Unraveling the mechanism of DNA methylation regulation during early embryo epigenetic reprogramming

Kristján Hólm Grétarsson

Thesis for the degree of Philosophiae Doctor

Supervisor/s:
Dr. Jamie A. Hackett

Doctoral committee:
Dr. Mathieu Boulard
Dr. Erna Magnúsdóttir
Dr. Alexander Aulehla

October 2020

UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

Afhjúpun stjórnunarferla DNA-metýleringar í endurforritun umframerfða snemma í fósturþroskun

Kristján Hólm Grétarsson

Ritgerð til doktorsgráðu

Leiðbeinandi/leiðbeinendur:
Dr. Jamie A. Hackett

Doktorsnefnd:
Dr. Mathieu Boulard
Dr. Erna Magnúsdóttir
Dr. Alexander Aulehla

Október 2020

UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

ISBN 978-9935-9516-8-7

Printing by Háskólaprent.

Reykjavik, Iceland 2020

Ágrip

Umframerfðastjórnprótein taka þátt í að stjórna mismunandi genatjáningu og koma þannig up sérhæfðu umframerfðamengi í frumum sem allar hafa sama erfðamengið. Það er mikilvægt að skilja hvernig umframerfðamerkjum er stjórnað snemma í fósturþroskun spendýra. Á fyrstu skrefum fósturþroskunar spendýra undirgengst fóstrið alhliða endurröðun umframerfða sem fjarlægir áunnin umframerfðamerki og jafnar út umframerfðamengin í báðum kynfrumum foreldranna. Slík endurröðun umframerfða felur í sér víðamiklar breytinar á histónumörkum ásamt því að fjarlægja og endursetja DNA-metýleringu sem tryggir þroskunarhæfni fóstursins. Skilningur á þeim ferlum sem stjórna endurröðun eða umframerfðastöðu er mikilvægur og hefur þýðingu fyrir umframerfðir, fjölhæfni og umframerfðastjórnun. Til að varpa ljósi á þá ferla sem stjórna heildarfærslum á DNA-metýleringu, var eins-frumu DNA-metýleringar næmu klögugeni byggt á hlutfallslegri tjáningu komið upp. Þetta klögugen var síðan notað í heilmengis CRISPR skimun í stofnfrumum úr fósturvísum músa (mESC). Lykilgen og stjórnferlar sem orsaka heildartap á DNA-metýleringu í frumum voru fundnir og staðfestir. Með þessu fundust yfir 20 gen sem að þegar þau voru slegin út hvert fyrir sig, ollu röskun í að fjarlægja heildar DNA-metýleringu í frumum ásamt því að nánari greining var gerð á hlutverkum *Cop1* og *Dusp6* í þessu ferli.

Auk þess voru DPPA2 og DPPA4 skilgreind sem nauðsynleg protín til að viðhalda staðbundnu umframerfðaástandi í umframerfðatilfærslum í fjölhæfum frumum. Með því að fjarlægja *Dppa2* eða *Dppa4* varð tap á H3K4me3 og aukning á DNA-metýleringu á genum með hlutverk í fósturþroska og þróunarfræðilega ungum LINE1 röðum sem DPPA2 binst við. Þessu fylgdi að gen með hlutverk í fósturþroska áunnu sér bælandi umframerfðaminni sem olli því að þessi gen voru ekki virkjuð í síðari vefjasérhæfingu. Þessi aukning á DNA-metýleringu sem var greind gæti verið mikilvæg fyrir stöðuga umframerfðaþöggun á DPPA2 bundnum genum með hlutverk í fósturþroskun og LINE1 röðum, því að fjarlæging á DNA-metýleringu var nóg til að endurvirkja genatjáningu og H3K4me3 að huta til þegar DPPA2 var ekki til staðar. DPPA2/4 móta því umframerfðamengið í fjölhæfum frumum með því að koma upp H3K4me3 og virka gegn aukningu á DNA methyleringu, þetta ferli hefur síðan verið aðlagað að þróunarfræðilega ungum LINE1 to að forðast umframerfðaþöggun í fjölhæfum frumum. Með þessari ritgerð er skilningur okkar aukinn á því hvernig umframerfðamerkjum er bæði stjórnað víðfemt yfir

erfðamengið og á staðbundin hátt í erfðamenginu snemma í fósturþroskun til að tryggja þroskunarhæfni.

Lykilorð:
DNA-metýlering, histónumörk, CRISPR-screen, fjölhæfni, umframerfðir

Abstract

Epigenetic regulators contribute to the control of distinct gene expression patterns by establishing specialized epigenomes in cells that share the same genome. How epigenetic marks are initially shaped in the early development is of great importance and incompletely understood. During mammalian development, the embryo undergoes genome-wide epigenome remodelling which equalises the distinct parental epigenomes and resets genomic potential. This epigenetic reprogramming requires extensive chromatin modification changes, including DNA methylation erasure and reacquisition, which enables developmental competence. Understanding the mechanisms that mediate either reprogramming or persistence of epigenetic states is currently an outstanding question and has implications for our understanding of inheritance, pluripotency, and epigenetic regulation. To uncover the mechanisms that regulate DNA methylation dynamics, I established a single-cell ratiometric global DNA methylation reporter and coupled it with genome-wide CRISPR screening in murine embryonic stem cells (ESCs) undergoing DNA demethylation transition. Key genes and regulatory pathways that drive global DNA hypomethylation were identified and validated, revealing over 20 gene perturbations that result in impaired global demethylation. I further characterised the roles for *Cop1* and *Dusp6* in this process in detail.

Additionally, I identified *Dppa2* and *Dppa4* as essential safeguards of focal epigenetic states during the epigenome transitions associated with pluripotency. Specifically, I found that deletion of *Dppa2* or *Dppa4* resulted in loss of H3K4me3 and gain of ectopic *de novo* DNA methylation at developmental genes and evolutionarily young LINE1 elements, which are specifically bound by DPPA2. As a result, developmental genes acquire a repressive epigenetic memory that leads to a failure to induce their appropriate activation in future lineage specification. Furthermore, the gain of DNA methylation might be important for stable epigenetic silencing at DPPA2 bound developmental genes and LINE1 elements as removal of DNA methylation causes their partial reactivation and gain of H3K4me3 in the absence of *Dppa2*. DPPA2/4 thereby sculpt the pluripotent epigenome by facilitating H3K4me3 and bivalency to counteract *de novo* methylation, a function co-opted by evolutionarily young LINE1 to evade epigenetic silencing during pluripotency. This thesis represents an advance in our understanding of both how epigenomic states are globally modulated, and how they are established

at specific loci, during early development to ensure genome competence multi-lineage differentiation.

Keywords:
DNA methylation, histone modificaitons, epigenetics, pluripotency, CRISPR-screen

Acknowledgements

The work in this project was done at the at the European Molecular Biology Laboratory (EMBL), Epigenetics & Neurobiology Unit in Rome, Italy.

I would like to thank my supervisor Jamie Hackett for all the support and guidance throughout my PhD. It truly has been a great time and learning experience and I am thankful for having had the opportunity to carry out research in the Hackett lab. I would also like to thank all current and past members of the Hackett lab for all the technical help and making my stay at EMBL Rome enjoyable along with all the other friends I met at EMBL.

Additionally, I would like to thank the amazing core facilities at EMBL, GAVEF (Jim Sawitzke), Genecore (Vladimir Benes) and Flow Cytometry Facility (Cora Chadick) for their help and contribution to the project.

I would like to thank the members of my PhD committee, Mathieu Boulard, Erna Magnúsdóttir and Alexander Aulehla for their scientific input, support and advice during the years of the project. Special thanks go to Erna Magnúsdóttir for her help in communicating with the University of Iceland and helping out with the thesis submission and planning of the defense.

Last but not least, I am especially thankful for my family, friends and my girlfriend Nainika for their love and support throughout my PhD.

Contents

List of abbreviations

H3K27me3	Histone 3 lysine 27 trimethlation
H3K4me3	Histone 3 lysine 4 trimethlation
H3K9me3	Histone 3 lysine 9 trimethlation
H3K36me3	Histone 3 lysine 36 trimethlation
mRNA	messenger Ribonucleic acid
KO	Knock out
TAM	Tamoxifen
CRISPR	clustered regularly interspaced short palindromic repeats
GFP	green fluorescent protein
BSL2	Biosafety level 2
PCR	Polymerase chain reaction
EM-seq	Enzyme Methyl sequencing
FBS	Fetal Bovine Serum
TE	Transposable elements
ESC	Embryonic Stem Cells
LIF	Leukemia inhibitory factor
ERK	Extracellular signal-regulated kinase
ZGA	Zygotic Genome Activation
LINE	Long interspersed nuclear elements
IAP	Intracisternal A-particle
EpiLC	Epiblast LikeCell
PGC	Primordial Germ Cell
H2B	Histone 2B
FGF	Fibroblast Growth Factors
pPB	plasmid PiggyBac
sgRNA	short guide RNA
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
5caC	5-carboxylcytosine
PCA	Principal component analysis
CUT&RUN	Cleavage Under Targets and Release Using Nuclease
RNA	Ribonucleic acid sequencing
TSS	Transcription Start Site

List of figures

List of tables

List of original papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

I Kristjan H. Gretarsson, Jamie A. Hackett.
Dppa2 and *Dppa4* counteract de novo methylation to establish a permissive epigenome for development. *Nat Struct Mol Biol* **27,** 706–716 (2020).

In addition, some unpublished data is presented.

Furthermore, in the appendix, addional publication contributions made during my PhD are listed.

Declaration of contribution

I performed all the experiments presented here and the bioinformatic analysis and wrote the thesis. The lentiviral production for CRISPR KO gRNA library was made by the GAVEF at EMBL Rome and NGS library preparation for RNAseq and EMseq samples were done at EMBL Genecore. I contributed to hypothesis generation, experimental planning and interpretation of results along with my supervisor, Jamie A. Hackett.

For original publication I, I performed all the experiments the majority of the bioinformatic analysis and contributed to the manuscript, written by Jamie. A. Hackett.

1 Introduction

Although almost every cell in a multicellular organism share the same DNA sequence, which originates from the initial fertilized oocyte, these cells have developed specialised functions. The emergance of this specialisation is one of the key questions in developmental biology. Over the last few decades we have gained valuable insight into the importance of careful regulation of gene expression. More specifically, the research of gene regulation and its inheritance in both development and in diseases make up the field of epigenetics. During development different cell types emerge as gene expression patterns are instigated and stabilized by epigenetic modifications (Reik, 2007). These epigenetic modifications can be directly on DNA (such as methylation) or various modifications of histone tails (Turner, 2007). Another layer of epigenetic regulation is ensuring competence for later development. By silencing genes or priming genes for later activation, cells ensure that they can undergo differentiation at the appropriate time. Of special interest are the events in the early embryo in mammals. A wave of epigenetic reprogramming takes place to erase parental traits of the germ cells and generate the naïve pluripotent state that has the capability of giving rise to different cell lineages through accumulation of different epigenetic marks. Additionally, dysregulation of epigenetic mechanisms can result in various diseases, for example in cancer cells DNA methylation is commonly altered often contributing directly to oncogenesis. The importance of understanding the basics of epigenetic regulation is therefore of immense value to learn more about development and disease progression. Here I will cover some of the key epigenetic modifications and their reprogramming in the early embryo.

1.1 DNA methylation in mammals

The addition of a methyl group to the fifth carbon of cytosine bases (5mC) is often found in the context of symmetrical CpG dinucleotides. This modification of DNA (DNA methylation) is a heritable and reversible epigenetic process (Schubeler, 2015). DNA methylation is generally associated with stable transcriptional repression, displaying important roles in stable repression at the inactive X chromosome, centromeres, imprinted genes, transposons and developmentally regulated genes (Greenberg & Bourc'his, 2019). How DNA methylation inhibits transcription is not entirely known as the methyl group does not induce silencing on its own (Schubeler, 2015). Major classes of transcription factors, including bHLH, bZIP and ETS, are sensitive to DNA methylation and show decreased DNA binding to their motif upon DNA

methylation (Yin et al., 2017). This offers an example of the silencing role of DNA, preventing the binding of activating transcriptional factors at their motifs and consequently inhibiting gene expression. It should be noted that vice versa, transcription factor binding can be instructive for local DNA methylation (Stadler et al., 2011). This has been shown by systematically inserting different DNA elements into the same loci which resulted in change in DNA methylation depending on the methylation-determining regions (MDRs) of the elements (Lienert et al., 2011). Additionally, DNA methylation is important for heterochromatin formation as the de novo methyltransferases function in complex with chromatin remodelers, H3K9me3 methyltransferases and histone deacetylases, ensuring repression at developmental genes during lineage commitment (Dennis et al., 2001; Fuks et al., 2001; Myant et al., 2011; Zhu et al., 2006). Other class of proteins can directly be recruited to methylated DNA, such as the methyl-CpG-binding domain (MBD) proteins and zinc finger proteins, to induce heterochromatin formation and gene silencing (Baubec et al., 2015; Hendrich & Bird, 1998; Y. W. Liu et al., 2012; Ren et al., 2018; Sasai et al., 2010). Interestingly the increased MBD protein binding at regions with higher methylation density is compatible with the discovery that DNA methylation at CpG rich regions correlates with stronger repression than DNA methylation at at CpG poor sequences (Baubec et al., 2013; Boyes & Bird, 1992).

In the mammalian genome, most CpG sites are found in small clusters (<1kb) with high CpG density (CpG islands or CGIs) and are generally demethylated while CpG sites outside of these clusters are sparse and highly methylated (Greenberg & Bourc'his, 2019). Promoters can be classified depending on their CpG density into classes of high, intermediate and low CpG density promoters (Messerschmidt et al., 2014). This classification is relevant for both the role of the promoters and for the presence and effect of DNA methylation on transcription. Over two thirds of mammalian promoters are CGI promoters, with housekeeping genes and some developmentally regulated genes often containing CGI promoters (Messerschmidt et al., 2014). While many unmethylated CGI promoters are ubiquitously active, like housekeeping genes, developmental genes are more regulated and inhibited by H3K27 trimethylation by PRC2 (Boyer et al., 2006). CGIs are not only found at promoters, over half of CGIs in the genome are found outside of promoters and are more frequently methylated during development (Illingworth et al., 2010). The role of DNA methylation at CpG poor regulatory regions outside of repeats and transposons is less clear but promoters with lower CpG density tend to be silenced and acquire DNA methylation during development (Schubeler, 2015; Stadler et al., 2011). Heterochromatin formation at those promoters has been suggested to be initiated by H3K9 methylation, followed

by de novo DNA methylation establishing stable silencing (Smith & Meissner, 2013). Therefore, the density and methylation status of CpG sites at regulatory elements in the genome is important for their functional state and regulation during development.

1.1.1 DNA methylation modifiers

DNA methylation can be separated into three different stages: Addition (or de novo), maintenance and removal of DNA methylation (Greenberg & Bourc'his, 2019). DNA methylation plays an important role in maintaining gene expression patterns and cell identity (Broske et al., 2009; Oda et al., 2013), indicating the need for faithful inheritance of DNA methylation patterns. DNA replication of symmetrical CpG methylated DNA results in two methylated parental strands and two daughter strands with unmethylated cytosine. For faithful inheritance of DNA methylation, the hemimethylated CpGs need to be fully methylated. The maintenance DNA methylation transferase, DNMT1, is recruited to the hemi methylated sites through UHRF1 (Bostick et al., 2007; Sharif et al., 2007). UHRF1 binds hemimethylated CpGs and H3K9me2 resulting in the ubiquitination of H3 residues by the E3 ligase activity of UHRF1 (Nishiyama et al., 2013; Rothbart et al., 2012). UHRF1 then binds DNMT1 which releases the autoinhibitory configuration of DNMT1 (Song et al., 2011; Takeshita et al., 2011), allowing it to bind the newly ubiquitinated H3K18 and H3K23 and methylating the hemi methylated CpG sites (Ishiyama et al., 2017). The importance of DNA methylation maintenance can be seen both in early development and in somatic cells. DNA methylation maintenance is essential for mammalian development as the knock out of either DNMT1 or UHRF1 results in early embryonic lethality in mice (E. Li et al., 1992a; Sharif et al., 2007). This is caused by failure in establishing lineage specification following the demethylated pluripotency stage. In fact, pluripotent cells are unique in their tolerance for lack of DNA methylation as most differentiated cells in culture are not sustainable without DNA methylation, where it has a role in repressing the transcription of unwanted genes and transposons (Jackson et al., 2004; Walsh et al., 1998). Therefore, DNA methylation maintenance is essential in development and in most cell types.

DNA methyltransferases (DNMTs) are a group of enzymes that catalyse the methyl group transfer to the cytosine. The DNA methyltransferase enzymes DNMT3A and DNMT3B are responsible for de novo methylation in mammals (Okano et al., 1999; Okano et al., 1998). In addition to a methyltransferase domain, DNMT3A/B contain ATRX-DNMT3-DNMT3L (ADD) and PWWP domains. DNMT3s are recruited to chromatin through binding of the ADD domain to unmethylated Lys 4 on histone H3 (H3K4) (Ooi et al., 2007; Otani et al., 2009; Y. Zhang et al., 2010). Additional layer of

regulation by autoinhibition is observed by the DNMT3s. The ADD domain binds the catalytic domain of DNMT3A (ADD-CD, interaction) and binding of ADD to histone through unmethylated H3K4 releases the enzyme from its autoinhibitory conformation and enables methylation transferase activity (X. Guo et al., 2015). The PWWP domain has been shown to have affinity for methylated H3K36 (Dhayalan et al., 2010). Moreover, the PWWP domain of DNMT3B binds H3K36me3 at transcribed genebodies, resulting in DNA methylated genebody of active genes and inhibition of cryptic transcriptional initiation (Baubec et al., 2015; Neri et al., 2017). Unlike DNMT3B, the PWWP domain of DNMT3A has higher affinity towards H3K36me2 which is more commonly found at intergenic regions, causing H3K36me2 intergenic regions to be DNA methylated by DNMT3A (Weinberg et al., 2019), demonstrating different roles for DNMT3A and DNMT3B. Double knock out of *Dnm3ta* and *Dnmt3b* results in early embryonic lethality as they carry out de novo methylation in the post implantation embryo (Okano et al., 1999), supporting the essential role of DNA methylation for mammalian development. Another *Dnmt3* gene, *Dnmt3l* can stimulate the de novo methyltransferases but lacks the methyltransferase domain (Suetake et al., 2004). DNMT3L is involved in establishing genomic imprints in both oocytes and sperm along with DNMT3A (Kaneda et al., 2004).

