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol* **23**: 218-226.

Graf T, Enver T. 2009. Forcing cells to change lineages. *Nature* **462**: 587-594.

Grob S, Cavalli G. 2018. Technical Review: A Hitchhiker's Guide to Chromosome Conformation Capture. *Methods Mol Biol* **1675**: 233-246.

Groh S, Schotta G. 2017. Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci* **74**: 2055-2065.

Groth A, Rocha W, Verreault A, Almouzni G. 2007. Chromatin challenges during DNA replication and repair. *Cell* **128**: 721-733.

Gu T, Elgin SC. 2013. Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. *PLoS Genet* **9**: e1003780.

Guo M, Zhang Y, Zhou J, Bi Y, Xu J, Xu C, Kou X, Zhao Y, Li Y, Tu Z et al. 2019. Precise temporal regulation of Dux is important for embryo development. *Cell Res* **29**: 956-959.

Gurdon JB. 1962. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol* **10**: 622-640.

Gurdon JB, Elsdale TR, Fischberg M. 1958. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature* **182**: 64-65.

Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297**: 2232-2237.

Hamatani T, Carter MG, Sharov AA, Ko MS. 2004. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6**: 117-131.

Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. 2011. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet* **7**: e1002266.

Harrison SE, Sozen B, Christodoulou N, Kyprianou C, Zernicka-Goetz M. 2017. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science* **356**.

Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**: 519-532.

Hendrickson PG, Dorais JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL 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.

Henikoff S, Greally JM. 2016. Epigenetics, cellular memory and gene regulation. *Curr Biol* **26**: R644-648.

Henikoff S, Smith MM. 2015. Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* **7**: a019364.

Hiragami-Hamada K, Soeroes S, Nikolov M, Wilkins B, Kreuz S, Chen C, De La Rosa-Velazquez IA, Zenn HM, Kost N, Pohl W et al. 2016. Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin. *Nat Commun* **7**: 11310.

Hoijman E, Hakkinen HM, Tolosa-Ramon Q, Jimenez-Delgado S, Wyatt C, Miret-Cuesta M, Irimia M, Callan-Jones A, Wieser S, Ruprecht V. 2021. Cooperative epithelial phagocytosis enables error correction in the early embryo. *Nature* **590**: 618-623.

Home P, Ray S, Dutta D, Bronshteyn I, Larson M, Paul S. 2009. GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. *J Biol Chem* **284**: 28729-28737.

Hota SK, Bruneau BG. 2016. ATP-dependent chromatin remodeling during mammalian development. *Development* **143**: 2882-2897.

Hu Z, Tan DEK, Chia G, Tan H, Leong HF, Chen BJ, Lau MS, Tan KYS, Bi X, Yang D et al. 2020. Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat Cell Biol* **22**: 175-186.

Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, Higgs DR. 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**: 205-212.

Ishiuchi T, Enriquez-Gasca R, Mizutani E, Boskovic A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM, Torres-Padilla ME. 2015. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol* **22**: 662-671.

Ishiuchi T, Torres-Padilla ME. 2013. Towards an understanding of the regulatory mechanisms of totipotency. *Curr Opin Genet Dev* **23**: 512-518.

Iturbide A, Torres-Padilla ME. 2020. A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming. *Curr Opin Genet Dev* **64**: 26-30.

Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, Schultz DC, Psulkowski E, Fredericks WJ, White DE, Maul GG et al. 2007. PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell* **28**: 823-837.

Iyengar S, Farnham PJ. 2011. KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem* **286**: 26267-26276.

Jachowicz JW, Bing X, Pontabry J, Boskovic A, Rando OJ, Torres-Padilla ME. 2017. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet* **49**: 1502-1510.

Jacobs SA, Taverna SD, Zhang Y, Briggs SD, Li J, Eissenberg JC, Allis CD, Khorasanizadeh S. 2001. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J* **20**: 5232-5241.

Janssen A, Colmenares SU, Karpen GH. 2018. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol* **34**: 265-288.

Jedrusik A, Parfitt DE, Guo G, Skamagki M, Grabarek JB, Johnson MH, Robson P, Zernicka-Goetz M. 2008. Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes Dev* **22**: 2692-2706.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**: 1497-1502.