Removal of DNA methylation can occur in at least two different ways in mammals. Removal can be achieved by blocking the DNA maintenance system which results in hemimethylated DNA on both daughter strands and causing DNA methylation to be eventually diluted through series of replications (von Meyenn et al., 2016). Alternatively, DNA methylation can be removed through active removal by the TET enzymes. Methylated cytosine (5mC) is oxidized to 5hmC and further to 5fC and 5caC by the TETs (He et al., 2011; Ito et al., 2010; Ito et al., 2011; Popp et al., 2010; Tahiliani et al., 2009). 5fC and 5caC can then be demethylated through base excision repair pathway after base removal by TDG (He et al., 2011; Weber et al., 2016). DNA demethylation can also occur as a mix of both. In fact, DNMT1 has 60x lower affinity towards hemihydroxymethylated DNA than hemimethylated DNA, demonstrating a role for the TETs in passive demethylation (Hashimoto et al., 2012). The TET proteins has been shown to be important for development, meiosis, imprinting and enhancer regulation (Schubeler, 2015). Loss of *Tet1* or *Tet2* is compatible with pluripotency but does impact differentiation (Dawlaty et al., 2013; Dawlaty et al., 2011; Koh et al., 2011), demonstrating a role for TETs in establishing lineage specific gene expression patterns. Moreover, triple knock out embryos, $Tet1^{-/-}$; $Tet2^{-/-}$; $Tet3^{-/-}$ are incompatible with embryo development and exhibit gastrulation failure (Dawlaty et al., 2014). Similar to de novo and DNA methylation maintenance the ability to locally remove

methylation to regulate the tissue specific methylome is essential during development.

1.2 Histone Modifications

DNA is wrapped around histone proteins making up the nucleosome and along with other non-histone proteins and RNA together form the chromatin. Just as DNA can be modified with methylation, histones can be modified by a range of different chemical groups, methyl or acetyl or protein groups such as ubiquitin (Jambhekar et al., 2019). Histone modifications can regulate gene activation or repression and various cellular processes, but it is not just the type of modification, the location of the modification on the histone tail is also important. Establishing, maintaining and removing these modifications is important to establish proper cellular transcriptional networks and their inheritance for development and cell fate decisions. It is important for the histone modifications to be read correctly to carry out the appropriate message by initiating structural changes or recruit functional effectors. Proteins or even domains can read histone methylation; plant homeodomain (PHD) fingers is a class of readers that binds lysine residues and Tudor domain bind both methylated lysine and arginine residues (Friesen et al., 2001; Huang et al., 2006; Shi et al., 2006). Here the focus will be on different histone methylations with roles in both activation and repression of gene expression.

1.2.1 H3K4me3

Methylation of the lysine 4 (K4) Histone H3 (H3K4) correlates with gene activity with mono-methylation typically marking enhancers and tri-methylation found at active promoters (Heintzman et al., 2007). H3K4 is methylated by the SETD1/MLL family of methyltransferases and functions in larger group of proteins called COMPASS or COMPASS related complexes and have been shown to have specialised role in H3K4 methylation (Piunti & Shilatifard, 2016). Set1A/B are the main H3K4 tri-methylases in mammals and put down default H3K4me3 at most active promoters (Wu et al., 2008). The SET1A/B complex are specially targeted to CGIs where they are recruited by the binding of CFP1 (CXXC) to unmethylated DNA (Thomson et al., 2010). Whether H3K4me3 is instructive for transcriptional activation or not has been debated. Deletion of CFP1 in ESC causes loss of H3K4me3 at most active promoters without loss of gene expression and aberrant gain of H3K4me3 at regulatory elements (Clouaire et al., 2012). This demonstrates the importance of targeting of H3K4me3 and that H3K4me3 might only be needed for initiation of expression or low levels of H3K4me3 could be sufficient to maintain transcriptional activity (Clouaire et al., 2012). KMT2B (MLL2) has been shown to specifically methylate H3K4 at bivalent regions while SET1 is missing at those regions (Denissov et al., 2014; D. Hu et al., 2013; Y. L. Xiang et al., 2020). H3K4me3

at bivalent promoters is not lost upon CFP1 deletion as KMT2A/B contain CXXC domain and are able to bind unmethylated CpG on their own. A recent study observed the uncoupling of H3K4me3 and transcription at KMT2B dependent promoters upon removal of inhibiting modifications (Douillet et al., 2020) suggesting a role for H3K4me3 in limiting the presence of inhibiting marks, such as DNA methylation or H3K27me3.

H3K4 mono-methylation that marks active enhancers is established by KMT2C and KMT2D functioning in COMPASS related complexes (D. Q. Hu et al., 2013). Recent studies in using catalytically deficient KMT2C/D have demonstrated that KMT2C/D are important for enhancer activity and promoter transcription independently of H3K4me1 (Dorighi et al., 2017; Rickels et al., 2017). With the catalytically mutant fly being viable without any H3K4me1 (Rickels et al., 2017), overall indicating an alternative role of KMT2C/D in maintaining enhancer activation and viability outside of its methyltransferase activity.

After H3 lysine methylation has been put down, it can be removed with lysine specific demethylases but eighteen jumonji domain class demethylases have been identified (Jambhekar et al., 2019). These demethylases are important for the regulation of histone methylation by removing or preventing unwanted methylation.

1.2.2 H3K27me3

H3K27me3 is repressive and reversible histone modification, allowing for dynamic regulation at genes making it important in development when genes need to be activated or repressed at high rate (Boyer et al., 2006; Jambhekar et al., 2019). H3K27 is methylated by a group of PcG proteins first identified in Drosophila, that form the polycomb repressive complex 2 (PRC2) (Lee et al., 2006). PRC2 core components of PRC2 include the histone methyltransferase EZH2 (or EZH1) which targets H3K27 for methylation, along with EED and SUZ12 (Gokbuget & Blelloch, 2019). Methylation of H3K27me3 by PRC2 results in the binding of another polycomb complex, PRC1 which maintains silencing by chromatin compaction and combined PRC1 and PRC2 are essential for polycomb-mediated transcriptional repression (Piunti & Shilatifard, 2016) and both are essential for early embryonic development (O'Carroll et al., 2001; Voncken et al., 2003). In Drosophila PRC2 is recruited to PcG responsive elements (PRE) which are few hundred base pairs and contain binding sites for multiple transcription factors that recruit PRC2 (Chiang et al., 1995). How PRC2 localise to specific regions in mammalian cells is unclear, as common PREs have not been identified in mammalian cells. PRC2 is known to bind H3K27me2/3 through its EED subunit, in fact this binding

leads to activation of the methyltransferase activity of PRC2 (Margueron et al., 2009). PRC2 has been shown to bind by default to unmethylated CGI promoters (Mendenhall et al., 2010) and effective targeting of PRC2 to CGI depends on PRC2 sub-complex with MTF2 and JARID2 (H. J. Li et al., 2017; Oksuz et al., 2018). Long range interactions dependening on PRC2 can spread the H3K27me3 mark from the nucleation sites and form large repressive domains (Oksuz et al., 2018). Furthermore, noncanonical PRC1 can recruit PRC2 on chromatin through H2AK119ub (Blackledge et al., 2014), further reinforcing the binding of each other to chromatin. However, multiple mechanisms limit PRC2 binding or activity. Both active transcription and DNA methylation counter PRC2 recruitment (Di Croce & Helin, 2013; Riising et al., 2014). In fact, global loss of DNA methylation causes a decrease of H3K27me3 at CpG islands through the ability of PRC2 to bind newly unmethylated regions, causing eventual dilution of H3K27me3, suggesting that the DNA methylome is required for proper H3K27me3 distribution (A. B. Brinkman et al., 2012; Reddington et al., 2013; Schwammle et al., 2016). Conversely, loss of EED results in gain of DNA methylation at non-CGI regions in naïve ESC, involving PRC2 in regulating the DNA methylome in naïve ESC and prevent widespread DNA methylation (van Mierlo et al., 2019b). Moreover, PRC2 activity and H3K27me3 are limited by H3K36me2, often accumulated at intragenic regions, as depletion of the H3K36 methyltransferase NSD1 resulted in spreading of H3K27me3 domains (Streubel et al., 2018). Overall, this indicates that the epigenetic landscape is important in defining PRC2 binding and H3K27me3 domains.

PRC2 is targeted to key developmental genes, such as the Hox clusters, and their enhancers in pluripotent cells and ensures effective silencing to maintain pluripotency in ESC (Boyer et al., 2006). Remarkably naïve mESC are viable without H3K27me3 or PRC2 (Chamberlain et al., 2008; Riising et al., 2014; van Mierlo et al., 2019a) but acquire early priming marks such as increased DNA methylation, suggesting a role for PRC2 in maintaining the naïve epigenome and repressing early priming (van Mierlo et al., 2019a). The reversible nature of PRC2 regulation allows for swift and effective activation of these developmental genes during differentiation. Comparison of the epigenomic landscapes of pluripotent and differentiated human cells revealed the expansion of repressive H3K27me3 and H3K9me3 regions in the differentiated fibroblasts (Hawkins et al., 2010). The expansion of repressive chromatin induced the silencing of pluripotency related genes and developmental genes of alternative lineage commitment (Hawkins et al., 2010). Additionally, deletion of core components of PRC2 results in embryonic failure during gastrulation when tissue specification and differentiation take place, but is compatible with pre-implantation development (O'Carroll et al.,

2001; Piunti & Shilatifard, 2016). This highlights the essential role of PRC2 and H2K27me3 in the silencing of developmental genes and specification of tissue specific transcriptional networks.

1.2.3 Bivalency

Both the repressive H3K27me3 and the active H3K4me3 can co-exist at some promoters (bivalent promoters) of developmental genes in pluripotent cells (Bernstein et al., 2006). These promoters are kept in a poised transcriptional state with both activating and repressing marks theorised to be quickly resolved to either activated or repressed state upon differentiation (Voigt et al., 2013). Loss of H3K4me3 at bivalent domains in ESC lacking KMT2B hinders their ability to differentiate (Mas et al., 2018) and lack of KMT2B is not compatible with early development (Glaser et al., 2006) but remarkably loss of KMT2B after E11.5 is sustained (Glaser et al., 2009). Interestingly, recent data indicates SET1A (not KMT2B) as the H3K4 methyltransferase at resolved bivalent promoters, suggesting a switch between KMT2B to SET1A during pluripotency to differentiation transition (Sze et al., 2017) and potentially explaining why KMT2B is only briefly required *in vivo*. In the early mammalian embryo, developmental genes acquire bivalency in the epiblast and loss of KMT2B causes failure to induce proper activation of a subset of key bivalent genes (Y. L. Xiang et al., 2020). The global gain of DNA methylation observed during the transition from the pre-implantation to epiblast embryo could be important in establishing H3K27me3, and therefore bivalency, at developmental promoters. However, this model of bivalency and its importance in lineage commitment has been challenged as cells that don't have differentiation potential retain several bivalent domains (Y. L. Xiang et al., 2020), indicating the need for further research on the role and regulation of bivalent promoters.

1.3 Transposons and their function in the mammalian genome

Transposons and related truncated transposon derived sequences make up a large fraction of the mammalian genomes. Transposons are mobile elements in the genome and the most abundant class of transposons in the mouse genome are retrotransposons (Deniz et al., 2019). Retrotransposons move through an RNA intermediate and contain elements such as LINEs, SINEs, and LTRs (Rodriguez-Terrones & Torres-Padilla, 2018). Although most transposon insertions are harmless for the host, some can be destructive, therefore their regulation is important for genome stability (Galli et al., 2005; Shukla et al., 2013). In mammals, a major pathway for silencing of transposable elements (TE) consists of the Kruppel-associated box zinc finger proteins (KRAB-ZFPs) that can bind to the transposable elements in a

sequence specific manner (Imbeault et al., 2017; Jacobs et al., 2014). KRAB-ZFP binding at transposable elements recruit TRIM28 that function as a scaffold to recruit further repressive modifiers including SETDB1 to silence expression through H3K9me3 and eventual heterochromatin formation (Matsui et al., 2010; Rowe et al., 2010). Furthermore, this heterochromatin formation results in DNA methylation that is often seen at transposable elements. In fact, transposons are thought to be a major driver of the evolution of DNA methylation patterns in mammals (Yoder et al., 1997). In the germline, specialised pathways have evolved as a consequence of an evolutionary conquest between the host and the transposons during this critical period. In the mouse germ line retrotransposons are silenced through Piwi interacting RNA-mediated mechanisms resulting in DNA methylation of the transposon element (Aravin et al., 2007; Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008). Recently a novel DNA methyltransferase, DNMT3C, was identified that has a specific role in methylating retrotransposons in male germ cells (Barau et al., 2016). Overall, this represents the evolutionary pressure on the host to suppress retrotransposons in the germ line.

However, the exact role of DNA methylation in the repression of transposable elements is not always clear. In mESC endogenous retroviruses (ERVs) and intracisternal A particle (IAP) elements are highly methylated but their repression depends on SETDB1 but is independent of DNA methylation (Hutnick et al., 2010; Karimi et al., 2011; Matsui et al., 2010; Rowe et al., 2010). However, recent studies using conditional KOs or DNA methylation transitional models have reported transient activation of transposon elements upon global loss of DNA methylation in mESC that were eventually silenced by repressive chromatin (H3K27me3 or H3K9me3) or endosiRNAs (Berrens et al., 2017; Sharif et al., 2016; Walter et al., 2016). DNA methylation at transposon elements such as IAP even persists through the global wave of DNA demethylation in the PGCs, (Hackett et al., 2012; Seisenberger et al., 2012) demonstrating the importance of DNA methylation in maintaining transposon silencing. Finally the role of DNA methylation in supporting genome stability or accelerating mutagenic deamination at transposons leading to their eventual inactivation at transposons should not be overlooked (Greenberg & Bourc'his, 2019).

In the pre-implantation embryo, the emergence of pluripotency is followed by a transcriptional activation of retrotransposons (Fadloun et al., 2013). This activation of transposons was previously thought to be a side effect of the loss of heterochromatin domains during epigenetic reprogramming (Fadloun et al., 2013; Friedli et al., 2014). However, LINE1 has been suggested to play an essential role in both mESC and in pre-implantation development (Beraldi et

al., 2006; Jachowicz et al., 2017; Percharde et al., 2018). Firstly, LINE1 RNA acts as a scaffold with Nucleolin and TRIM28 to silence Dux and activate rRNA synthesis (Percharde et al., 2018). Secondly, the expression of LINE1 has a role in controlling chromatin accessibility. Repression of LINE1 results in a decrease in chromatin accessibility and prolonged activation of LINE1 prevents chromatin compaction (Jachowicz et al., 2017). These recent findings highlight the potential altruistic role of LINE1 expression, indicating a possible co-adaptive role for LINE1 in development.

Transposable elements (TE) are not conserved across species and within mammalian genomes there is high variation of TE classes. Notable differences are between the mouse and human genomes, for example, while the human genome has abundance of SINE and DNA transposons, the mouse genome is largely depleted of these elements (Barau et al., 2016). This evolutionary mobility of transposons can lead to adaption of their derived sequences as functional regulatory elements in the host genome (Kunarso et al., 2010; Lu et al., 2014; Lynch et al., 2011; Schmidt et al., 2012; T. Wang et al., 2007). In fact, TE have been shown to contain species specific binding sites of transcriptional factors (Imbeault et al., 2017; Schmidt et al., 2012; T. Wang et al., 2007). Additionally, 25% of OCT4 and NANOG binding sites in ESC were at least partly occupied by ERVs, contributing to the different binding profiles of the transcription factors between human and mouse (Kunarso et al., 2010). New or alternative promoters have originated from TE sequences often providing functional promoters for tissue specific expression of genes. In fact, 6-30% of transcripts initiate within TE sequences (Faulkner et al., 2009). Transposable elements can therefore shape the regulatory network and drive species specific differences (Lu et al., 2014; Lynch et al., 2011). Just as TE shape the host genome, TE have the ability to use bound host transcription factors to increase expression and the potential to be transposed in the germline (Kano et al., 2009). This has been shown for the DUX transcription factor which activates early embryo related genes along with human and mouse LTR elements, ERV (De Iaco et al., 2017a; Hendrickson et al., 2017b). Overall, this demonstrates how transposon derived sequences have been adapted in the genome.

1.4 Epigenetic reprogramming in the early mammalian embryo

The development of an organism from two gametes involves multiple carefully controlled events such as cell divisions, differentiation and linage commitment. Epigenetic marks like DNA methylation and histone modifications carry information for cell identities, restricting the transcriptional network and requiring the need to be reprogrammed to generate the epigenetic plasticity

required for later development. Epigenetic reprogramming in the early embryo converges with pluripotency and is thought to establish epigenetic competence for later development (Reik, 2007; Q. H. Xu & Xie, 2018). That is, the ability to transcriptionally respond to future inductive signals for multiple different lineages. Developmental genes are repressed during pluripotency but become appropriately activated upon differentiation.

During early embryo reprogramming the zygote must lose the germ cell specific DNA methylation from both parental genomes. Initial models of demethylation in the zygote indicated active removal of DNA methylation from the highly methylated paternal genome via TET3 resulting in increase in 5hmC (Gu et al., 2011; Iqbal et al., 2011), followed by passive, replication dependent dilution of DNA methylation in both parental genomes. Recent findings using whole genome bisulfate sequencing with single nucleotide resolution of DNA methylation challenged these initial findings (F. Guo et al., 2014; Shen et al., 2014; L. Wang et al., 2014). In fact, they demonstrated that most zygotic DNA methylation is lost through a replication independent manner. More recent work suggested a role for TET3 in countering active de novo methylation in the early embryo and the paternal genome may undergo active demethylation through an unknown mechanism (Amouroux et al., 2016). At the blastocyst stage the pluripotent inner cell mass (ICM) achieves the lowest point of DNA methylation during pre-implantation epigenetic reprogramming (Smith et al., 2012). Along with imprinted regions, DNA methylation is also found at IAP retrotransposons and centromeric heterochromatin after epigenetic reprogramming in the early embryo (Messerschmidt et al., 2012; Smith et al., 2012; L. Wang et al., 2014). Maintaining DNA methylation at these regions is essential for development by inhibiting extensive retrotransposon activation, and maintaining chromosomal stability and methylation asymmetry at imprinted genes. Epigenetic reprogramming might balance genome-wide resetting of epigenetic marks to establish competent state for development and locally regulated maintenance.

To form the totipotent zygotic epigenome the two parental germ cells need to undergo extensive epigenetic reprogramming and their vastly different epigenomes become compatible for this transformation (Q. H. Xu & Xie, 2018). In oocytes, de novo methylation is acquired after birth and in a transcription dependent manner, resulting in a unique methylome (Smallwood et al., 2011; Veselovska et al., 2015). Intergenic and repressed genes are poorly methylated, forming partially methylated domains (PMDs) but transcribed gene bodies are fully methylated (FMDs). As oogenesis proceeds noncanonical patterns of histone modifications are established (Stewart et al., 2015). H3K4me3 forms broad domains at promoters and intergenic sites as oocytes

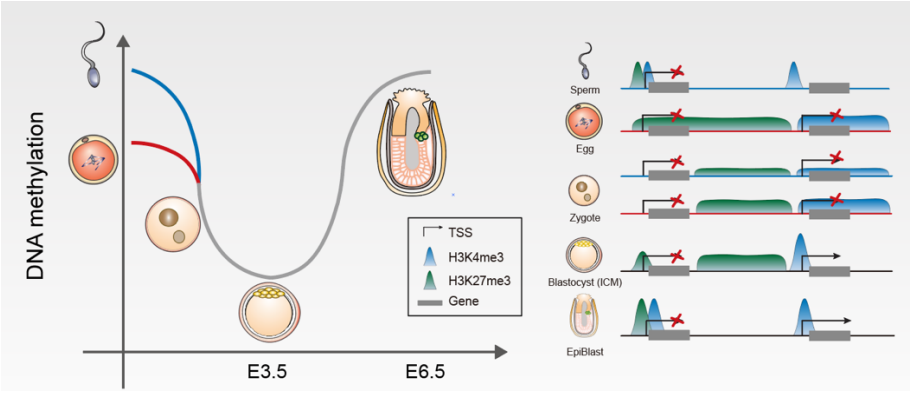


Figure 1. Epigentic reprogramming in the early mouse embryo.