Jones KW. 1970. Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature* **225**: 912-915.

Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M et al. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**: 676-687.

Kato H, Goto DB, Martienssen RA, Urano T, Furukawa K, Murakami Y. 2005. RNA polymerase II is required for

RNAi-dependent heterochromatin assembly. *Science* **309**: 467-469.

Kato M, Takemoto K, Shinkai Y. 2018. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. *Nat Commun* **9**: 1683.

Kelly SJ. 1977. Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J Exp Zool* **200**: 365-376.

Kelsey G, Stegle O, Reik W. 2017. Single-cell epigenomics: Recording the past and predicting the future. *Science* **358**: 69-75.

Klenov MS, Sokolova OA, Yakushev EY, Stolyarenko AD, Mikhaleva EA, Lavrov SA, Gvozdev VA. 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci U S A* **108**: 18760-18765.

Komatsu K, Fujimori T. 2015. Multiple phases in regulation of Nanog expression during pre-implantation development. *Dev Growth Differ* **57**: 648-656.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.

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

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.

Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, Graf T. 2006. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity* **25**: 731-744.

Lakadamyali M, Cosma MP. 2020. Visualizing the genome in high resolution challenges our textbook understanding. *Nat Methods*.

Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. 2017. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**: 236-240.

Larson AG, Narlikar GJ. 2018. The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. *Biochemistry* **57**: 2540-2548.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Toth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* **27**: 390-399.

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13**: 1192-1200.

Leidenroth A, Clapp J, Mitchell LM, Coneyworth D, Dearden FL, Iannuzzi L, Hewitt JE. 2012. Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma* **121**: 489-497.

Leung CY, Zernicka-Goetz M. 2015. Mapping the journey from totipotency to lineage specification in the mouse embryo. *Curr Opin Genet Dev* **34**: 71-76.

Levet F, Hosy E, Kechkar A, Butler C, Beghin A, Choquet D, Sibarita JB. 2015. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nat Methods* **12**: 1065-1071.

Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**: 400-403.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289-293.

Liu K, Luo Y, Lin FT, Lin WC. 2004. TopBP1 recruits Brg1/Brm to repress E2F1-induced apoptosis, a novel pRb-independent and E2F1-specific control for cell survival. *Genes Dev* **18**: 673-686.

Liu L, Bailey SM, Okuka M, Munoz P, Li C, Zhou L, Wu C, Czerwiec E, Sandler L, Seyfang A et al. 2007. Telomere

lengthening early in development. *Nat Cell Biol* **9**: 1436-1441.

Liu L, Michowski W, Kolodziejczyk A, Sicinski P. 2019. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nat Cell Biol* **21**: 1060-1067.

Lomberk G, Bensi D, Fernandez-Zapico ME, Urrutia R. 2006a. Evidence for the existence of an HP1-mediated subcode within the histone code. *Nat Cell Biol* **8**: 407-415.

Lomberk G, Wallrath L, Urrutia R. 2006b. The Heterochromatin Protein 1 family. *Genome Biol* **7**: 228.

Lu F, Liu Y, Jiang L, Yamaguchi S, Zhang Y. 2014. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* **28**: 2103-2119.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**: 251-260.

Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R et al. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**: 1012-1025.

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**: 57-63.

Machida S, Takizawa Y, Ishimaru M, Sugita Y, Sekine S, Nakayama JI, Wolf M, Kurumizaka H. 2018. Structural Basis of Heterochromatin Formation by Human HP1. *Mol Cell* **69**: 385-397 e388.

Maeda I, Okamura D, Tokitake Y, Ikeda M, Kawaguchi H, Mise N, Abe K, Noce T, Okuda A, Matsui Y. 2013. Max is a repressor of germ cell-related gene expression in mouse embryonic stem cells. *Nat Commun* **4**: 1754.

Mandai M, Watanabe A, Kurimoto Y, Hirami Y, Morinaga C, Daimon T, Fujihara M, Akimaru H, Sakai N, Shibata Y et al. 2017. Autologous Induced Stem-Cell-Derived Retinal Cells for Macular Degeneration. *N Engl J Med* **376**: 1038-1046.

Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A

et al. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**: 590-604.

Martienssen RA, Kloc A, Slotkin RK, Tanurdzic M. 2008. Epigenetic inheritance and reprogramming in plants and fission yeast. *Cold Spring Harb Symp Quant Biol* **73**: 265-271.

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

Martire S, Banaszynski LA. 2020. The roles of histone variants in fine-tuning chromatin organization and function. *Nat Rev Mol Cell Biol* **21**: 522-541.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**: 927-931.

McGrath J, Solter D. 1984. Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. *Science* **226**: 1317-1319.

Meilhac SM, Adams RJ, Morris SA, Danckaert A, Le Garrec JF, Zernicka-Goetz M. 2009. Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. *Dev Biol* **331**: 210-221.

Meng J, Zhang HH, Zhou CX, Li C, Zhang F, Mei QB. 2012. The histone deacetylase inhibitor trichostatin A induces cell cycle arrest and apoptosis in colorectal cancer cells via p53-dependent and -independent pathways. *Oncol Rep* **28**: 384-388.

Messerschmidt DM, Kemler R. 2010. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev Biol* **344**: 129-137.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**: D419-D426.

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA et al. 2015. Mapping long-range

promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**: 598-606.

Mikhailov A, Shinohara M, Rieder CL. 2004. Topoisomerase II and histone deacetylase inhibitors delay the G2/M transition by triggering the p38 MAPK checkpoint pathway. *J Cell Biol* **166**: 517-526.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553-560.

Morgani SM, Brickman JM. 2014. The molecular underpinnings of totipotency. *Philos Trans R Soc Lond B Biol Sci* **369**.

Morgani SM, Canham MA, Nichols J, Sharov AA, Migueles RP, Ko MS, Brickman JM. 2013. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep* **3**: 1945-1957.

Morris SA, Daley GQ. 2013. A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell Res* **23**: 33-48.

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**: 110-113.

Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**: 490-503.

Nichols J, Ying QL. 2006. Derivation and propagation of embryonic stem cells in serum- and feeder-free culture. *Methods Mol Biol* **329**: 91-98.

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**: 379-391.

Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, O'Carroll D, Firestein R, Cleary M, Jenuwein T, Herrera RE et al. 2001. Rb targets histone H3 methylation and HP1 to promoters. *Nature* **412**: 561-565.

Niwa H, Burdon T, Chambers I, Smith A. 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**: 2048-2060.

Nolte H, MacVicar TD, Tellkamp F, Kruger M. 2018. Instant Clue: A Software Suite for Interactive Data Visualization and Analysis. *Sci Rep* **8**: 12648.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381-385.

Nottke A, Colaiacovo MP, Shi Y. 2009. Developmental roles of the histone lysine demethylases. *Development* **136**: 879-889.

Ooga M, Fulka H, Hashimoto S, Suzuki MG, Aoki F. 2016. Analysis of chromatin structure in mouse preimplantation embryos by fluorescent recovery after photobleaching. *Epigenetics* **11**: 85-94.

Otterstrom J, Castells-Garcia A, Vicario C, Gomez-Garcia PA, Cosma MP, Lakadamyali M. 2019. Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo. *Nucleic Acids Res* **47**: 8470-8484.

Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**.

Papaioannou VE, Mkandawire J, Biggers JD. 1989. Development and phenotypic variability of genetically identical half mouse embryos. *Development* **106**: 817-827.

Park SJ, Komata M, Inoue F, Yamada K, Nakai K, Ohsugi M, Shirahige K. 2013. Inferring the choreography of parental genomes during fertilization from ultralarge-scale whole-transcriptome analysis. *Genes Dev* **27**: 2736-2748.

Passier R, Orlova V, Mummery C. 2016. Complex Tissue and Disease Modeling using hiPSCs. *Cell Stem Cell* **18**: 309-321.

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004. Retrotransposons

regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7**: 597-606.

Peng H, Begg GE, Schultz DC, Friedman JR, Jensen DE, Speicher DW, Rauscher FJ, 3rd. 2000. Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J Mol Biol* **295**: 1139-1162.

Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen X, Ramalho-Santos M. 2018. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**: 391-405 e319.

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M et al. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**: D442-D450.

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551-3567.

Pesaresi M, Sebastian-Perez R, Cosma MP. 2019. Dedifferentiation, transdifferentiation and cell fusion: in vivo reprogramming strategies for regenerative medicine. *FEBS J* **286**: 1074-1093.

Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, Kohlmaier A, Opravil S, Tachibana M, Shinkai Y et al. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* **12**: 1577-1589.

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A et al. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**: 323-337.

Pezic D, Manakov SA, Sachidanandam R, Aravin AA. 2014. piRNA pathway targets active LINE1 elements to

establish the repressive H3K9me3 mark in germ cells. *Genes Dev* **28**: 1410-1428.

Piotrowska-Nitsche K, Perea-Gomez A, Haraguchi S, Zernicka-Goetz M. 2005. Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**: 479-490.

Plachta N, Bollenbach T, Pease S, Fraser SE, Pantazis P. 2011. Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat Cell Biol* **13**: 117-123.

Platero JS, Hartnett T, Eissenberg JC. 1995. Functional analysis of the chromo domain of HP1. *EMBO J* **14**: 3977-3986.

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131-135.

Plusa B, Frankenberg S, Chalmers A, Hadjantonakis AK, Moore CA, Papalopulu N, Papaioannou VE, Glover DM, Zernicka-Goetz M. 2005. Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *Journal of cell science* **118**: 505-515.

Plusa B, Piliszek A, Frankenberg S, Artus J, Hadjantonakis AK. 2008. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**: 3081-3091.

Pommier Y. 2006. Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer* **6**: 789-802.

Posfai E, Schell JP, Janiszewski A, Rovic I, Murray A, Bradshaw B, Yamakawa T, Pardon T, El Bakkali M, Talon I et al. 2021. Evaluating totipotency using criteria of increasing stringency. *Nat Cell Biol* **23**: 49-60.

Probst AV, Almouzni G. 2011. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends Genet* **27**: 177-185.

Probst AV, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G. 2010. A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev Cell* **19**: 625-638.

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.

Puschendorf M, Terranova R, Boutsma E, Mao X, Isono K, Brykczynska U, Kolb C, Otte AP, Koseki H, Orkin SH et al. 2008. PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* **40**: 411-420.

Qiu L, Burgess A, Fairlie DP, Leonard H, Parsons PG, Gabrielli BG. 2000. Histone deacetylase inhibitors trigger a G2 checkpoint in normal cells that is defective in tumor cells. *Molecular biology of the cell* **11**: 2069-2083.

Ragunathan K, Jih G, Moazed D. 2015. Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348**: 1258699.

Ralston A, Rossant J. 2008. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev Biol* **313**: 614-629.

Ram PT, Schultz RM. 1993. Reporter gene expression in G2 of the 1-cell mouse embryo. *Dev Biol* **156**: 552-556.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665-1680.

Raulf A, Spahn CK, Zessin PJ, Finan K, Bernhardt S, Heckel A, Heilemann M. 2014. Click chemistry facilitates direct labelling and super-resolution imaging of nucleic acids and proteinsdaggerElectronic supplementary information (ESI) available. See DOI: 10.1039/c4ra01027bClick here for additional data file. *RSC Adv* **4**: 30462-30466.

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.

Reinberg D, Vales LD. 2018. Chromatin domains rich in inheritance. *Science* **361**: 33-34.

Ricci MA, Manzo C, Garcia-Parajo MF, Lakadamyali M, Cosma MP. 2015. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**: 1145-1158.

Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, Shinkai Y, Allis CD. 2003. Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol Cell* **12**: 1591-1598.

Richards EJ, Elgin SC. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**: 489-500.

Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res* **21**: 147-163.

Ringrose L, Paro R. 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* **38**: 413-443.

Rivron NC, Frias-Aldeguer J, Vrij EJ, Boisset JC, Korving J, Vivie J, Truckenmuller RK, van Oudenaarden A, van Blitterswijk CA, Geijsen N. 2018. Blastocyst-like structures generated solely from stem cells. *Nature* **557**: 106-111.

Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. 2000. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* **25**: 338-342.