Epigentic reprogramming in the early mouse embryo, demonstrating DNA methylation and histone modificaitons dynamics.

become full-grown and reach a silenced state (Dahl et al., 2016; B. Zhang et al., 2016). In fact, it has been suggested that this H3K4me3 landscape is involved in genome silencing in full grown oocytes (Andreu-Vieyra et al., 2010; B. Zhang et al., 2016). At intergenic regions and gene deserts large domains of H3K27me3 are established (Zheng et al., 2016). Both H3K27me3 and H3K4me3 overlap with different PMDs which is in line with the antagonism between DNA methylation and both H3K4me3 and H3K27me3 (Ooi et al., 2007; Reddington et al., 2013). FMDs in oocytes are however marked with H3K36me3 and defects in H3K36me3 lead to oocyte maturation defects and one cell arrest after fertilization (Q. Xu et al., 2019). Loss of H3K36me3 leads to aberrant DNA methylome, invasion of H3K4me3 and H3K27me3 into H3K36me3 marked regions and H3K4me3 at methylated imprinting control regions. This indicates that the non-canonical establishment of the histone modifications and DNA methylation during oocyte maturation is essential for embryonic development. During spermatogenesis most histones are replaced by protamines, however histone modifications are still retained and are mostly canonical (Lesch et al., 2016). After fertilization protamines of the paternal genome are replaced with histones. These histones are hyperacetylated and have more monomehtylation marks H3K4me, H3K9me3, H3K27me and H4K20me rather than the trimethylated (me3) marks and less heterochromatin features in general (Burton & Torres-Padilla, 2014). Thus, the two differentiated germ cells display greatly different epigenetic landscapes.

During pre-implantation development H3K4me3 and H3K27me3 show different dynamics. The oocyte specific and noncanonical broad H3K4me3 domains are removed in the two-cell stage with subsequent re-establishment of strong H3K4me3 peaks at promoters (Dahl et al., 2016; B. Zhang et al.,

2016). These H3K4me3 peaks are stable until the blastocyst with only minor fluctuations on weak promoters and distant peaks. This removal of broad H3K4me3 depends on zygotic transcription but not DNA replication (B. Zhang et al., 2016). However, the paternal genome is depleted of H3K4me3 suggesting extensive reprogramming following fertilization. Paternal H3K4me3 is seen again from the 2-cell stage and by post-implantation allelic H3K4me3 differences are not seen (B. Zhang et al., 2016). Oocyte H3K27me3 patterns are only inherited at non-promoter distal regions and are depleted at developmental gene promoters such as Hox following fertilization (Zheng et al., 2016). Therefore, bivalency is absent in the pre-implantation embryo and the lack of H3K27me3 at developmental promoters implies another mechanism that keeps them silenced during that time. However, bivalency is observed after implantation, following the re-establishment of H3K27me3 at developmental promoters and strong enrichment of H3K4me3 at their promoters in the E6.5 epiblast (Y. Xiang et al., 2020). Thus, epigenetic reprogramming establishes the epigenetic plasticity needed for totipotency and involves the decompaction of condensed genomes of the germ cells.

1.4.1 Imprinted promoters

Imprinted genes are genes expressed exclusively from either the maternal or the paternal allele, forming a monoallelic expression pattern that is essential for development (Ferguson-Smith, 2011). The genes are found in clusters controlled by an imprinted control region (ICR), which are often CpG-rich. The expression pattern is regulated by parentally different epigenetic states at the imprinted control regions which control the expression of neighbouring genes (Lin et al., 2003; Wutz et al., 1997). The establishment in the parental germlines and the maintaining of epigenetic parental asymmetry at imprinted control regions is essential for proper development. Most ICRs are targeted by de novo methylation in the unique methylome of oocytes and overlap with particularly CpG rich CGIs while the 3 paternal ICR are CpG poor (B. Zhang et al., 2016). Loss of DNA methylation at the ICR results in loss of imprinted gene expression (Bourc'his et al., 2001). After fertilization, the parental epigenomes are equalised apart from imprinted control regions which retain their parent of origin epigenetic status (Smith et al., 2012; L. Wang et al., 2014). During preimplantation reprogramming and global loss of DNA methylation a role for the zing finger proteins ZFP57 and ZFP445 and Trim28 in protecting imprints by maintaining DNA methylation has been demonstrated (X. J. Li et al., 2008; Messerschmidt et al., 2012; Quenneville et al., 2011; Takahashi et al., 2019). TRIM28 recruits DNMT proteins and the H3K9me3 methyltransferase SETDB1 to ensure the maintenance of imprints through both histone modification and DNA methylation. Chromatin associated with ICR are allele specific, H3K9me3, H4K20me3 and H3K27me3 are found at the methylated ICR alleles

while the unmethylated ICR carry H3K4me3 and H3/H4 acetylation (Kacem & Feil, 2009). Recently, a set of non-canonical imprinted genes dependent and marked by H3K27me3 rather than DNA methylation were identified (Inoue et al., 2017). These genes are expressed from the paternal allele only in the pre-implantation embryo and are mostly gone in the epiblast (E6.5) as opposite to the life-long maintained imprinted genes. Eventually in the germline, active removal of DNA methylation by TETs is needed to reset the DNA methylation status at imprinted genes (Hackett, Sengupta, et al., 2013). The epigenetic regulation at imprinted genes demonstrate an interesting case of epigenetic memory during development.

1.5 Stem cells and pluripotency

Pluripotent cells have the potential to give rise to all the lineages of the embryo and the germline, gathering a great interest in their behaviour and ability. Cells derived from the inner cell mass of the pre-implantation blastocyst can be maintained in vitro and their pluripotency potential is preserved as they can contribute to chimeras after replacement to another embryo (Nichols & Smith, 2009). Cultured cells can be massively expanded, making them easier to study and perform experiments that have traditionally relied on a large number of cells. Cells derived from the mouse inner cell mass of the pre-implantation blastocyst, mouse embryonic stem cells (mESC), have been studied intensively to understand the nature of epigenetic and pluripotency regulation. mESC can be cultured *in vitro* by using leukaemia inhibitory factor (LIF) or by co-culture of fibroblasts (feeders) to activate STAT3 and serum to induce the inhibition of differentiation signalling. However, by maintaining ESC in serum a number of different and conflicting signalling pathways are activated which causes transcriptional and morphological heterogeneity of serum cultured ESC (Marks et al., 2012). The pluripotency of ESC is based on the core transcription factors for pluripotency, OCT4, SOX2 and NANOG (Silva & Smith, 2008a). This regulatory circuitry maintains pluripotency but at the same time activates ERK signalling triggering differentiation cues. In fact, ESC in serum can fluctuate between different epigenetic states through active FGF/ERK signalling which is constantly reinforcing differentiation signalling resulting in spontaneous differentiation. Serum grown ESC have therefore been labelled as being in a ‘metastable state’ with the heterogeneous expression of many developmental genes (Marks et al., 2012) (Galonska et al., 2015; Kolodziejczyk et al., 2015).

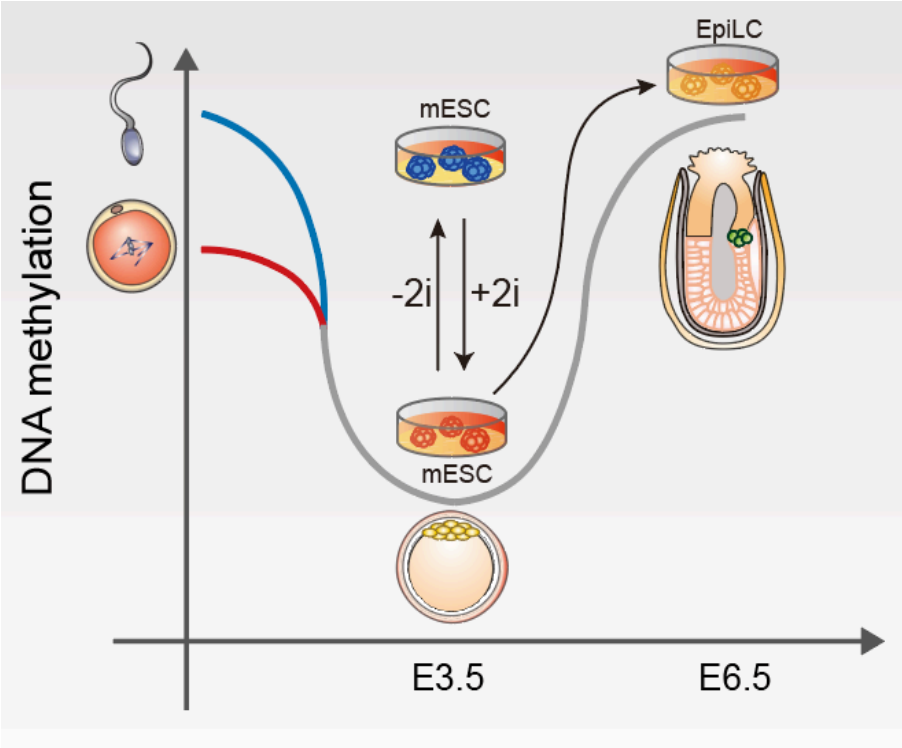


Figure 2. *In vitro* model for DNA methylation reprogramming in the early embryo.

Schematic of the *in vitro* model for DNA methylation reprograming in the early embryo and the corresponding states *in vivo*.

Culturing ESC without serum and with two cell signalling inhibitors (2i), the FGF/ERK signalling inhibitor known as PD0325901 and the GSK3 inhibitor known as CHIRON (CHIR99021), enhances the pluripotency network and blocks differentiation networks (Ying et al., 2008). The GSK3 inhibitor establishes the canonical Wnt pathway and maintains self-renewal of ESC while the MERK/ERK signalling inhibitor blocks differentiation signalling (Ying et al., 2008). 2i ESC are considered to have ground state pluripotency, with more homogeneous transcriptional and epigenetic state along with *Nanog* and *Prmd14* expression and a more permissive epigenetic state resembling the epiblast stage of the late blastocyst, including hypomethylated genome and stable silencing of differentiation genes (Marks et al., 2012) (Galonska et al., 2015; Kolodziejczyk et al., 2015). During development, a naïve pluripotent stage marked by expression of key pluripotency factors and derestricted epigenome emerges in the epiblast cells at the blastocyst stage. This transient state of pluripotency then becomes primed with the beginning of the post-implantation development. The classical serum grown ESC are thought to be

primed while 2i/LIF ESC are closer to the naïve epiblast in the late blastocyst (Hackett & Surani, 2014). However, prolonged 2i/LIF culture of ESC causes loss of parental imprints and impaired autonomous development. However, reducing the concentration of the ERK signalling inhibitor, PD0325901, provided culture conditions with higher global DNA methylation than in full 2i/LIF conditions and stability of genomic imprints (Choi, Huebner, et al., 2017; Yagi et al., 2017).

One of the key differences between serum grown ESC and naïve ESC is the global DNA methylation levels. Naïve ESC are hypomethylated paralleling the methylation status of the ICM while serum grown ESC are generally hypermethylated (60-70%) resembling the postimplantation blastocyst cells (Ficz et al., 2013; Habibi et al., 2013). Low global DNA methylation is generally associated with naive pluripotency (Leitch et al., 2013). In fact, lack of the DNA methylation maintenance system or the de novo methylation system are compatible with ESC derivation and proliferation. However, ESC without the potential to gain DNA methylation in such KO models fail to differentiate and therefore these cells are not pluripotent as they cannot contribute to all embryonic lineages (Jackson et al., 2004). This demonstrates the requirement for careful regulation of DNA methylation in early reprogramming of methylation as compatible methylome is needed at multiple stages. The DNA hypomethylation in naive ESC has been suggested to be regulated in part by PRDM14. PRDM14 is a transcription factor that co-occupies and regulates a number of genes with the core pluripotency factors (OCT4, SOX2 and NANOG) and its expression is limited to PGCs and pluripotent cells (Yamaji et al., 2008). Moreover, PRDM14 protects against somatic fates by repressing ERK signalling to safeguard a naïve pluripotent state (Grabole et al., 2013; Yamaji et al., 2013). Compared to Serum/Lif ESC naïve ESC upregulate PRDM14 which in turn represses the de novo methyltransfereses through direct transcriptional silencing (Ficz et al., 2013) and G9a mediated protein degradation (Sim et al., 2017). PRDM14 through combined de novo methylation repression and TET recruitment at target loci, ensures global hypomethylation in naïve ESC(Okashita et al., 2014). In fact, Prdm14 perturbation in 2i/LIF ESC results in elevated global DNA methylation (Leitch et al., 2013). However, the *Prdm14* knock out only partially explains the gain of DNA methylation and over expression of *Prdm14* in serum grown ESC is not sufficient to drive genome wide demethylation suggesting alternative pathways that could regulate the methylome in ESC (Hackett, Dietmann, et al., 2013).

1.6 The Dppa family

During this work, *Dppa2* and *Dppa4* and their role in DNA methylation were investigated. Additionally, the interest in the role of *Dppa3*, another Dppa family member, in DNA methylation has re-emerged. All three proteins belong to the family of developmental pluripotency associated (DPPA) proteins which were identified as novel markers of ESC with expression pattern similar to Oct4, that is expression in pluripotent cells and not in somatic cells (Bortvin et al., 2003). These Dppa family members share a SAP DNA binding domain which implicates wide range of functions regarding chromosomal organization and RNA biology, similar expression pattern and overlapping perturbation phenotypes.

1.6.1 Dppa2 and Dppa4

Dppa2 and *Dppa4* were identified as genes with specific expression in ESC, initially termed ES cell associated transcripts (*Ecat15-1* and *Ecat15-2*) (Mitsui et al., 2003). Their expression is limited to pluripotent cells, ICM of the mouse blastocyst, and the developing primordial germ line and their expression is swiftly repressed during differentiation (Maldonado-Saldivia et al., 2007). *Dppa2* and *Dppa4* are closely related with 32% identity and 47% similarity in their amino acid sequences and both contain putative SAP domains (Aravind & Koonin, 2000; Maldonado-Saldivia et al., 2007). Moreover, a conserved 73 amino acid sequence is found at the C terminal regions of both *Dppa2* and *Dppa4* (Maldonado-Saldivia et al., 2007) and the C terminal region of DPPA4 has been shown to interact with histone H3 and the N terminal with DNA (Masaki et al., 2010).

Earlier studies using short hairpin RNA (shRNA) mediated knockdown noted induced differentiation when targeting *Dppa2* or *Dppa4* (J. Du et al., 2010; Masaki et al., 2007). Other studies using targeted disruption of *Dppa2* and *Dppa4* did not report such effects (Madan et al., 2009; Nakamura et al., 2011)). However, *Dppa2* and *Dppa4* regulate gene expression as number genes are downregulated upon their knockout (De Iaco et al., 2019; M. Eckersley-Maslin et al., 2019; Madan et al., 2009). In fact, two recent studies demonstrated the role of DPPA2 and DPPA4 in positively regulating the 2C-like state, a subset of ESC that resemble the two-cell embryo both epigenetically and transcriptionally (De Iaco et al., 2019; M. Eckersley-Maslin et al., 2019). Lack of *Dppa2* and *Dppa4* resulted in failure to upregulate *Dux* and *Zscan4* and ZGA related genes, indicating DPPA2 and DPPA4 as master regulators of the ZGA transcriptional program. However, this remains to be demonstrated in vivo. *Dppa2* and *Dppa4* null mice are viable during early mouse development but are born with deficiency in lung alveolar formation and die postnatally (Nakamura et al., 2011). Mutants develop an altered and

repressive epigenetic environment during early development. The repressive epigenetic state persisted at the few tested loci in somatic lung tissues where *Dppa2* and *Dppa4* are not expressed, indicating a role for DPPA2 and DPPA4 in ensuring a permissive epigenetic state in the early embryo (Nakamura et al., 2011). In agreement, DPPA4 is associated with active chromatin (Masaki et al., 2007) potentially functioning with an ESC specific SWI/SNF chromatin remodelling complex (esBAF), that both DPPA2 and DPPA4 interact with (Ho et al., 2009). Furthermore, DPPA2 genome occupancy revealed association with H3K4me3 but low overlap with H3K27ac marks indicting a preference for promoters with low activity and demonstrating a different binding pattern than other pluripotency factors such as OCT4 or ESRRB but more similar to *Kdm5b* and the polycomb repressor proteins (Engelen et al., 2015). This association of DPPA2 and DPPA4 with active chromatin was further explored during iPSC generation. Overexpression of *Dppa2* and *Dppa4* enhances iPSCs generation kinetics and efficiency as they function to ensure permissive chromatin environment during the transition to pluripotency (Hernandez et al., 2018).

1.6.2 Dppa3

Dppa3 (also known as Stella or PGC7) was initially identified as a gene involved in PGC specification by expression patterns in mice embryos (Sato et al., 2002) and is essential for normal pre implantation development (Bortvin et al., 2004; Payer et al., 2003). *Dppa3* expression continues throughout oocyte development and is also activated in early embryo development after ZGA before being repressed upon differentiation in the epiblast by increased DNA methylation (Hayashi et al., 2008). The expression pattern of *Dppa3* is associated with pluripotency and therefor been used as a marker for naïve pluripotency in ESC resembling the inner cell mass of blastocysts. DPPA3 is a small protein with SAP-like domain which can bind both DNA and RNA, additionally it contains nuclear localization and nuclear export signal, implying a wide range of functional potential of DPPA3.

The relationship of DPPA3 with DNA methylation is complicated with various studies showing conflicting results. Initially DPPA3 was shown to protect the maternal pronuclei from TET3 mediated active DNA demethylation using immunostaining methods (Nakamura et al., 2007). Zygotes lacking *Dppa3* displayed 5hmC, the by-product of TET oxidation of 5mC, on both parental genomes compared to just the paternal genome gaining 5hmC and undergoing active removal of 5mC in wild type zygotes. Specifically, DPPA3 was shown to preserve DNA methylation at imprinted control regions and transposable elements. Functionally DPPA3 protects DNA methylation against TET3 mediated removal by binding to chromatin in a H3K9me2 dependent manner, potentially affecting chromatin configuration to inhibit TET3 binding

(Nakamura et al., 2012). Contrasting to its role in protecting DNA methylation at imprinted regions in the early embryo, *Dppa3* KO in mouse PGCs resulted in slower TET mediated DNA demethylation kinetics at retrotransposons (LINE1 and IAP) indicating a role for Dppa3 in driving DNA demethylation during epigenetic reprogramming of PGC (Nakashima et al., 2013). These opposing observations might be explained by reduced amount of H3K9me2 in PGCs compared to the zygote and/or different TETs; TET1 in PGCs compared to TET3 in the zygote (Kagiwada et al., 2013; Nakashima et al., 2013).