Robinson HM, Bratlie-Thoresen S, Brown R, Gillespie DA. 2007. Chk1 is required for G2/M checkpoint response induced by the catalytic topoisomerase II inhibitor ICRF-193. *Cell Cycle* **6**: 1265-1267.

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

Rodriguez-Terrones D, Hartleben G, Gaume X, Eid A, Guthmann M, Iturbide A, Torres-Padilla ME. 2020. A

distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep* **21**: e48354.

Rodriguez-Terrones D, Torres-Padilla ME. 2018. Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet* **34**: 806-820.

Rodriguez TA, Srinivas S, Clements MP, Smith JC, Beddington RS. 2005. Induction and migration of the anterior visceral endoderm is regulated by the extra-embryonic ectoderm. *Development* **132**: 2513-2520.

Rossant J. 1976. Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *J Embryol Exp Morphol* **36**: 283-290.

Rowbotham SP, Barki L, Neves-Costa A, Santos F, Dean W, Hawkes N, Choudhary P, Will WR, Webster J, Oxley D et al. 2011. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol Cell* **42**: 285-296.

Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D. 2013a. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**: 519-529.

Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J et al. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**: 237-240.

Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D. 2013b. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* **23**: 452-461.

Rowley MJ, Corces VG. 2018. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**: 789-800.

Rozhkov NV, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev* **27**: 400-412.

Rust MJ, Bates M, Zhuang X. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**: 793-795.

Sachs P, Ding D, Bergmaier P, Lamp B, Schlagheck C, Finkernagel F, Nist A, Stiewe T, Mermoud JE. 2019. SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. *Nat Commun* **10**: 1335.

Saintigny Y, Delacote F, Vares G, Petitot F, Lambert S, Averbeck D, Lopez BS. 2001. Characterization of homologous recombination induced by replication inhibition in mammalian cells. *EMBO J* **20**: 3861-3870.

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.

Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. 2004. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**: 55-63.

Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS et al. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**: 53-61.

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G. 2002. Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* **21**: 1121-1131.

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.

Schrode N, Saiz N, Di Talia S, Hadjantonakis AK. 2014. GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Dev Cell* **29**: 454-467.

Schrode N, Xenopoulos P, Piliszek A, Frankenberg S, Plusa B, Hadjantonakis AK. 2013. Anatomy of a blastocyst: cell behaviors driving cell fate choice and

morphogenesis in the early mouse embryo. *Genesis* **51**: 219-233.

Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G. 2017. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**: 34-57.

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell* **128**: 735-745.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ, 3rd. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**: 919-932.

Schultz DC, Friedman JR, Rauscher FJ, 3rd. 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev* **15**: 428-443.

Schultz RM. 1993. Regulation of zygotic gene activation in the mouse. *Bioessays* **15**: 531-538.

Selker EU. 1998. Trichostatin A causes selective loss of DNA methylation in Neurospora. *Proc Natl Acad Sci U S A* **95**: 9430-9435.

Seller CA, Cho CY, O'Farrell PH. 2019. Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in Drosophila. *Genes Dev* **33**: 403-417.

Sexton T, Cavalli G. 2015. The role of chromosome domains in shaping the functional genome. *Cell* **160**: 1049-1059.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**: 458-472.

Sienski G, Donertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**: 964-980.

Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith

A. 2009. Nanog is the gateway to the pluripotent ground state. *Cell* **138**: 722-737.

Silva J, Smith A. 2008. Capturing pluripotency. *Cell* **132**: 532-536.

Simmons DG, Cross JC. 2005. Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. *Dev Biol* **284**: 12-24.

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**: 1348-1354.

Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**: 688-690.

Soares ML, Torres-Padilla ME, Zernicka-Goetz M. 2008. Bone morphogenetic protein 4 signaling regulates development of the anterior visceral endoderm in the mouse embryo. *Dev Growth Differ* **50**: 615-621.

Sozen B, Amadei G, Cox A, Wang R, Na E, Czukiewska S, Chappell L, Voet T, Michel G, Jing N et al. 2018. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat Cell Biol* **20**: 979-989.