Recently, further work on the interplay of DPPA3 and DNA methylation has revealed an unexpected role of DPPA3 in regulating global DNA methylation. Overexpression of *Dppa3* in cultured somatic cells caused globally reduced DNA methylation. DPPA3 binds to UHRF1 causing subcellular relocation of UHRF1, resulting in failure in the DNA maintenance methylation system and global hypomethylation (Funaki et al., 2014). Further work has demonstrated a similar relationship in the germ line and early development. In metaphase II (MII) oocytes, DPPA3 regulates the subcellular distribution of UHRF1 and keeps it out of the nucleus to protect against DNA hypermethylation, with only moderate transcriptional changes (Y. Li et al., 2018). However, zygotes derived from *Dppa3* KO MII oocytes and wild type sperm resulted in hypermethylated genome producing developmental defects in the early embryo, demonstrating the contribution of lack of maternal DPPA3 to embryo lethality. This early developmental failure was in agreement with a previous study on *Dppa3* null embryos (Huang et al., 2017). These embryos confirmed a failure to upregulate key zygotic genes, affecting maternal to zygotic transition and show widespread misregulation of transposon expression, including failure to upregulate MuERV-L elements at the 2-cell stage. The role of DPPA3 in regulating DNA methylation is therefore context dependent. It has a role in protecting the oocyte from aberrant hypermethylation and in protecting against active demethylation of imprinted genes in the early embryo.

A recent pre-print (Mulholland, et al. 2020) has suggested a central role for DPPA3 in promoting the genome wide DNA demethylation seen in mammalian naïve pluripotency using ESC, linking active TET mediated and passive demethylation. Similar to *Dppa2* and *Dppa4*, *Dppa3* is also regulated by DNA methylation and expression of TET genes in pluripotent genes results in hypomethylation at the *Dppa3* promoter and its subsequent upregulation. In concordance to previous studies, KO of *Dppa3* in ESC was shown to cause DNA hypermethlation, by relocating UHRF1, resulting in perturbation of DNA maintenance methylation, phenocopying TET KO ESC. However, it should be noted that previous studies of TET KO have not shown such dramatic hypermethylation and that further studies are needed to rule out cell line

specific occurrence. More interestingly, expression of *Dppa3* in Xenopus, an organism lacking both *Dppa3* and global reduction of DNA methylation, resulted in DNA demethylation following fertilization in in early development. This demonstrates DPPA3 as a key evolutionary functional protein in mammals with role in generating the genome wide DNA demethylated environment associated with naïve pluripotency.

1.7 Genome-wide CRISPR-Cas9 screening

The ability to perform whole genome, loss of function genetic screens is a powerful approach to amplify our understanding of cellular functions by identifying their key genetic components. Historically, such techniques have been done by generating random mutations in a model system and then observing a change in a phenotype followed by the identification of the causing mutation to link a phenotype to a genotype. However, the ability to directly target a gene of interest for perturbation, first with RNA interference methods (Berns et al., 2004; Boutros & Ahringer, 2008; Elbashir et al., 2001) and later with CRISPR-Cas9 technology (Doench et al., 2016b; Shalem et al., 2014), has allowed for powerful loss of function genetic screens being done in mammalian systems.

In short, lentivirus carrying a pooled library of short guide RNAs (sgRNAs) targeting every gene in the genome are used to integrate the sgRNAs into the host's genome. This is done at an appropriate virus to cell ratio that ensures that most cells only receive a single sgRNA, cells without integrated sgRNAs are selected against using antibiotics. After a global genetic perturbation in a pooled CRISPR screen, the affected cells carrying a potentially functional gene knock out need to be separated from the unaffected cells. This is done by connecting the phenotype of interest to a fluorescent reporter or antibiotic resistance expression that can be selected using FACS or antibiotics. The genetic perturbations of the selected population are then identified by amplification of the integrated sgRNAs by next generation sequencing and compared to the distribution of sgRNA in the unselected population to identify depleted or enriched sgRNAs. Finally, a successful screen includes validation of candidate genes and further investigation into their function in the phenotype in interest (Doench, 2018; Joung et al., 2017a).

CRISPR-CAS9 targeting results in a complete knock out of a gene expression compared to a partial knock down by RNA interference (Doench, 2018; Joung et al., 2017a). Thus, CRISPR screens have greater range as they can detect perturbations that would normally be functional at very low expression and not identified by RNA interference methods. Moreover, CRISPR targeting has higher efficiency than RNA interference methods and

therefor CRISPR screening can be used at a lower coverage, therefor fewer cells need to be transduced and maintained to reach the same screening strength. CRISPR technology is versatile and instead of loss of function it can be modified to induce systematic transcriptional activation or repression and epigenetical changes. Using modified CRISPR in a genome wide screening opens the door for different questions to be asked for the role of genes in a specific phenotype (Doench, 2018).

2 Aims

As discussed earlier, DNA methylation undergoes a major resetting in the early embryo with a low point in global DNA methylation converging with pluripotent cells in the inner cell mass (ICM). Epigenome remodeling is central to the establishment of developmental competence, but how this contributes to increased developmental plasticity and competence is interesting. Our understanding of how epigenetic mechanisms are regulated during pre-implantation reprogramming is lacking. More specifically, how DNA methylation is regulated during this period, both globally and focally is not completely understood. To investigate the molecular mechanism behind epigenetic reprogramming and at the same time epigenetic inheritance I will develop a real-time reporter of global DNA methylation to register the methylation status of single mouse embryonic stem cells. By coupling the reporter to genome wide CRISPR knockout screening during demethylation transitioning, I aim to identify genes with role in regulating DNA methylation during epigenetic reprogramming. Finally, candidate genes identified from the screens will be validated and further functionally tested. Overall, this could help further our understanding of the important mechanisms in epigenetic regulation in the early mouse embryo.

To summarise the specific aims are:

1. Establish independent cell lines carrying a real-time reporter of global DNA methylation that responds to DNA methylation transitions with a normalizer optimized for CRISPR-Cas9 screening.

2. Couple the optimized reporter of global DNA methylation with genome wide CRISPR-Cas9 screening in ESC to identify genes with role in regulation of DNA methylation.

3. Validation and analysis of candidate genes identified for their role in escapee protection or reprogramming from the CRISPR screens with further functionally testing.

3 Materials and methods

3.1 Cell culture and differentiation

3.1.1 Mouse embryonic stem cells and HEK293T culture

Mouse embryonic stem cells (mESCs) (mixed 129:B6, XY) were maintained on gelatin (0.1%) coated wells in titrated 2i/LIF (t2i/LIF) culture media (NDIFF 227 media supplemented with PD0325901 (200nM), CHIR99021 (3μM), LIF (1000 U/ml), Penicillin/Streptomycin (1%) and 1% FBS, filtered through 0. 22μM filter) in a humidified CO_2 incubator at 37ºC. ESCs were passaged every two or three days by dissociation with TrpLE. DNA hypomethylation was induced by culturing the ESC in full 2i/L (same as ti2/L except with 1μM PD0325901) for 12 days (naïve mESC). HEK 293T cells were maintained in DMEM with 10% FBS and Penicillin/Streptomycin (1%) and passaged every 3 to 4 days.

3.1.2 Epiblast-like cell (EpiLC) differentiation

EpiLC differentiation was induced by seeding $3.5x10^4$ naïve ESCs per cm^2 on fibronectin (Millipore, #FC010) coated wells in EpiLC media (NDIFF 227 media supplemented with 1% KnockOut Serum Replacement, ActivinA (20 ng/ml), bFGF (12.5 ng/ml) and Penecillin/Streptamycin (1%)). EpiLC media replaced after 24 hours and cells collected after 44 hours.

3.1.3 Endoderm differentiation

To induce endodermal differentiation 2500 to 20000 ESC (depending on time point collection) were seeded per cm2 on gelatin (0.1%) coated wells in 2i/Lif+1% FBS media. After overnight culture the 2i/LIF ESC media was removed followed by 3 washes with PBS and replaced with endodermal media (RPMI media supplemented with L-Glutamine (2mM), FBS (0.2%), Inducer of definitive endoderm 1 (IDE1)(5uM)(Borowiak et al., 2009) penicillin/streptomycin (1%). Endodermal media was changed every two days or when needed with samples collected after 0 days (ESC), 3 days (Endodermal day 3), 6 days (Endodermal day 6) and 12 days (Endodermal day 12). Genomic DNA and total RNA were collected from each time point.

3.2 Molecular cloning and generation of the eRGM plasmids

3.2.1 Generation of the DNA methylation sensitive reporter

H2B-GFP-SV40pA fragment from the H2B-GFP plasmid (addgene #11680) was amplified using the primers H2B_GFP_Fw and H2B_GFP_Rev (Table 1). Minimal promoter of the mouse gene *Kcnq1ot1* (mm10, Chr7:143,296,371-143,296,745) was amplified using the primers Kcnq1ot1_Fw and Kcnq1ot1_Rev. The H2B-GFP-SV40pA and the *Kcnq1ot1* promoter fragments were cloned into an empty ampicillin resistant piggyBac backbone digested with EcoRI and BamHI using the InFusion assembly system (pPB-*Kcnq1ot1*-H2B-GFP). A genomic DNA methylation sensor region derived from the mouse *Dazl* promoter region (mm10, chr17:50,293,285-50,294,435) was amplified using the primers Dazlas_Fw and Dazlas_Rev and inserted in antisense upstream of the *Kcnq1ot1* promoter on EcoRV digested pPB-*Kcnq1ot1*-H2B-GFP construct using InFusion assembly, generating pPB-Dazlas-*Kcnq1ot1*-H2B::GFP. Similarly, the other tested sensor regions and the IAP were inserted upstream of *Kcnq1ot1*-H2B-GFP with the following mm10 coordinates; Dst6 exon (chr1:34,200,844-34,201,413), Asz1 Promoter (chr6:18,108,776-18,109,711), Chr18 Intergenic (chr18:83,180,129-83,181,250) and IAP (chr8:47,453,950-47,454,931). Confirmation of the correct assembly was done with sanger sequencing.

3.2.2 Generation of the DNA methylation insensitive reporter

The H2B region of the H2B-GFP plasmid (addgene #11680) was amplified using the primers H2B_GFP_Fw2 and H2B_GFP_Rev2 (Table 1). The EF1a promoter region of the lenti MS2-P65-HSF1_Hygro plasmid (addgene, #61426) was amplified using the primers EF1a_Fw and EF1a_Rev. The mCherry-pA region of the pMK293 plasmid (addgene #72831) was amplified using the primers mCherry_Fw and mCherry_Rev (Table 1). The H2B fragment, the *EF1a* promoter fragment and the mCherry2-pA fragments were cloned into an empty ampicillin resistant piggyBac backbone digested with EcoRI and BamHI using InFusion assembly generating pPB-*EF1a*-H2B::mCherry. Confirmation of the correct assembly was done with sanger sequencing.

Table 1. InFusion primers used to generate eRGM plasmids.

Primer	Sequence (5'-3')
H2B_GFP_Fw	AAGTAACAAAGGATCTAAGATACATTGATGAGTTTGGACAAACCACAAC
H2B_GFP_Rev	TTTCTCTCCACCATGCCAGAGCCAG
Kcnq1ot1_Fw	TCGTCTTCAAGAATTACAAGCTCACCCAATCCAAATGC
Kcnq1ot1_Rev	CATGGTGGAGAGAAAAGCACACTAAGGTGTACTAGAC
Dazlas_Fw	AAGACGATCGACGATGCTAGCTTATGCCCTCTCCCC
Dazlas_Rev	TTTCTTGTTATAGATCTCGAGCCAAGCACCCT
EF1a_Fw	TCGTCTTCAAGAATTGGCTCCGGTGCCCG
EF1a_Rev	ATGGTGGCTCACGACACCTGAAATGGAAGAA
H2B_GFP_Fw2	GTCGTGAGCCACCATGCCAGAGCCA
H2B_GFP_Rev2	TTCTCCTTGGTGGCGACCGGTGGA
mCherry_Fw	CGCCACCAAGGAGAAGAGTGCTTGTCC
mCherry_Rev	AAGTAACAAAGGATCTAAGATACATTGATGAGTTTGGACAAACCACAAC
IAP_Fw	TTTCTTGTTATAGATGCTAGCTCGTTAGAAAGTTCAAGGC
IAP_Rev	AAGACGATCGACGATCTCGAGGCATAAACACACGTGC

3.3 Generation of eRGM cell lines

mESC lines containing floxed *Dnmt1* alleles (Jackson-Grusby et al., 2001) were transfected with the following plasmids: pPB-as*Dazl-Kcnq1ot1*-H2B::GFP (*in silico* DNA methylated using M.SssI (NEB, #M0226S)), pPB-*EF1a*-H2B::mCherry, pPB-*spCas9*-Hygro (Hackett et al., 2018) and PBase using Lipofectamin 3000 (Life Technologies, #L3000015) following the manufactures recommendations. The transfected cells were selected for *spCas9* integration and co-integration of GFP and mCherry in titrated 2i/Lif using Hygromycin (250 µg/ml) for 5 days. Hygromycin resistant cells were seeded for single clone picking. After clonal isolation and expansion, the lines were tested by comparing the expression of GFP and mCherry using flow analysis after culturing the cells for 7 days in t2i/Lif or 2i/Lif. The clonal lines

demonstrating the best dynamic range of GFP expression and minimal changes in mCherry expression up on demethylation with 2i/Lif were picked, generating eRGM line 1 and eRGM line 2 (Figure 4E). Additionally, the eRGM response was confirmed by inducing conditional Dnmt1 KO using of 4-Hydroxytamoxifen (800nM) (Sigma-Aldrich #H7904).

3.4 Gene editing using CRISPR-Cas9

3.4.1 Generation of clonal knock-out lines

To generate clonal knockout (KO) lines the CRISPR/Cas9 system was utilized to generate frame-shifting mutations. ESC carrying eRGM (including *spCas9*) and cultured in t2i/LIF were transiently transfected with a sgRNAs cassette targeting coding exon of the genes of interest (*Dppa2, Dppa4, Dusp6, Cop1*) (Table 2) for gRNA primers). 12 hours after transfection, successfully transfected cells were enriched by puromycin (1.2µg/ml) selection for 60 hours. Puromycin resistant cells were seeded at low density, 1000 cells per 9.6cm^2 in t2/LIF without puromycin for clonal isolation. Following clonal expansion knockout lines were initially probed using functional testing by investigating the effect of the KO on eRGM dynamics using flow cytometry (section Methods 3.5). Finally, the homozygous knockout was confirmed using immunoblotting (see section Methods 3.9 , done for *Dppa2* and *Dppa4* KO lines) or Sanger sequencing using the Tracking of Indels by DEcomposition (TIDE) tool (E. K. Brinkman et al., 2014) (done for *Dusp6* and *Cop1* KO lines).

Table 2. sgRNA used to generate clonal KO lines.

Gene target	gRNA sequence (5'-3')
Dppa2	GATAGATACCTGGTGGTGTG
Dppa4	CTGCAAAGGCTAAAGCAACG
Cop1	AAGCTCCTTCTCCATCACAC
Dusp6	CTCTTCCAACACGTCCAAGT
Non-targeting Control	GCTTTCACGGAGGTTCGACG

3.4.2 Generation of population scale knock out lines

Population scale knock of multiple candidate factors from both screens (n=24 and n=26), selected for validation of the screening results, were generated by transfecting both eRGM line1 and line2 with piggyBac cassettes carrying gRNAs targeting the genes of interest (Table 3 and Table 4 for list of gRNA used to generate population KO lines) and PBase using lipofectamine 3000,

following the manufactures recommendations. To generate stably integrated cell lines, the cells were selected for successful integration of the gRNAs for 7 days with puromycin (1,2g/ml) to generate bulk KO population before assaying eRGM dynamics in t2i/LIF and 2i/LIF.

Table 3. sgRNAs used to generate population KO lines for DNA methylation maintenance candiates.

Gene	**gRNA sequence (5'-3')**
Arid1a	CCTGCTGGCCATACGCACTG
Atrx	ACATACCTAACTGCTCTCTG
Bap1	ACCTGTCTGAGTGCACTCAG
Caprin1	GCGTAATCGACAAGAAACTT
Cpsf6	ACGTTGTGTACACTTACACT
Dis3	GATATTACCACTAGTCTCCG
Dnajc2	GAAACTTCGGAACTCGTGCA
Dnmt1	ACCTCGGGCCAATCAATCAG
Dohh	AAGCCCCCGAAGTGTAAACA
Elp3	TATTCAACCCAGTCTTACAC
Elp6	CTTCAGTTCGCAGCACACCG
Hnrnpc	GGATGTACAGTTACCCAGCG
Ipo11	ATGGGATTGATCGTTACTGG
L3mbtl2	ATGAAGTACCCTTTCCGACA
Max	GACTCAGTCCCATCACTCCA
Meaf6	CGCCGGGTGTCTGGGATCTG
Mga	ACTGGAATCAACAACAATCG
Morc2a	TAGGCTACTCCAGCCAACTG
Nudt21	CCGAACTGGTTGACCCCCCG
Shoc2	ATTATGTAACCTCATTACCC
Smarce1	ATGAGGTACAGCAGAAAGGT
Socs3	CGGATAAGAAAGGTGCCCGC
Tceb2	CGAACTGAAGCGCATCGTCG
Tgif1	GTTCACGATTTCCCGCCGTG
Thoc3	GTTGATGCAACCATTGCCGT
Trmt6	GCCATGATGGAACGAATGGG

Vhl	TTAACCTGGCAATGTGATGT
Non-targeting Control	GCTTTCACGGAGGTTCGACG

Table 4. sgRNAs used to generate population KO lines for DNA methylation reprogramming candidates.

Gene	**gRNA sequence (5'-3')**
Brd4	AGCCATGGCTCGAAAACTCC
Cbfa2t2	AACACCGAGAAGTTCGTGAG
Cnot10	TGCCAAGTCCAAGATACACC
Cop1	AAGCTCCTTCTCCATCACAC
Dcaf15	GTGGCCTAGCGACGCTTCCA
Dda1	AAAAGACGAACATACTTCTG
Dppa2	GATAGATACCTGGTGGTGTG
Dppa4	CTGCAAAGGCTAAAGCAACG
Dusp6	CTCTTCCAACACGTCCAAGT
Hdac1	TAAAGGGAGTTCTCACCCGT
Hnrpa1	GGGAACACTAACAGACTGTG
Hsp90aa1	GCTTCAGCTTGGAATTCACG
Ints10	TTCTACATCACCATACACAT
Jak1	AAACATATAGTGTACCTCTA
Kdm3a	AGGGGAACTTTGCTGGACCA
Klf2	GCCTTCACTAGCCGCCCGGG
Lifr	TGTTTCGTGTTCATTGAACG
Med24	TCGACTCCGAGAGGATCACC
Nufip2	AACTGGGGGTGGCTTAAACA
Prdm14	TAGTAGGGAAGCAATTACCG
Rqcd	GCCTACTGCACTAGCCCAAG
Virma	GATATGAAGTGGTACCTGCG
Vwa9	ATCCCTAAAGTCATTAACAC
Ythdf2	GGGATTGACTTCTCAGCATG
Non-targeting Control	GCTTTCACGGAGGTTCGACG

3.5 Flow analysis and cell sorting

Cell were collected by dissociation using TrpLE and resuspended in PBS + 1% FBS (FACS media) and filtered (0.2um) before being run on Attune Nxt (ThermoFisher) for analysis. Fluorescent activated cell sorting (FACS) was done using a FACS Aria III (Becton Dickinson) and FACS Diva software. Data were analysed using FlowJO v10.5.3 (Tree Star, Inc.).

3.6 Genome-wide CRISPR-Cas9 screen in eRGM

To generate lentiviral particles carrying the Brie gRNA library (Doench et al., 2016a), Lenti-X HEK 293T cells were transfected with *pPax2* plasmid, *pMD2.G* plasmid and the Brie gRNA plasmid library using lipfectamine 3000 in a BSL2 culture facility. The supernatant, enriched in lentivirus, was harvested 48- and 72-hours post transfection and filtered through 0.22µm low protein-binding filter unit. The lentiviral particles were concentrated using Lenti-X concentrator and resuspended in NDIFF 227. Finally, the lentiviral transduction efficiency was established by transducing mESCs across a titration curve followed by measuring cell survival after puromycin selection (1.2µg/ml) for 2 days compared to ESC culture in culture without puromycin. Cell lines carrying gRNA knockout library were generated by transducing $7x10^7$ *eRGM line 1* and *line 2* cells cultured in t2i/LIF with the previously generated lentiviral particles carrying the genome-wide CRISPR knock out gRNA (n=78,637) library. Optimized number of lentivirus was used to ensure ~45% efficiency and therefore >400 fold gRNA coverage in the transduced population. The transduced cell lines were maintained in t2i/LIF + puromycin (1.2µg/ml), while keeping the total cells over $3.2x10^7$ at every passage (>400 fold coverage maintained). After 10 days of t2i/LIF culture the GFP positive cells (the highest 1% of GFP expression) that also that also maintained normal mCherry expression, indicative of loss of epigenetic silencing, were purified by flow cytometry (114455 and 78508 for eRGM #1 and eRGM #2, respectively). Another portion of transduced cells (likewise maintained $>3.2x10^7$) in t2i/LIF were demethylated by transitioning the cell lines into 2i/LIF. At 12 days GFP-negative cells (defined as the lowest 1% of GFP-expression), indicative of incomplete epigenetic resetting, were purified by flow cytometry (291,248 and 237,121 for eRGM #1 and eRGM #2, respectively). Moreover, total unsorted cells ($>3x10^7$) from both t2i/LIF and 2i/LIF cultures at the time of cell sorting were collected from both eRGM line 1 and line 2 as controls. Genomic DNA was collected from all purified and unselected (control) cell populations using either a Quick-DNA Microprep Plus Kit (Zymo Research, #D3020) or a DNeasy blood and tissue kit (Qiagen, # 69504). Integrated gRNAs were amplified using primers with the P7 flow cell overhang: 5'-

CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT-3' (8bp Barcode) and P5 overhang: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTTTGTGGAAAGGACGAAACACCG-3' using Q5 Hot Start High-Fidelity 2X Master Mix (NEB, #M0494S) for 21 to 24 cycles in multiple 50µl reactions (up to 4µg of gDNA per 50µl). PCR fragments purified with SPRI beads (Beckman Coulter, #B23318) and quantified with Qubit III were pooled and sequenced with a Nextseq500 Illumina system (single-end 75nt). For counting and analyzing the sgRNA hits the Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout (MAGeCK, v0.5.9) (W. Li et al., 2014) tool was used. First the reads were trimmed using cutadapt (v1.15) (*cutadapt -g TTGTGGAAAGGACGAAACACCG*) and quality checked using FastQC. Next the gRNAs were counted and normalized to total reads within the sample (MAGeCK -count -norm-method total), and finally the sorted and unsorted control samples were compared using the -test command in MAGeCK, which identifies significantly enriched/depleted gRNAs between samples. Final candidate lists were generated by using FDR threshold of <0.05 and a fold-change (FC) gRNA frequency threshold of >3 (to select larger effect-size candidates), and by intersecting the candidates from the two independent cell lines (eRGM line1 and line2).

3.7 Luminometric DNA Methylation Assay

The LUminometric DNA Methylation Assay (LUMA) was used to measure global DNA (CpG) methylation levels. For each sample, 200-500ng of genomic DNA were split evenly into two 4 hour digestions at 37C. Digestion A: gDNA digested with HpaII and EcoRI and Digestion B: digested with MspI and EcoRI, both done in 15µl reactions using 10x CutSmart digestion buffer (NEB). Using the PyroMark Q24 advanced system, an equal amount of annealing buffer (15µl) was added to the digestion and the samples were loaded into a PyroMark Q24 pyrosequencer platform using the dispensation order GTGTGTCACACAGTGTGT to quantitate the overhangs generated in each digestion. The % CpG methylation was calculated by comparing the EcoRI normalized HpaII ratio to the normalized MspI ratio using the formulas:

$$HpaII\ ratio: \frac{Dispensation\ 7+13}{\left(\frac{Dispensation\ 8+14}{2}\right)} \quad MspI\ ratio: \frac{Dispensation\ 7+13}{\left(\frac{Dispensation\ 8+14}{2}\right)}$$

$$\%\ DNA\ methylation: 100\%x\left(1-\frac{HpaII}{MspI}\right)$$

3.8 CUT&RUNseq

The recently developed Cut and Run protocol (Skene et al., 2018) was used here to study protein-DNA binding and histone modification locations. mESC were dissociated using TrypLE, counted and 250.000 cells collected by centrifugation at 600xg for 3 minutes at room temperature. The supernatant was removed, and the cells were washed twice with 1ml of WASH Buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5mM Spermidine and Protease Inhibitor tablet (Roche)) followed by resuspension 1ml of wash buffer. 10µl of Concanavalin A-Coated beads (Bangs Laboratories, #BP531), washed and resuspended in Binding buffer (20mM Hepes-KOH, pH 7.9, 10mM KCl, 1mM $CaCl_2$, 1mM $MnCl_2$), were added to the cells and the samples rotated for 10 minutes at RT. The bead-bound cells were placed on a magnet stand, the supernatant removed and 300µl of antibody buffer (Wash buffer plus 0.02% digitonin and 2mM EDTA) containing 0.5µg of antibody (see Table 5. for antibodies used) was added to the beads and incubated with rotation at 4°C overnight. Supernatant was discarded and samples were washed with 1mL of cold Dig-Wash buffer (Wash buffer plus 0.02% digitonin) using a magnetic stand and the samples resuspended in 300µl of cold Dig-Wash buffer. Protein-A::MNase (pA-MNase) was added to a final concentration of 700ng/ml and the samples were rotated for 1 hour at 4C. After discarding the supernatant, the samples were washed twice in 1ml cold Dig-Wash buffer before resuspending the samples in 50µl of Dig-Wash buffer by flicking the 1.5ml Eppendorf tubes. The samples were placed in iced water to cool them to 0C. Digestion was initiated by adding 2µL 100mM $CaCl_2$ to the samples, followed by gently flicking the tubes to mix before returning the samples to iced water. After 30 minutes (histone modifications CnR) or 5 minutes (DPPA2 CnR) pA-MNase digestion at 0°C, 50µl of 2XSTOP buffer (340 mM NaCl, 20mM EDTA, 4mM EGTA, 0.02% Digitonin, 250µg Rnase A, 250µg Glycogen) was added and the samples mixed. Samples were incubated at 37°C for 10 minutes to release CUT&RUN fragments from the insoluble nuclear chromatin followed by 16000xg centrifugation for 5 minutes at 4°C. The supernatants were transferred to a new tube and the beads discarded. After addition of 2µl 10% SDS and 2.5µl 20mg/ml Proteinase K (Thermo Fisher, #AM2546) the samples were incubated for 10 minutes at 70°C. DNA was purified and size selected using SPRIselect beads (Beckman Coulter, #B23318) following the recommended protocol for double selection; 0.5x volume of SPRIselect beads to sample ratio (50µl of beads to 100µl sample) followed by 1.3x ratio (130µl of SPRI beads). DNA was eluted from the SPRIselect beads with 30µl of 0.1xTE. Purified CUT&RUN DNA was quantified with Qbuit III and up to 10ng of CUT&RUN DNA fragments were used to make libraries using the NEBNext Ultra II DNA Library Prep Kit for Illumina (#E7645S, NEB). The libraries were

made using the following PCR program: 98°C 30s, 98°C 10s, 65°C 10s and 65°C 5min, steps 2 and 3 repeated 12-14 times depending on input DNA. Library samples were sequenced on the Nextseq Illumina sequencing system (paired-end 40 sequencing).

Table 5. Antibodies used in the Cut and Run protocol

Antibody	AntibodySupplier	Catalog Number
Anti-DPPA2	Sigma	AB4356
Anti-H3K4me3	Diagenode	C15410003
Anti-H3K27me3	Sigma	07-449
Anti-H3K9me3	Merck	07-442
Anti-Rabbit IgG	Cell Signaling	7074

3.9 Immunoblotting

ESC at sub-confluence were subjected to protein extraction using RIPA buffer (Sigma, #R0278) with protease inhibitors (Roche, #4693159001) at 4°C for 30 minutes. Cellular lysis was centrifuged at 12,000xg and the supernatants collected and the protein concentration measured using the Pierce Rapid Gold BCA Protein Assay Kit (Thermo Fisher, #A53226) following the manufactures instructions. 10µl of 4x Bolt LDS sample buffer (Thermo Fisher, #B0007) and 4 µl of 10x bolt reducing agent (Thermo Fisher, #B0004) were added to 20µg of protein samples and filled up to final volume of 40µl with H2O. The samples were incubated at 70°C for 10 minutes before being loaded onto 4-12% Bis-Tris gel (Thermo Fisher, NW04125BOX), the proteins were separated with 150V electrophoresis for 30 minutes and blotted on a PVDF membrane using the iBlot dry 2 blotting system. The membrane was blocked in 5% milk/PBS for 1 hour at RT followed by primary antibody staining using 1:500 to 1:1000 diluted antibody (see Table 6. for antibodies used) with 5% milk at 4°C overnight. After washing with PBS 0.1% Tween the membrane was stained for 1 hour at RT with a HRP-linked secondary antibody diluted 1:10000 in 5% milk. The membrane was washed for 5 minutes thrice with PBS 0.1% Tween and Pierce western blot plus solution (Thermo Scientific, #32132) was added on the membrane for 5 minutes before imaging using the ChemiDoc XRS+ system (Bio-Rad) or ImageQuant 800 (AMERSHAM).

Table 6. Antibodies used for immunoblotting

Antibody	Antibody Supplier	Catalog Number
Anti-DPPA2	Sigma	AB4356
Anti-DPPA4	R&D Systems	AF3730
Anti-TUBULIN	Sigma	T4026
Anti-Mouse IgG	Abcam	Ab6709
Anti-Goat IgG	Fisher Scientific	PI31402

3.10 RNA-seq

ESCs were dissociated using TrpLE and collected with a 3 minutes 600xg centrifugation. Total RNA was collected using the Qiagen RNeasy kit. Total RNA was measured using Quibit III and quality checked with Bioanalyser 2100 (Agilent) and samples with RIN>8.5 were subjected to mRNA enrichment using the NEBNext Poly(A) mRNA magnetic isolation module and prepped into stranded (directional) libraries using the NEBnext Ultra II directional RNA library prep kit following all manufactures guidelines (performed at EMBL Genomics Core Facility, GeneCore). Amplified libraries were multiplexed and sequenced on Illumina Nextseq for paired-end 40 (PE40) or single-end 75 (SE75).

3.11 EM-seq (Bisulfite sequencing)

ESCs were collected with a 3 minute 600xg centrifugation and genomic DNA isolated using Zymo Microprep DNA kit (Zymo Research, #D4074). DNA methylation libraries were generated using the NEBnext Enzymatic Methyl-seq (EM-seq) kit following the manufacturer's instructions (performed at EMBL Genomics Core Facility, GeneCore). In short, up to 200ng of genomic DNA was mechanically fragmented to an average size of 240-290bp, then end-repaired and ligated to EM-seq adapters. Genomic 5mC and 5hmC were oxidised using TET2 to 5caC to protect methylated sites against deamination. Subsequently, APOBEC was used to deaminate unmethylated cytosines to uracils, whilst oxidized forms of the originally methylated cytosines are protected. This enzymatic DNA conversion generates a conversion system identical to bisulfite conversion but with higher yields and lower duplication and therefore better detection of CpGs with fewer sequencing reads. The libraries

were amplified using Q5 polymerase, multiplexed and sequenced on an Illumina Nextseq for single-end 75 (SE75).

3.12 Real-time qPCR

Total RNA was collected with RNeasy (Qiagen) used to synthesize cDNA with mixture of random hexamers and reverse transcriptase after an erasure of genomic DNA (TAKARA PrimeScript™ RT Reagent Kit with gDNA Eraser). Briefly, up to 1µg of total RNA, measured with NanoDrop 2000, was subjected to DNAase treatment: 2µl 5x gRNA eraser Buffer, 1µl gDNA eraser, 1µg total RNA and RNase Free dH2O to 10µl for 2 minutes at 42°C followed by reverse transcription reaction: 10µl of the DNAase treated RNA reaction, 4µl 5X PrimeScript Buffer 2, 4µl RNase Free H2O, 1µl PrimeScript RT Enzyme Mix I and 1µl RT primer Mix, kept at 37°C for 15min and then at 85°C for 5 seconds. Diluted cDNA was used (2.5ng RNA equivalent per reaction) in triplicate quantitative PCR reactions with primers in Table 7. and qPCRbio SYgreen Blue Mix on a QuantStudio 5 (Applied Biosystems) thermal cycler. Results were analysed using the $2^{-\Delta\Delta Ct}$ (relative quantitation) method with normalisation to the housekeeping gene *Rplp0*. Statistical significance was determined using the students t-test with Holm-Sidak correction in Graphpad Prism8 with alpha = 0.05 to identify differences in gene expression from two biological replicates.

Table 7. Primers used in RT-qPCR

Primer Name	Primer sequence (5'-3')
RplP0_F	TCCAGAGGCACCATTGAAATT
RplP0_R	TCGCTGGCTCCCACCTT
Nanog_Fw	AAGGATGAAGTGCAAGCGGT
Nanog_Rev	ATACTCCACTGGTGCTGAGC
Prdm14_Fw	CGCCTCCGGATCCATATTCT
Prdm14_Rev	TGTGCTTGTTCAGGCTGGAA
Dnmt3b_Fw	TCGAGAATGTTGTGGCCATGA
Dnmt3b_Rev	TGGCATCGATCATCACTGGG
Fgf5_Fw	TGTGTCTCAGGGGATTGTAGG
Fgf5_Rev	TTACAGTCATCCGTAAATTTGGC
Emb_Fw	TGAGGGCGATCCCACAGAT
Emb_Rev	CCGTCACTGAGATATTACAGCTC

FoxA1_Fw	ATGAGAGCAACGACTGGAACA
FoxA1_Rev	TCATGGAGTTCATAGAGCCCA
Hand1_Fw	CAAAAAGACGGATGGTGGTCGC
Hand1_Rev	TGCGCCCTTTAATCCTCTTCTCG
Col16a1_Fw	GCCAGGTTTTGACAACTGTGCC
Col16a1_Rev	AACAGCCAGGTTCTCCCTCACT
Nkx2-5_Fw	CATTTTACCCGGGAGCCTACGGTGA
Nkx2-5_Rev	CTTTGTCCAGCTCCACTGCCTTCTG
Gnmt_Fw	GTTTTGCTCACTTGCCAGACT
Gnmt_Rev	GTAGTTGCGGTGGTCGATCA
Tekt2_Fw	GAGAGGATCGACACTGTCAACC
Tekt2_Rev	GTTTTGCTCCGCCGATTCTTT
Cdcp1_Fw	AGAGAAGTGGCGTCACAGTG
Cdcp1_Rev	CCAGGTCTGATAATCAACTCGGT
L1-T_Fw	CAGCGGTCGCCATCTTG
L1-T_Rev	CACCCTCTCACCTGTTCAGACTAA
L1-A_Fw	GGATTCCACACGTGATCCTAA
L1-A_Rev	TCCTCTATGAGCAGACCTGGA
IAPEz_Fw	CGGGTCGCGGTAATAAAGGT
IAPEz_Rev	ACTCTCGTTCCCCAGCTGAA
Dppa2_Fw	GGGAGTGGTGTCGGTATCAT
Dppa2_Rev	GAGTGTCTCCGAAGTCTCAAA
Dppa4_Fw	GTCTAGTCAACCAAGCACGG
Dppa4_Rev	GGTTTTCCCTCCTTTGGTTTACA

3.13 Pyrosequencing

Genomic DNA (50-300ng) was purified from mESC or EpiLC using DNeasy blood and tissue kit (Qiagen, #69504) and sodium bisulfite converted using EZ DNA Methylation-Gold Kit (#D5005 Zymo Research), eluting with 11µl of H_2O. 1µl of Bisulfite converted DNA used as template in 25µl PCR reactions with primers in Table 8 and polymerase from PyroMark PCR kit (Qiagen, #978703) following the recommended PCR conditions (annealing temperature 56°C). The presence of a single dsDNA fragment was confirmed for all samples after PCR by running 5µl of the reaction on a 2% agarose gel. 10µl of PCR reactions

were subsequently used for pyrosequencing with the PyroMark Q24 Advanced Reagents (Qiagen, #970902) and the sequencing primers from Table 8 and ran on PyroMark Q24 Advanced (Qiagen) platform with target specific dispensation orders. The ratio of sequenced unconverted cytosine to converted thymine at each CpG site was analysed using the PyroMark Q24 Advanced program (v3.0.0). Statistical significance between samples was determined using students t-test with Holm-Sidak correction in Graphpad Prism 8 with alpha = 0.05 to identify differences in DNA methylation of the average methylation of all tested CpG sites from two biological replicates.

Table 8. Primers used for Pyrosequencing

Primer Name	Primer sequence (5'-3')
Col16a1_PyroSeq_Fwb	GTGTTTTATTAGTTTTTTGGTAGGTT[Btn]
Col16a1_PyroSeq_Rev	CAACCCTAAAACTACCAACTATAA
Col16a1_Seq	ATCTACCCTAACCCCAACTT
Nkx2-5_PyroSeq_Fw	GTTGTTTATTTTTGGGGATAGTTT
Nkx2-5_PyroSeq_Revb	CCTCTTCTACCCCTAATACTC[Btn]
Nkx2-5_Seq	AGTTTTTTAGTATGGTGTT
Hand1_PyroSeq_Fwb	GAGGTAGGGTTAAATAGGAAGGG[Btn]
Hand1_PyroSeq_Rev	TAACAACACAAACCAAAAACCTCTC
Hand1_Seq	AAAAACCTCTCCAACATA
L1-T_PyroSeq_Fw	GGTTGGGGAGGAGGTTTAAGTTATA
L1-T_PyroSeq_Rev	CTACCTATTCCAAAAACTATCAAATTCTT[Btn]
L1-T_Seq	GGGAGGAGGTTTAAGTTATAGTA

3.14 Bioinformatics analysis

3.14.1 RNAseq analysis

First, raw RNAseq reads were quality trimmed using TrimGalore (0.4.3.1, -phred33 --quality 20 --stringency 1 -e 0.1 --length 20) before being mapped to the mouse mm10 genome assembly using RNA Star (2.5.2b-0, default parameters except for --outFilterMultimapNmax 1000). Mapped reads were imported into seqmonk (Primary alignment only, MAPQ > 20). The data was quantified using the RNA-seq quantification pipeline in the seqmonk software; directional library on antisense strand, normalized to reads per million (RPM) or gene length-adjusted to reads per kilobase million (RPKM). The samples

were normalized using the match distribution quantitation in seqmonk when appropriate. Differentially expressed genes were determined using DEseq2 (v.1.24.0) using raw mapping counts and multiple-testing adjusted P<0.5 along with fold-change difference >2 as significance threshold.