Spangler A, Su EY, Craft AM, Cahan P. 2018. A single cell transcriptional portrait of embryoid body differentiation and comparison to progenitors of the developing embryo. *Stem Cell Res* **31**: 201-215.

Stadhouders R, Filion GJ, Graf T. 2019. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**: 345-354.

Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. 2017. Phase separation drives heterochromatin domain formation. *Nature* **547**: 241-245.

Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, Rossant J. 2005. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**: 2093-2102.

Surani A, Tischler J. 2012. Stem cells: a sporadic super state. *Nature* **487**: 43-45.

Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM. 2004. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol* **269**: 276-285.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P et al. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **45**: D362-D368.

Tabansky I, Lenarcic A, Draft RW, Loulier K, Keskin DB, Rosains J, Rivera-Feliciano J, Lichtman JW, Livet J, Stern JN et al. 2013. Developmental bias in cleavage-stage mouse blastomeres. *Curr Biol* **23**: 21-31.

Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. 2001. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* **276**: 25309-25317.

Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H et al. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **16**: 1779-1791.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663-676.

Tarkowski AK. 1959. Experiments on the development of isolated blastomers of mouse eggs. *Nature* **184**: 1286-1287.

Tarkowski AK, Suwinska A, Czolowska R, Ozdzenski W. 2010. Individual blastomeres of 16- and 32-cell mouse embryos are able to develop into foetuses and mice. *Dev Biol* **348**: 190-198.

Tatsumi D, Hayashi Y, Endo M, Kobayashi H, Yoshioka T, Kiso K, Kanno S, Nakai Y, Maeda I, Mochizuki K et al. 2018. DNMTs and SETDB1 function as co-repressors in MAX-mediated repression of germ cell-related genes in mouse embryonic stem cells. *PLoS One* **13**: e0205969.

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**: 196-199.

Thompson PJ, Dulberg V, Moon KM, Foster LJ, Chen C, Karimi MM, Lorincz MC. 2015. hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. *PLoS Genet* **11**: e1004933.

Thornton TM, Rincon M. 2009. Non-classical p38 map kinase functions: cell cycle checkpoints and survival. *Int J Biol Sci* **5**: 44-51.

Tie CH, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, Rodriguez-Plata MT, Mickute G, Gifford R, Towers GJ et al. 2018. KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. *EMBO Rep* **19**.

Torres-Padilla ME. 2020. On transposons and totipotency. *Philos Trans R Soc Lond B Biol Sci* **375**: 20190339.

Torres-Padilla ME, Parfitt DE, Kouzarides T, Zernicka-Goetz M. 2007. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445**: 214-218.

Tosolini M, Brochard V, Adenot P, Chebrout M, Grillo G, Navia V, Beaujean N, Francastel C, Bonnet-Garnier A, Jouneau A. 2018. Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. *Sci Rep* **8**: 5776.

Tsunoda Y, McLaren A. 1983. Effect of various procedures on the viability of mouse embryos containing half the normal number of blastomeres. *J Reprod Fertil* **69**: 315-322.

Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D. 2014. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res* **24**: 1260-1270.

Uchimura Y, Ichimura T, Uwada J, Tachibana T, Sugahara S, Nakao M, Saitoh H. 2006. Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. *J Biol Chem* **281**: 23180-23190.

van Steensel B, Belmont AS. 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**: 780-791.

Vermaak D, Malik HS. 2009. Multiple roles for heterochromatin protein 1 genes in Drosophila. *Annu Rev Genet* **43**: 467-492.

Vierbuchen T, Wernig M. 2011. Direct lineage conversions: unnatural but useful? *Nat Biotechnol* **29**: 892-907.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833-1837.

Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, Chen C, Kou X, Zhao Y, Chen J et al. 2018. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol* **20**: 620-631.

Wang JC. 2002. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* **3**: 430-440.

Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. 2016. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**: 598-602.

Wang SH, Elgin SC. 2011. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci U S A* **108**: 21164-21169.

Wennekamp S, Mesecke S, Nedelec F, Hiiragi T. 2013. A self-organization framework for symmetry breaking in the mammalian embryo. *Nat Rev Mol Cell Biol* **14**: 452-459.

Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ. 2017. Conservation and innovation in the DUX4-family gene network. *Nat Genet* **49**: 935-940.

Wicklow E, Blij S, Frum T, Hirate Y, Lang RA, Sasaki H, Ralston A. 2014. HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS Genet* **10**: e1004618.

Wiles ET, Selker EU. 2017. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr Opin Genet Dev* **43**: 31-37.

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**: 684-687.

Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**: 810-813.

Wisniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat Methods* **6**: 359-362.

Wolf D, Goff SP. 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**: 46-57.

Wolf G, Yang P, Fuchtbauer AC, Fuchtbauer EM, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS. 2015. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev* **29**: 538-554.

Wray J, Kalkan T, Smith AG. 2010. The ground state of pluripotency. *Biochem Soc Trans* **38**: 1027-1032.

Wu G, Lei L, Scholer HR. 2017. Totipotency in the mouse. *J Mol Med (Berl)* **95**: 687-694.

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

Wu J, Yamauchi T, Izpisua Belmonte JC. 2016b. An overview of mammalian pluripotency. *Development* **143**: 1644-1648.

Xie H, Ye M, Feng R, Graf T. 2004. Stepwise reprogramming of B cells into macrophages. *Cell* **117**: 663-676.

Yadav T, Quivy JP, Almouzni G. 2018. Chromatin plasticity: A versatile landscape that underlies cell fate and identity. *Science* **361**: 1332-1336.

Yamashita Y, Shimada M, Harimoto N, Rikimaru T, Shirabe K, Tanaka S, Sugimachi K. 2003. Histone deacetylase inhibitor trichostatin A induces cell-cycle

arrest/apoptosis and hepatocyte differentiation in human hepatoma cells. *Int J Cancer* **103**: 572-576.

Yan YL, Zhang C, Hao J, Wang XL, Ming J, Mi L, Na J, Hu X, Wang Y. 2019. DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program. *PLoS Biol* **17**: e3000324.

Yang F, Huang X, Zang R, Chen J, Fidalgo M, Sanchez-Priego C, Yang J, Caichen A, Ma F, Macfarlan T et al. 2020. DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* **26**: 234-250 e237.

Yang J, Ryan DJ, Wang W, Tsang JC, Lan G, Masaki H, Gao X, Antunes L, Yu Y, Zhu Z et al. 2017a. Establishment of mouse expanded potential stem cells. *Nature* **550**: 393-397.

Yang Y, Liu B, Xu J, Wang J, Wu J, Shi C, Xu Y, Dong J, Wang C, Lai W et al. 2017b. Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell* **169**: 243-257 e225.

Yiangou L, Grandy RA, Morell CM, Tomaz RA, Osnato A, Kadiwala J, Muraro D, Garcia-Bernardo J, Nakanoh S, Bernard WG et al. 2019. Method to Synchronize Cell Cycle of Human Pluripotent Stem Cells without Affecting Their Fundamental Characteristics. *Stem Cell Reports* **12**: 165-179.

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

Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee SL, Stagg CA, Hoang HG, Yang HT, Indig FE et al. 2010. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**: 858-863.

Zeman MK, Cimprich KA. 2014. Causes and consequences of replication stress. *Nat Cell Biol* **16**: 2-9.

Zeng F, Baldwin DA, Schultz RM. 2004. Transcript profiling during preimplantation mouse development. *Dev Biol* **272**: 483-496.

Zernicka-Goetz M, Morris SA, Bruce AW. 2009. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10**: 467-477.

Zessin PJ, Finan K, Heilemann M. 2012. Super-resolution fluorescence imaging of chromosomal DNA. *J Struct Biol* **177**: 344-348.

Zhang Q, Dan J, Wang H, Guo R, Mao J, Fu H, Wei X, Liu L. 2016. Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells. *Sci Rep* **6**: 19852.

Zhang Y, Huang Y, Dong Y, Liu X, Kou X, Zhao Y, Zhao A, Sun J, Su Z, Li Z et al. 2021. Unique Patterns of H3K4me3 and H3K27me3 in 2-Cell-like Embryonic Stem Cells. *Stem Cell Reports* **16**: 458-469.