3.14.2 Analysis of TE expression

Using seqmonk, mapped RNAseq reads were not filtered using MAPQ to enable multimapping reads but taking the primary alignment only. Mapped reads that overlapped genes (+/- 2kb) were filtered to exclude signal derived from genic and/or TE transcription, which could confound results. Repeatmasker was used to identify repeat locations for the mm10 genome and quantitate the filtered reads overlapping either full-length (>5 kb for LINE, >3 kb for LTR) or truncated (<5 kb for LINE, <3 kb for LTR) annotated TEs. All reads overlapping each class of repeat element were summed up and then corrected for total genome length of the same TE class and the sequencing depth of the library. This generate a log2 RPM expression value for each TE class as previously described (Berrens et al., 2017).

3.14.3 EMseq analysis

Raw fastQ sequences were quality trimmed using Trim Galore (0.4.3.1) and reads aligned to mm10 using Bismark (0.20.0, enabled discarding of the first 8 bp from the 5' end and the last 2 bp from the 3' of a single-end read). Bismark methylation extractor tool was used to generate genome-wide methylation calls from mapped reads. These calls were analysed using the Seqmonk software (1.44.0). The genome was binned into tiles containing 50 consecutive CpGs and their methylation status counted using the DNA methylation pipline in seqmonk. Differentially methylated regions were identified as tiles with that satisfied both statistical filters: read depth corrected logistic regression p<0.05 and binomial test p<0.01 with minimal reads of 10 per probe and minimal 10% methylation difference. Using bins over all refseq gene transcriptional start sites (+/- 1kb) and at the 5' end of LINE1 elements > 5kb (+/- 500bp) (using repeatmasker annotations) differentially methylated promoters (DMPs) and differentially methylated LINE1s (DMLs) were identified (Logistic regression p<0.05 and binomial p<0.01 with minimal reads of 10 per probe and minimal 10% methylation difference). To generate final DMR, DMP and DML datasets, significant hits from ESC and EpiLC *Dppa2*$^{-/-}$ were collated.

3.14.4 CUT&RUNseq analysis

Raw Fastq sequences were quality- and adapter- trimmed with TrimGalore (v0.4.3.1, -phred33 --quality 20 --stringency 1 -e 0.1 --length 20) and aligned to the mouse mm10 genome using Bowtie2 (v2.3.4.2, -I 50 -X 800 --fr -N 0 -L 22 -i 'S,1,1.15' --n-ceil 'L,0,0.15' --dpad 15 --gbar 4 --end-to-end --score-min 'L,-0.6,-0.6'). Mapped sequences were analysed using seqmonk and reads

with a MAPQ score <20 were discarded. DPPA2 binding peaks were identified using MACS2 ($p<1^{e-5}$, 200nt fragments) with IgG CUT&RUN as input control. Significant genomic feature overlap of DPPA2 binding sites were determined from one sample t-test compared to three random genomic probe sets of similar size distribution as the DPPA2 binding sites. Histone modification CUT&RUN was analysed by quantifying the normalized reads over genomic features (DPPA2 binding sites, and the previously identified DMPs, DML, DMRs (see section Methods 3.14.3)) to generate density and trend plots using seqmonk. For analysis of repetitive elements (e.g. LINE) multi-mapping (MAPQ<20), was allowed taking only the primary alignment.

3.14.5 Gene ontology analysis

Gene Ontology analysis were performed using DAVID (v.6.8) for differentially expressed genes (WT versus *Dppa2* knockout and *Dppa4* knockout in ESCs and EpiLCs), genes associated with DMPs and for screen candidates. FDR values for selected Gene Ontology terms from BP_all are displayed.

3.14.6 Data Availability

All data from the RNA-seq, Cut&Run-Seq, EM-seq and CRISPR screening experiments are available through the Gene Expression Omnibus database under the accession code GSE146863.

4 Results

4.1 Setting up and validating an in vitro model for DNA methylation reprogramming

To unravel the mechanisms at play in either erasure or persistence of epigenetic states during developmental reprogramming, a system to recreate epigenetic reprogramming in vitro was needed (Figure 2). Traditionally mouse embryonic stem cells (mESC) have been cultured in high serum (~10%) conditions (Serum/LIF) which induces DNA hypermethylation (70%, Figure 3A) and transcriptionally heterogeneous environment. Maintaining mESC in serum low (1%) conditions using two cell signaling inhibitors (2i/LIF, see methods) can be used to recreate naïve pluripotency conditions and DNA hypomethylation (35%, Figure 1). Although a transition from Serum/LIF to 2i/LIF produces global DNA methylation changes, it is accompanied by a major transcriptional shift (Figure 3B-E). These changes in cellular state make it a less than an ideal system to study the regulation of developmentally induced DNA demethylation because of confounding effects of the cellular state on the global DNA methylation state. However, culturing ESC in titrated 2i/LIF (t2i/LIF) by reducing the amount of PD inhibitor used (from 5uM to 1uM, see methods) promotes high global levels of DNA methylation. Unlike Serum/LIF ESC, t2i/LIF ESC are highly transcriptionally comparable to full 2i/LIF ESC (Figure 3B-E). I therefore tested the transition of t2i/LIF cultured mESC to full 2i/ LIF media and marked induced global DNA demethylation over 12 days in 2i/ LIF culture. Thus, the global DNA demethylation following the t2i/LIF to 2i/LIF transition occurs with only minimal changes in global gene expression and functional ‘cell state’ thereby preventing indirect effects in the DNA demethylation model.

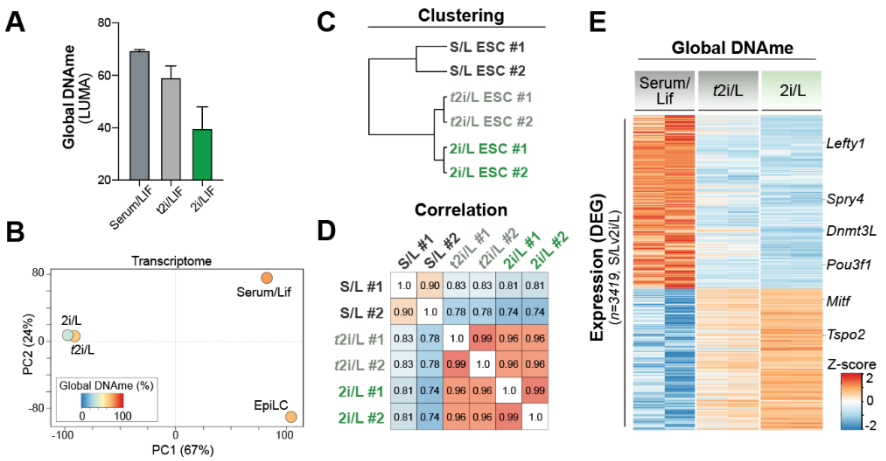


Figure 3. Transcriptional changes during demethylation.

A) Global DNA methylation levels of mESC cultured in Serum/LIF, t2i/LIF and 2i/LIF measured by LUMA, n=2, Data represent mean ± standard deviation (s.d.) B) Principal component analysis of the transcriptomes of ESC grown in Serum/LIF, t2i/LIF and 2i/LIF and EpiLC (n=2). C) Cluster analysis of the global transcriptomes of ESC grown in Serum/LIF, t2i/LIF and 2i/LIF in replicates. D) Correlation matrix of the global transcriptomes of mESC grown in Serum/LIF, t2i/LIF and 2i/LIF in replicates. E) z-value heatmap of all the differentially expressed genes (DEGs) between Serum/LIF and 2i/LIF for mESC grown in Serum/LIF, t2i/LIF and 2i/LIF.

4.2 Establishing an enhanced reporter of global DNA methylation (eRGM) cell lines

After establishing a system of global DNA demethylation a method to dynamically track the global methylation status at the single cell state was needed. To measure the DNA methylation state of single-cells we exploited the recently developed reporter for genomic DNA methylation (RGM) (Stelzer et al., 2015). The system is based on a GFP reporter driven by a methylation sensitive promoter which methylation status is regulated in *cis* by the methylation changes of an adjacent sensor region (Figure 4A). The strength of the reporter system depends on a unique feature of imprinted promoters. They display intrinsic sensitivity to DNA methylation of adjacent genomic regions and are not regulated by the DNA methylation machinery in a tissue specific manner (Ferguson-Smith, 2011). This is evident as the two parental loci of imprinted promoters are differently methylated and active depending on the *cis* DMR and independent on the cellular context (Lin et al., 2003). Therefore, the GFP expression is regulated by the methylation status of the sensor region and is independent from the underlying sequence of the sensor as the GFP is driven directly by the imprinted promoter.

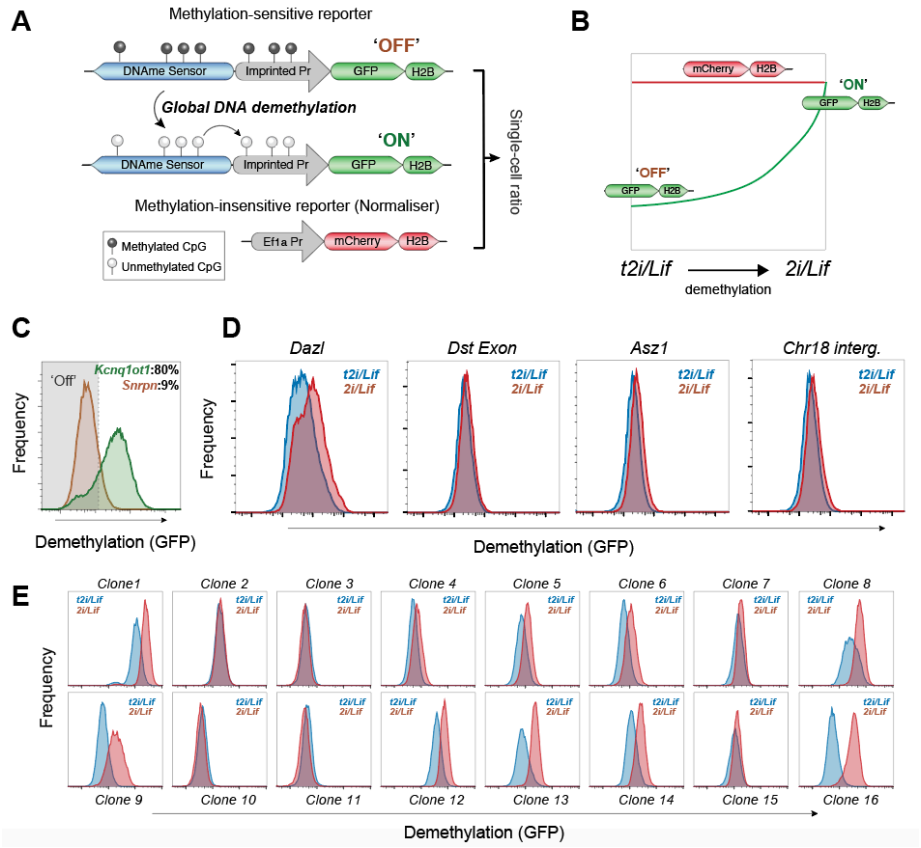


Figure 4. Generation of enhanced Reporter of Global Methylation (eRGM) mESC lines.

A) Schematic of the design of the eRGM system. B) Schematic of the GFP and mCherry changes during demethylation transition from t2i/LIF to 2i/LIF. C) GFP distribution of ESC in 2i/LIF carrying Kcnq1ot1-GFP or Snrpn-GFP reporters as measured by flow analysis used to optimize the selection of the imprinted promoter for eRGM. D) Comparison of GFP distribution of ESC carrying different sensors in antisense upstream of Kcnq1ot1-GFP reporter in both methylated (t2i/LIF) and demethylated (2i/LIF) conditions as measured by flow analysis used to optimize the selection of the sensor region for eRGM. E) Comparison of GFP distribution for different clonal integration of the Dazl-Kcnq1ot1-GFP reporter between methylated (t2i/LIF) and demethylated (2i/LIF) conditions.

To further optimize the system and to make it more compatible with CRISPR screening I made two changes to the previously published reporter of global methylation (RGM) (Stelzer et al., 2015). First, we compared the original *Snrpn* imprinted promoter and the *Kcnq1ot1* imprinted promoter. *Kcnq1ot1* enhanced the dynamic range of reporter activity (eRGM), enabling better separation of hypo- and hyper-methylated cells (Figure 4C) and was thus used from there on. Additionally, we introduced an additional Ef1a-mCherry reporter to the

system that is not sensitive to DNA methylation, converting the read-out to a ratiometric measure (Figure 4A). This enables a single-cell ratiometric score (GFP:Ef1a-mCherry) that normalises for general confounding effects on a reporter in a screen (e.g. disruption of translation factors) or inherent cell-cell variance (e.g. cell cycle stage) (Figure 4B). Finally, I needed a sensor that faithfully captured the global DNA methylation dynamics in the system. Two developmental promoters, *Dazl* and *Asz1*, an exon region on *Dst* and an intergenic region on chromosome 18, all regions that are known to be DNA methylation sensitive in ESC, were cloned in an antisense orientation upstream of Kcnq1ot1-GFP to ensure that the transcription of the reporter was not driven from the sensor region. These clones were *in silico* methylated and then integrated into mESC using the piggyBac system and tested by comparing GFP expression by flow cytometry in t2i/LIF (hypermethylated) and then after addition of full 2i (hypomethylated). Of the tested sensor regions, the *Dazl* promoter exhibited the greatest dynamics in GFP expression going from methylated to demethylated conditions (Figure 4D). Using the piggyBac system, the reporter is randomly integrated into the genome, generating high variation of the reporter activity between cells that is not indicative of DNAme changes. To expand the dynamic range of the *Dazl* reporter (eRGM), I decided to test single clonal integrations of the reporter by generating independent ESC lines carrying *Cas9* and the ratiometric eRGM system (Dazl-Kcnq1ot1-GFP + EF1a-mCherry). Two eRGM clones (clone 16 and clone 13, hereafter eRGM Cell lines 1 and 2) demonstrated strong separation of GFP expression between hypermethylated and hypomethylated conditions (Figure 4E) and were used here after.

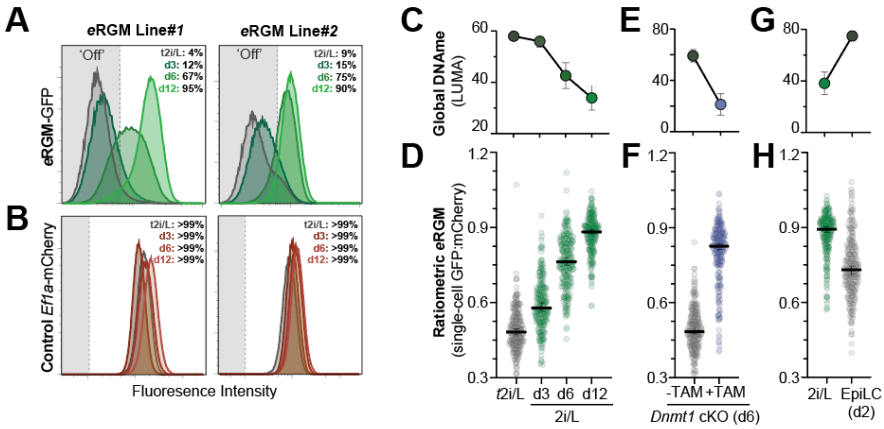


Figure 5. Dynamics of eRGM during DNA methylation transitions.

A-B) Changes in GFP A) or mCherry B) expression as measured by flow analysis during 12 day transition from t2i/L to 2i/L in independent eRGM cell lines (Line #1 and #2). C,E,G) Changes in global DNA methylation measured by LUMA C) during 12 transition from t2i/L to 2i/L E) after conditional DNMT1 KO and G) after EpiLC differentiation. Data represent mean ± s.d. (n=2) D, F, H) Plot showing ratiometric (GFP/mCherry) activation of eRGM of single cells C) during 12 transition from t2i/L to 2i/L E) after TAM induced *Dnmt1* KO and G) after EpiLC differentiation.

4.3 Validation of eRGM during DNA demethylation transition

I further explored the reporter using the established eRGM Cell Lines #1 and #2 over a 12-day transition from t2i/LIF hypermethylated to full 2i/LIF hypomethylated culture measuring both global DNA methylation using LUMA and GFP expression using flow analysis. eRGM Line #1 displayed better separation between hyper- and hypomethylated conditions and in t2i/L eRGM #1 was silenced in >95% of cells, consistent with high global DNA methylation (Figure 5A and 5C). In 2i/LIF the eRGM exhibited a progressive single cell activation (Figure 5D) concomitant with induced DNA methylation erasure (Figure 5C) leading to eRGM #1 activation in 12% of single-cells after 3 days, 67% after 6 days, and in >95% of cells upon complete DNA hypomethylation at 12 days. Independent eRGM line #2 exhibited consistent response to induced hypomethylation albeit quicker and reached almost complete separation after 6 days in 2i/LIF (Figure 5A). Notably, Ef1-mCherry did not alter expression during this transition for both eRGM lines enabling its use as a ratiometric normalizer (Figure 5B). To further validate the eRGM and its ability to detect methylation perturbation I utilized the fact that the system was integrated into ESC carrying conditional alleles for *Dnmt1*. Addition of tamoxifen induced CRE-mediated deletion of *Dnmt1* and without DNMT1 the global DNA demethylation achieved (Figure 5E) was picked up by the reporter, as a strong single cell activation of eRGM was observed (Figure 5F). Finally,

the reporter was tested for its ability to respond to gain of DNA methylation. Differentiation of the hypomethylated 2i/Lif ESC into hypermethylated EpiLC (Figure 5G) initiated silencing of the eRGM even over a short period of time (Figure 5H). Thus, the enhanced ratiometric reporter of genomic DNA methylation (eRGM) represents a single-cell read out for dynamic transitions of cellular DNA methylation status.