Zhou Q, Wang M, Yuan Y, Wang X, Fu R, Wan H, Xie M, Liu M, Guo X, Zheng Y et al. 2016. Complete Meiosis from Embryonic Stem Cell-Derived Germ Cells In Vitro. *Cell Stem Cell* **18**: 330-340.

Zilberman D, Cao X, Jacobsen SE. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716-719.

PART VI - ANNEX

ANNEX - PUBLICATIONS

Research article resulting from this PhD Thesis:

Sebastian-Perez R, Aranda A, Nakagawa S, Pesaresi M, Gomez-Garcia PA, Alcoverro-Bertran M, Gomez-Vazquez JL, Carnevali D, Borràs E, Sabidó E, Nissim-Rafinia M, Meshorer E, Di Croce L, Cosma MP. SMARCAD1 contributes to heterochromatin establishment at the transition from the 2C-like to the pluripotent state. *Manuscript under review*.

I collaborated in developing a single locus imaging technology for live and STORM microscopy exploiting the CRISPR-dCas9 system:

Neguembor MV, **Sebastian-Perez R**, Aulicino F, Gomez-Garcia PA, Cosma MP, Lakadamyali M. (Po)STAC (Polycistronic SunTAg modified CRISPR) enables live-cell and fixed-cell super-resolution imaging of multiple genes. *Nucleic Acids Research*, 2018 Mar 16; 46 (5):e30.

Neguembor MV, **Sebastian-Perez R**, Aulicino F, Gomez-Garcia PA, Cosma MP, Lakadamyali M. (Po)STAC (Polycistronic SunTAg modified CRISPR) enables live-cell and fixed-cell super-resolution imaging of multiple genes. Nucleic Acids Research, 46 (5):e30 (2018).

I collaborated in identifying signaling axes that enhances migration of adult stem cells toward the degenerating retina and demonstrating its potential use in stem cell-based therapies:

Pesaresi M, Bonilla-Pons SA, **Sebastian-Perez R**, Di Vicino U, Alcoverro-Bertran M, Cosma MP. Exogenous expression of Ccr5 and Cxcr6 improves mouse mesenchymal stem cell integration into the degenerating retina. *Molecular Therapy*, 2021 Feb 3; 29 (2):804-821.

Pesaresi M, Bonilla-Pons SA, **Sebastian-Perez R**, Di Vicino U, Alcoverro-Bertran M, Cosma MP. Exogenous expression of Ccr5 and Cxcr6 improves mouse mesenchymal stem cell integration into the degenerating retina. Molecular Therapy, 29 (2):804-821 (2021).

I collaborated in determining the role of transcription in 3D genome folding and loop extrusion:

Neguembor MV, Martin L, Castells-Garcia A, Gomez-Garcia PA, Vicario C, Carnevali D, AlHaj Abed J, Granados A, **Sebastian-Perez R**, Sottile F, Solon J, Wu CT, Lakadamyali M, Cosma MP. Transcription-mediated supercoiling regulates genome folding and loop formation. *Molecular Cell*, 2021 (in press).

I collaborated in elucidating the role of the transcription factor NKX1-2 in mouse early embryo development:

Nakagawa S,…, **Sebastian-Perez R**,…, Cosma MP. *Manuscript in preparation*.

I collaborated in exploring the use of stem cell therapy for human retina regeneration:

Bonilla-Pons SA,…, **Sebastian-Perez R**,…, Cosma MP. *Manuscript in preparation*.

Review article:

Pesaresi M, **Sebastian-Perez R**, Cosma MP. Dedifferentiation, transdifferentiation and cell fusion: in vivo reprogramming strategies for regenerative medicine. *FEBS Journal*, 2019 Mar; 286 (6):1074-1093.

Pesaresi M, **Sebastian-Perez R**, Cosma MP. Dedifferentiation, transdifferentiation and cell fusion: in vivo reprogramming strategies for regenerative medicine. FEBS Journal, 286 (6):1074-1093 (2019).

This PhD thesis was supported by a FI-AGAUR fellowship from the Government of Catalonia and the European Social Fund; the European Union's Horizon 2020 Research and Innovation Programme; and the Spanish Ministry of Science and Innovation.