4.4 CRISPR-Cas9 screen for DNA methylation maintenance regulators

To identify potential regulators of DNA methylation maintenance or reprogramming, unbiased CRISPR-Cas9 knock-out screening was performed using the independent eRGM cell lines #1 and #2. The eRGM lines carrying Cas9 were kept hypermethylated in t2i/LIF and transduced with a lentiviral knock out library of 78637 sgRNAs targeting 19674 mouse protein coding genes (4-fold sgRNA against each gene) (Doench et al., 2016a). To validate the effectiveness of the genome wide knock out screen the gRNAs counts from the unsorted populations were compared to the initial distribution of the gRNA library. Among the most depleted genes (FDR <0.01) from the unsorted population are *Sox2, Nanog* and *Oct4* along with genes that belong to essential cell functions which confirms the efficacy of the screen. Gene ontology analysis of depleted genes (FDR <0.001, total of 182 genes) revealed: Ribonucleoprotein 1.72^{-54}, ribosome 1.68^{-44} and translation 1.31^{-41} (Figure 6A). The unsorted population also had strong enrichment of *Trp53* (FDR = 0.00099), a tumor suppression gene often found to be enriched in mESCs in genome wide knock-out screens as its removal increases proliferation (Joung et al., 2017b). 10 days after transduction and puromycin selection for viral integration, the cells that had activated eRGM (top 1%) under hypermethylated conditions were isolated using flow cytometry sorting. Since the reporter is hypermethylated and silenced these GFP positive cells might contain a sgRNA that resulted in a knockout of a gene with role in maintaining eRGM silencing. Analysis was done by next generation sequencing of amplified sgRNAs from the unsorted control population and compared to the top 1 % sorted population using MAGeCK (Figure 6B). This comparison revealed enrichment of sgRNAs targeting *Dnmt1* (average rank=9.5 and average Robust Rank Aggregation (RRA) score 1.21e-7) and *Uhrf1* (average rank 82, average RRA score 3.45e-4). Picking up known factors of the DNA maintenance machinery confirms the ability of the approach to identify regulators of DNA methylation.

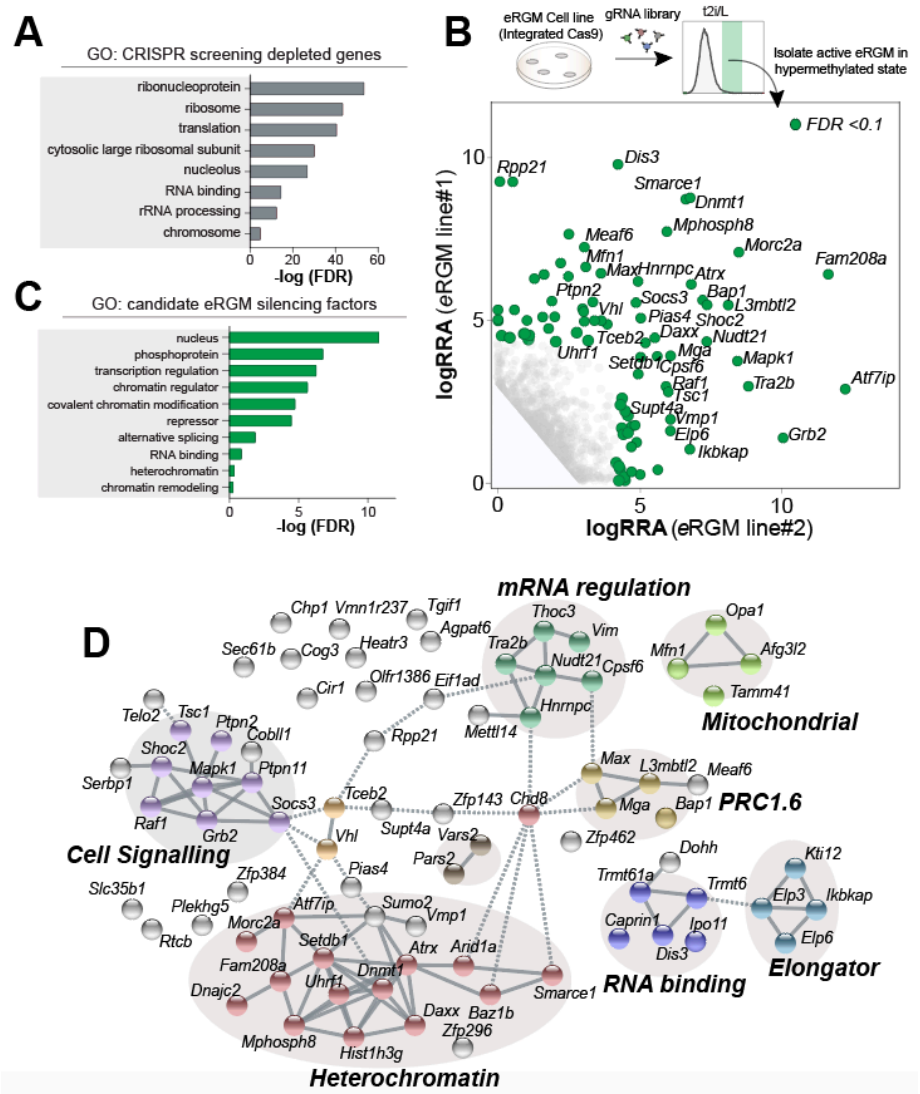


Figure 6. DNA methylation maintenance screen results.

A) Gene ontology analysis for the top 100 genes targeted by significantly depleted sgRNA from an unselected transduced population as compared to the library sgRNA distribution. B) Results of two independent CRISPR screens for gene KO that enable eRGM activation under hypermethylated conditions (t2i/L). C) Gene ontology and D) STRING clustering of the intersected list of significant (FDR <0.1) candidates required for eRGM repression from Figure 6B.

Moreover, the screen also revealed the regulators of chromatin-mediated silencing such as *Setdb1* (average rank 37.5, average RRA score 2.97e-5) and its cofactor *Atf7ip* (average rank 98.5, average RRA score 7.1e-4) and the HUSH complex (*Mophosph8*, *Morc2a* and *Fam208a*). By intersecting

significant hits (FDR<*0.01*, FC>3) from both independent screens a candidate list of 78 putative genes linked with maintaining the epigenome in ESC was generated. Gene ontology analysis of the hits revealed enrichment in the nucleus (FDR= 1.36^{-11}), chromatin regulator (FDR= 1.87^{-6}) and covalent chromatin modification (FDR= 1.53^{-5}) consistent with roles in epigenome regulation (Figure 6C). MCF clustering in STRING revealed clusters of genes enriched for functional interactions. In addition to the HUSH complex other genes with roles in heterochromatin formation and chromatin remodeling were identified (*Atrx, Daxx, Smarce1* etc.) as well as genes related to PRC1.6 mediated regulation (*Max*, *Mga*, *L3mbtl2 and Bap1*). A cluster of cell signaling linked genes (*Mapk1, Raf1, Shoc2* etc.) either acting as activators of MERK/ERK pathway or repressors of JAK/STAT were also identified. Finally, genes related to mRNA regulation (export, splicing and polyadenylation; *Tra2b, Thoc3, Nudt21 and Cpsf6*) were picked up (Figure 6D). These clusters of enriched genes provide potential means of epigenetic maintenance during development.

4.5 Validation of DNA methylation maintenance regulator candidates

To validate the candidates from the epigenetic maintenance screen I transfected gRNAs into both eRGM Line #1 and #2, targeting 26 candidates with role in maintaining epigenetic silencing. After introducing the targeting sgRNA the eRGM lines were kept in t2i/LIF for 10 days and the GFP activity was measured using flow cytometry. In the control, using non-targeting sgRNA, only 9% of cells were GFP positive. However, the knockout of 26 candidate genes resulted in a gradient increase in GFP positive cells ranging from 20.3% to 93.9% (Figure 7A). This disruption of candidate factors was confirmed in eRGM line #2 (Figure 7B), overall demonstrating the role of the candidate genes in epigenetic silencing at the reporter. Next, I investigated whether the observed eRGM activation is a direct consequence of the loss of global DNA methylation upon candidate KO. Assessing the global DNA methylation level in candidate knockout lines using LUMA showed that knockout of only 3 candidate genes, including *Dnmt1*, resulted in loss of global DNA methylation compared to non-targeting control in t2i/Lif in both independent eRGM lines (Figure 7C). Only genes with roles in cell signaling (*Shoc2* and *Socs3*) affected the methylome. However as demonstrated earlier the reporter is able to identify key methylation regulators (*Dnmt1* and *Uhrf1*) (Figure 6B), indicating that in this direction the screen was saturated for factors involved in epigenetic silencing without affecting the global level of DNA methylation.

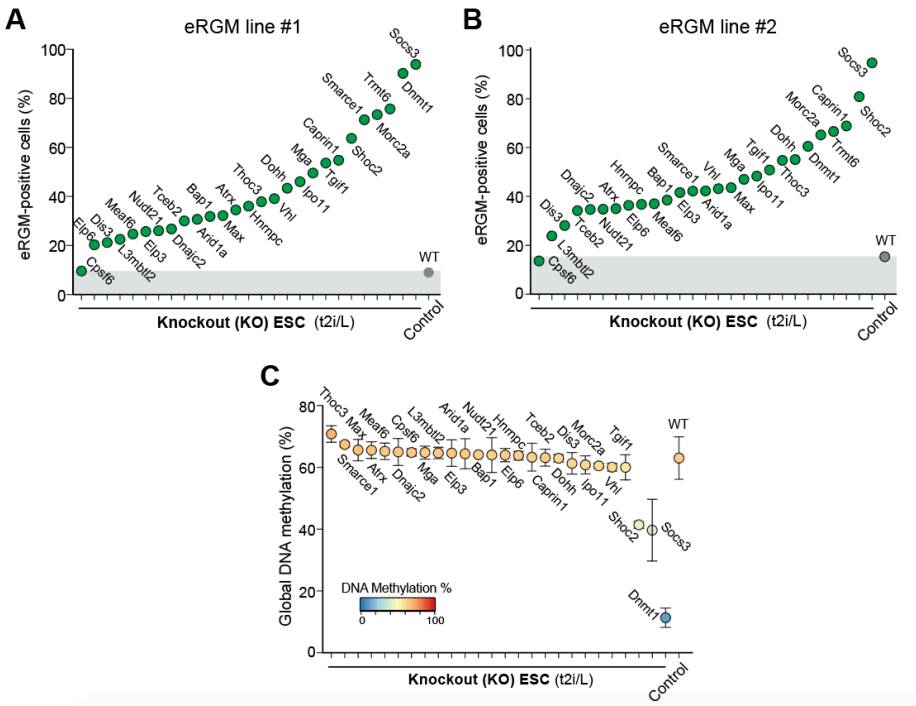


Figure 7. Validation of DNA methylation maintenance candidates.

A-B) Percent of single-cells with eRGM 'on' in hypermethylated t2i/L following knockout of indicated candidate genes (26) and control (non-targeting gRNA). C) LUMA quantification of global DNA methylation levels in indicated knockout lines of candidate genes compared to control (non-targeting gRNA) of mESC in t2i/LIF. Data represent mean ± s.d. (n=2).

4.6 CRISPR-Cas9 screen for DNA methylation reprogramming regulators

Next, to identify factors that contribute to resetting the epigenome in naïve cells, I induced DNA demethylation in the transduced eRGM cell lines #1 and #2 using full 2i/LIF. After 12 days in 2i/LIF culture, using flow cytometry sorting, the individual cells that failed to ratiometrically activate GFP with mCherry stable between conditions were isolated. Since the reporter becomes hypomethylated in 2i/LIF and GFP is activated, the cells that fail to turn on GFP have a knockout of a gene with a role in ensuring eRGM activation or reprogramming (Figure 8A). Analysis of this sorted population of GFP negative cells by comparing sequenced sgRNA to the unsorted control population using MAGeCK revealed enrichment of sgRNAs targeting *Prdm14* (average rank 19.5 and average RRA score 4.3^{-7}) which is a key regulator of naïve

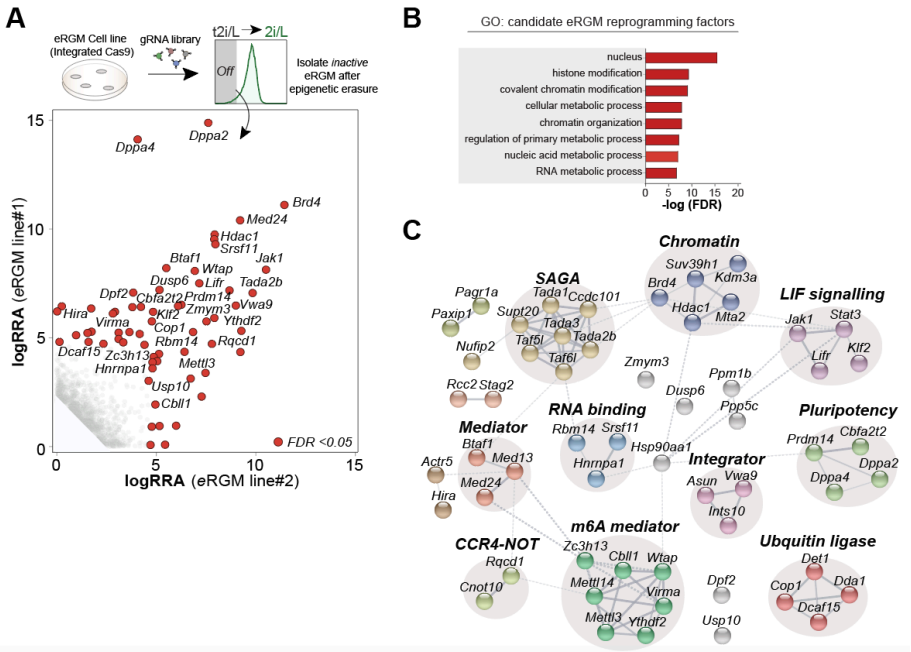


Figure 8. DNA methylation reprogramming screen candidates.

A) Gene ontology analysis for the top 100 genes targeted by significantly depleted sgRNA from an unselected transduced population as compared to the library sgRNA distribution. B) Results of two independent CRISPR screens for gene KO that enable eRGM activation under hypermethylated conditions (t2i/L). C) Gene ontology and D) STRING clustering of the intersected list of significant (FDR <0.1) candidates required for eRGM repression from Figure 4B.

pluripotency and its perturbation causes hypermethylation in ESC (Leitch et al., 2013; Okashita et al., 2014). Enrichment of *Prdm14* and its cofactor *Cbfa2t2a* (average rank 37.5, average RRA score 3.0^{-5}) demonstrates the power of the approach to identify reprogramming factors. To generate a final list of hits linked with resetting the epigenome in ESC, the significant hits (FDR<0.05, FC>3) from both independent screens were intersected, giving a list of 56 genes. Gene ontology analysis suggested of the 56 putative genes were preferentially enriched in the nucleus (FDR=3.2^{-16}), with top hit biological processes related to 'nucleic acid metabolic processes' (FDR=1.1^{-7}) and 'histone modification' (FDR=4.3^{-10}) in agreement with roles in epigenetic regulation (Figure 8B). MCF cluster analysis in STRING revealed 13 functional clusters (Figure 8C). The top hits *Dppa2* and *Dppa4* clustered as pluripotency linked genes with *Prdm14* and *Cbfa2t2a*. Opposite to the DNA maintenance screen, genes corresponding to the JAK/STAT pathway and inhibiting MERK/ERK signaling (*Lif, Jak, Stat3, Dusp6*) were picked up.

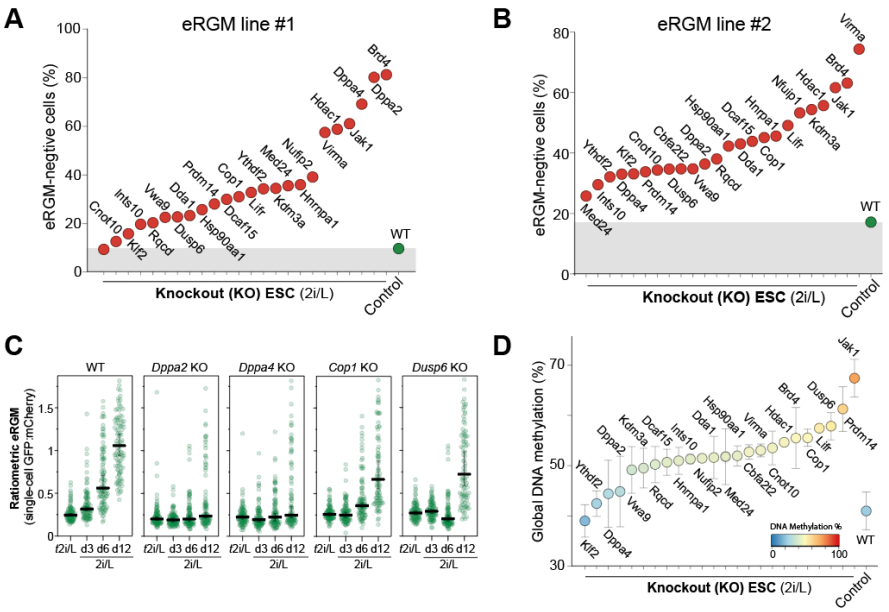


Figure 9. Validation of DNA methylation reprogramming candidates.

A-B) Percent of single-cells with eRGM 'off' in hypomethylated 2i/L following knockout of indicated candidate genes (24) and control (non-targeting gRNA). C) Dynamics of ratiometric eRGM activation amongst single-cells during t2i/L to 2i/L transition in WT, *Dppa2*, *Dppa4*, *Dusp6* and *Cop1* KO lines. D) LUMA quantification of global DNA methylation levels in indicated knockout lines of candidate genes compared to control (non-targeting gRNA) of mESC in 2i/LIF. Data represent mean ± s.d. (n=2).

Moreover, multiple candidate factors involved in m6a RNA methylation (*Virma, Ythdf2, Mettl14, Z3h13, Mettl3, Cbll1*), ubiquitin ligases (*Cop1, Det1, Dda1, Dcaf15*), and the mediator (*Btaf1, Med13, Med24*) were observed.

4.7 Validation of DNA methylation reprogramming candidates

To validate the factors with role in resetting the epigenome, KO ESC populations were generated for the top 24 candidates by integrating a targeting sgRNA with piggyBAC into both eRGM Line #1 and #2. After introducing the targeting sgRNA, the eRGM lines were cultured in 2i/Lif to induce global DNA demethylation and the GFP activity was tracked using flow cytometry after 3, 6 and 12 days in 2i/Lif. In the control, using non-targeting sgRNA, only ~9% off cells were still GFP negative after 12 days compared to most KOs of candidate genes which resulted in delayed or near complete block of GFP activation (Figure 9A). This disruption of candidate factors was confirmed in eRGM line #2 day 6 in 2i/Lif, in concordance with the maximum separation of GFP

between ti2/Lif and 2i/Lif being established after 6 days in eRGM line #2 (Figure 9B). This indicates altered epigenome remodeling upon perturbation of the candidate genes. Interestingly the dynamics of eRGM during induced epigenetic resetting varied amongst candidate KO lines, with some exhibiting a delayed response and/or partial impairment relative to control (*Dusp6, Rfwd2, Prdm14*), whilst others resulting in an absolute block in eRGM activation (*Brd4, Dppa2 and Dppa4*) (Figure 9C). This indicates that the deletion of different candidate regulators results in distinct effects on epigenetic reprogramming dynamics, suggesting different modes of action. Failure to activate eRGM in 2i/Lif could indicate only a focal repression at the reporter loci. Next, to investigate whether reduced eRGM activation is directly indicative of incomplete epigenomic reprogramming upon candidate KO, I used LUMA to quantitatively measure global DNA methylation. Assessing the global DNA methylation level in candidate knockout lines demonstrated that knockout of most candidate genes resulted in increased global DNA methylation compared to non-targeting control in 2i/LIF in both independent eRGM lines (Figure 9D). This validated the previously known epigenetic regulator *Prdm14*, which maintained 64-58% global 5mC relative to WT control (39%). Other novel candidates that exhibited substantially impaired DNA methylation resetting upon knockout include Dusp6 (60-56%), the tyrosine kinase *Jak1* (70-65%), the epigenetic regulator *Brd4* (59-51%), and the E3 ubiquitin ligase *Cop1* (56-54%). Thus, these data validate the effect of both known and novel factors on global DNA methylation and demonstrating the strength of the epigenetic resetting screen to identify regulators of DNA demethylation in naïve ESC.

4.8 *Cop1* and *Dusp6* regulate global DNA methylation

To further investigate the potential role of the novel candidates on DNA methylation regulation I generated *Dusp6* and *Cop1* clonal knockout ESC lines (*Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$) in duplicate. *Dusp6* and *Cop1* are among the top hits with validated effect on global DNA methylation but are expected to affect the epigenome through different mechanisms. DUSP6 is a phosphatase that acts downstream of MEK to attenuate the ERK signal cascade, whilst COP1 mediates ubiquitination and proteasomal degradation of target proteins (Dornan et al., 2004; C. Li et al., 2007). Measuring the global DNA methylome using EM-seq, an enhanced equivalent of whole genome bisulfite(WGBS)-seq, confirmed elevated DNA methylation, 19.5% and 11.0% increase, in Dusp6$^{-/-}$ and Cop1$^{-/-}$ mESC respectively. Further inspection of the DNA methylome revealed that both knockouts resulted in a general block to DNA demethylation across all genomic features tested (such as exons, promoters, repeats and intergenic regions) resulting in a DNA methylome more similar to S/L and EpiLC than 2i/Lif ESC (Figure 10A and B). To explore the potential

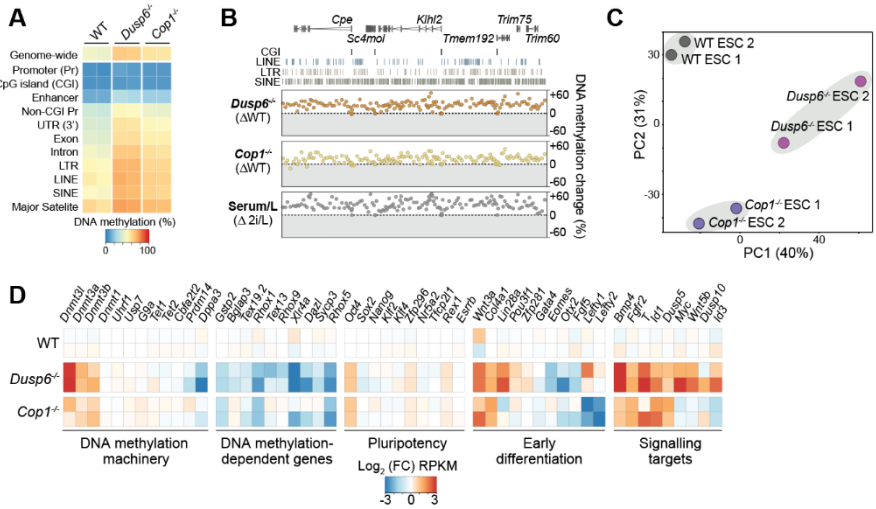
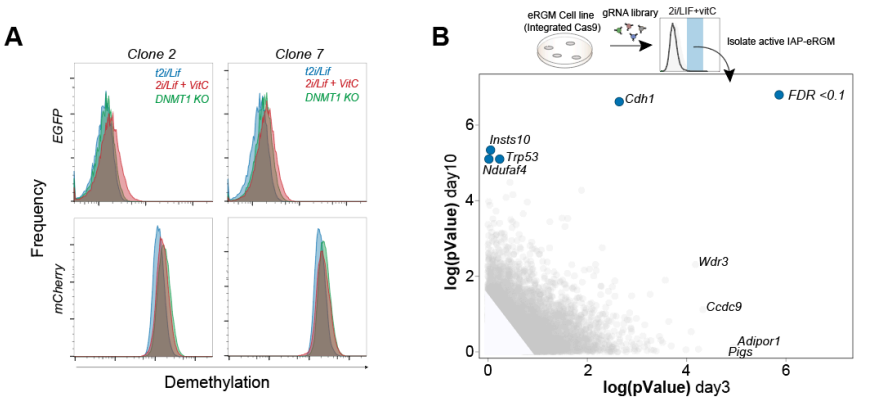


Figure 10. *Cop1* and *Dusp6* as regulators of global DNA methylation.

Heatmap showing DNA methylation levels genome-wide and at indicated genomic features in WT, *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ naïve ESC. B) Sample genome view of DNA hypermethylation in *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ 2i/LIF ESC and WT Serum/LIF ESC, shown as a delta over WT 2i/L ESC. Each datapoint represents average methylation change over a 100 CpG-tile. C) Principal component analysis of global transcriptomes in WT, *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ ESC. D) Heatmap of log fold changes in gene expression normalised to WT ESC for selected gene clusters in *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ ESC.

mechanisms underlying DUSP6 and COP1 function on DNA methylation, I examined the transcriptome of the Dusp6$^{-/-}$ and Cop1$^{-/-}$ lines using RNA-seq. Both knockouts exhibited upregulation of the de novo methylation machinery (*Dnmt3a*, *Dnmt3b*, *Dnmt3L*), although weaker in Cop1$^{-/-}$ (Figure 10D). Additionally, Dusp6$^{-/-}$ cells downregulated *Dppa3* which together with the upregulation of the de novo methylation machinery might contribute to impaired global demethylation. COP1 is involved in ubiquitination and stability regulation of proteins and thus one could expect changes at the protein level rather than the transcriptional upon *Cop1* perturbation. Thus, targets of COP1 might include upstream DNA methylation regulators such as UHRF1. Further studies could uncover the precise modes of action for both knockouts. In concordance with increased global DNA methylation strong repression of DNA methylation sensitive genes was observed while pluripotency genes were mostly unaffected (Figure 10D). Although overall relatively few genes were perturbed in both knockout lines, their transcriptomes were clearly separated from wild type 2i/LIF cells which may reflect their disrupted epigenetic state (Figure 10C). In summary, screening for factors with role in epigenetic reprogramming, I have identified and functionally confirmed novel candidates involved in DNA methylation regulation.



C IAP-eRGM top hits day 3

id	RRA	p-Value	FDR	Rank	Good gRNA	log FC
Adipor1	2.89E-06	1.13E-05	0.131188	1	4	2.6252
Pigs	3.43E-06	1.27E-05	0.131188	2	3	2.234
Ccdc9	1.19E-05	4.68E-05	0.272277	3	4	1.7988
Dhx8	1.80E-05	5.74E-05	0.272277	4	3	2.9643
Ddx56	2.56E-05	8.04E-05	0.276403	5	3	2.6806
Arhgap44	2.58E-05	1.02E-04	0.30198	6	3	1.6167
Wdr3	3.87E-05	6.60E-05	0.272277	7	2	0.08556
Phyh	5.28E-05	2.03E-04	0.520902	8	2	1.3578
Gm3376	6.69E-05	2.67E-04	0.55099	9	2	1.7372

D IAP-eRGM top hits day 10

id	RRA	p Value	FDR	Rank	Good gRNA	log FC
Cdh1	2.86E-09	2.41E-07	0.00495	1	4	2.4442
Insts10	1.26E-06	4.58E-06	0.04084	2	4	2.7844
Ndufaf4	2.16E-06	7.95E-06	0.04084	3	2	2.0003
Trp53	2.19E-06	7.95E-06	0.04084	4	4	1.2836
Zfp830	8.73E-06	3.30E-05	0.13564	5	2	-0.1245
Fam50a	1.63E-05	5.32E-05	0.18234	6	2	0.80351
Zfp420	2.64E-05	1.05E-04	0.22426	7	1	-2.0104
Xrn1	2.73E-05	1.09E-04	0.22426	8	2	1.7153
Nhlrc2	3.41E-05	1.32E-04	0.24618	9	4	1.9752

Figure 11. Screening for regulators of IAPs using IAP-eRGM cell lines

A) Distribution of GFP and mCherry expression for IAP eRGM clones 2 and 7 in t2i/L, 2i/L and 2i/L+vitamin C. B) Results of two CRISPR screens for genes that enable IAP eRGM activation under hypermethylated condiations. C-D) Top hits from IAP-eRGM C) day 3 and D) day 10.

4.9 CRISPR-Cas9 screen for IAP regulators

The eRGM system can be modified using different sensor regions to learn more about DNA methylation regulation at a specific region. Using a transposon element, intracisternal A-type particles (IAP), as a sensor for the eRGM coupled with genome wide CRISPR knock out screening I attempted to discover novel genes with roles in epigenetic regulation at IAP. To identify regulators of IAP, a 1kb evolutionarily conserved IAP LTR region was cloned in antisense upstream of Kcqn1ot1-GFP and the reporter was integrated into ESC using the piggyBac system. The IAP reporter clones were only partly activated (Figure 11A) during extensive demethylation using 2i/LIF + vitamin C (Blaschke et al., 2013) as is expected when using the reporter system with a transposable element like IAP that retains DNA methylation in global demethylation events (Seisenberger et al., 2012). The IAP-eRGM ESC were screened for gene knock outs that enable spontaneous GFP activation indicating a role in suppressing IAP. The top 2.5% GFP expressing cells as measured by flow cytometry were sorted after 3 days and 10 days in vitamin C. Few significant (0.1 <FDR) hits were identified on either day with low overlap of hits and no known regulators of IAPs (*Setdb1* and *Trim28*) (Matsui et al.,

2010; Rowe et al., 2010) were identified (figure 11B-D). Furthermore, none of the top candidates had any connection to epigenetic mechanisms and were most likely false positives. The IAP element used as a sensor in the reporter was highly methylated and was a conserved element most likely under redundant epigenetic repression that could not be relieved in this screen.

4.10 Altered DNA methylome in *Dppa2* and *Dppa4* mutants

Atypically for top hit candidates (e.g. *Brd4*), deletion of the paralogues *Dppa2* or *Dppa4* did not lead to significant impairment of global DNA demethylation in naïve ESC (Figure 9D). The screen is designed to identify both global and/or focal epigenome regulators. This could imply that DPPA2 and DPPA4 modulate the methylation landscape at specific genomic features. Analysis of the expression of *Dppa2* and *Dppa4* from single cell RNA-seq data revealed an expression pattern that coincides remarkably with epigenetic reprogramming during development; going from low expression in oocytes to specific activation during pre-implantation development before being silenced in post-implantation embryo (Figure 12A). To examine the effect of DPPA2 and DPPA4 on the DNA methylome, three independent *Dppa2*$^{-/-}$ or Dppa4$^{-/-}$ ESC lines, as well as three double KO (DKO) lines (*Dppa2*$^{-/-}$ and Dppa4$^{-/-}$) were generated and subjected to EM-seq (Figure 12B). Consistent with the previous LUMA measurement, *Dppa2/4* knock-out did not impact the globally DNA hypomethylated status of naïve ESC, on the contrary, it led to a minor reduction of global DNA methylation (Figure 12C). However, 1662 and 1725 differentially methylated regions (DMR) (logistic regression adjusted $p<0.05$ & binomial test $p<0.01$) in the *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ lines respectively were identified when the methylome was analysed locally using 50-CpG genomic tiles (Figure 12D). Remarkably, the vast majority of these DMRs (>96%) correspond to loci with elevated DNA methylation, suggesting a general acquisition of focal hypermethylation in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESCs. In summary, perturbation of *Dppa2* or *Dppa4* results in aberrant methylome with multiple loci gaining DNA methylation.

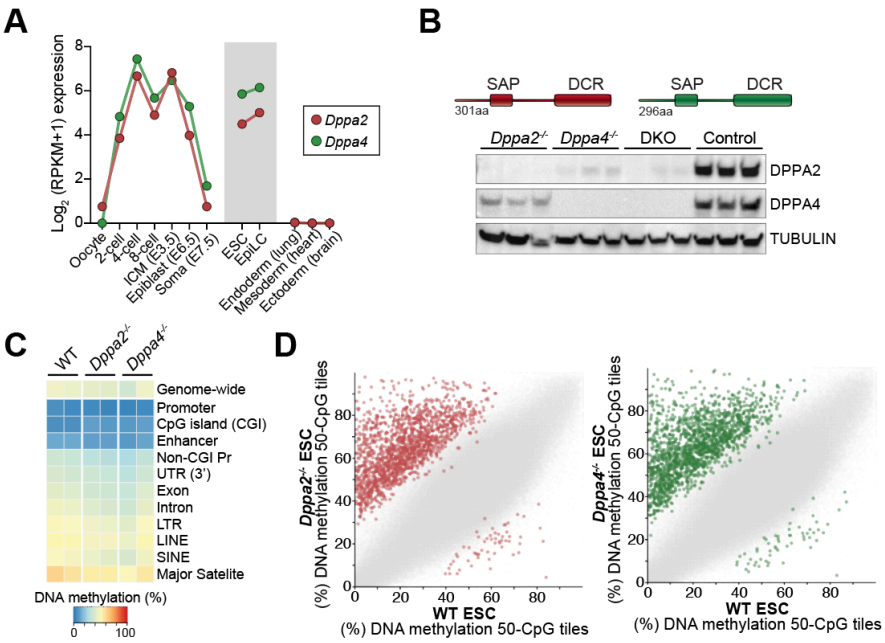


Figure 12. The DNA methylation landscape in *Dppa2* and *Dppa4* knockouts

A) Line plot of *Dppa2* and *Dppa4* expression dynamics during early development, in pluripotent ESC and EpiLC, and in selected somatic tissues. B) Western blot confirming loss of DPPA2, DPPA4, or both (DKO) protein(s) in clonal knockout ESC lines. C) Heatmap displaying DNA methylation levels genome-wide and at selected genomic features in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ naïve ESC. D) Scatter plot of genome-wide DNA methylation at sliding 50-CpG tiles in WT versus *Dppa2*$^{-/-}$ or *Dppa4*$^{-/-}$ ESC. Significant differentially methylated regions (DMR) are shown in red (Dppa2) or green (Dppa4). Significance by logistic regression ($p<0.05$) and binomial test ($p<0.01$)

4.11 DPPA2 and DPPA4 ensure focal hypometehylation

Differentiation of ESC to epiblast-like cells (EpiLC) corresponds to a post-implantation development when global DNA methylation has been re-established. To investigate the persistence of the aberrant DNA hypermethylation in *Dppa2* and *Dppa4* knockouts during differentiation I profiled the DNA methylome of WT, *Dppa2*$^{-/-}$ and Dppa4$^{-/-}$ EpiLC using EM-seq. The DNA hypermethylated genomic regions established in *Dppa2*$^{-/-}$ or *Dppa4*$^{-/-}$ ESC remained in knockout EpiLC, whilst a number of additional loci also acquire aberrant de novo methylation (Figure 13A). This demonstrates that the epimutated loci in *Dppa2/4* knockout ESC are kept relatively hypomethylated in WT EpiLC. Intersecting the DMRs from both ESC and EpiLC of *Dppa2*$^{-/-}$, I attempted to explore the underlying genomic regions affected by the aberrant gain of methylation. This revealed substantial overlap of DMRs at gene promoters, full length LINE1s and genebodies (Figure 13A).

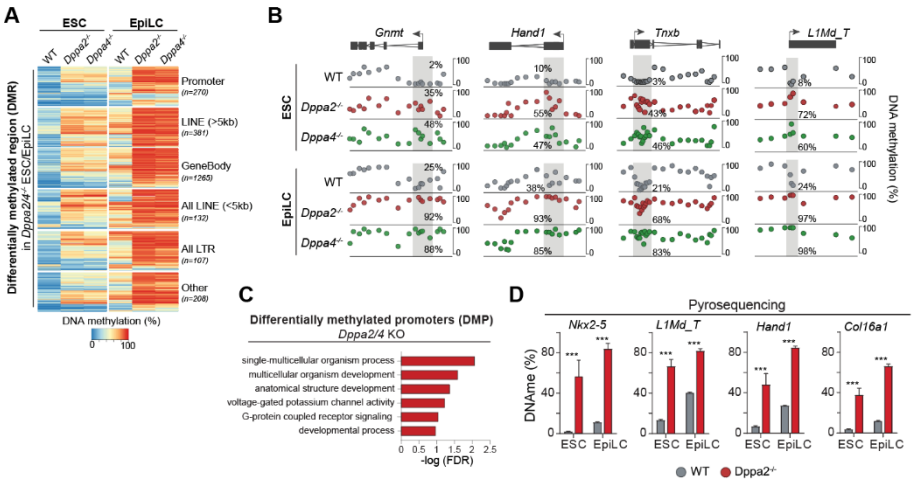


Figure 13. DPPA2 and DPPA4 ensure focal hypomethylation.

A) Heatmap showing methylation status of significant differentially methylated regions (from sliding 50 CpG windows) (DMR) in *Dppa2* and *Dppa4* knockout in ESC and EpiLC. B) Genome views showing DNA methylation patterns at selected gene promoters and LINE1 element in WT and *Dppa2/4* KO ESC and EpiLC. Each datapoint represents the windowed average methylation of 15-20 CpG sites. C) Gene ontology (GO) of differentially methylated promoter (DMP) associated genes. D) Bisufite pyrosequencing quantification showing DNA methylation at selected gene and L1Md_T promoters in WT and *Dppa2* KO in ESC and EpiLC. (n=2, each of multiple CpG sites). Student t test, adjusted p-value (Holm-Sidak), ns p>0.05, * p<0.05, ** p<0.01, *** p<0.001.

Gain of DNA methylation was seen at many gene promoters of developmental genes. As examples, the promoters of *Hand1*, *Gnmt1* and *Tnxb,* acquired significant promoter hypermethylation in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiL (Figure 13B). These developmental genes are kept repressed but usually remain hypomethylated during early development stages which implies that *Dppa2* and *Dppa4* knockouts result in gain of aberrant DNA methylation rather than a faulty demethylation. To generate a list of all differently methylated promoters (DMPs), probes at the 5' end (+/- 1kb) of all genes were used to identify 363 DMPs between *Dppa2*$^{-/-}$ and WT (see methods). Gene ontology analysis of DMPs revealed that these hypermethylated promoters are specifically enriched only for developmental processes (*multicellular organism development FDR=0.0053*; *developmental process FDR=0.024, anatomical structure development FDR=0.01*) (Figure 13C). A similar strategy was used to identity 1131 differentially methylated LINE1s (DMLs) using probes at all full length (>5kb) LINE1 promoters (+/- 500bp at the 5'). Pyrosequencing confirmed the gain of DNA methylation at the promoters of *Hand1*, *Nkx2-5*, *Col16a1* and young full length LINE1s (L1Md_T) in *Dppa2*$^{-/-}$ ESC and EpiLC (Figure 13D). These data imply that, in contrast to other candidates that have

a global effect, *Dppa2/4* could have a specific role in establishing or maintaining patterns of DNA methylation at specific loci such as developmental gene and young LINE1 promoters during epigenetic reprogramming.

4.12 DPPA2 occupies gene and full length evolutionary young LINE1 promoters

Since the gain of DNA methylation in the *Dppa2* and *Dppa4* mutants is focal it is important to reveal if the observed gain of DNA methylation is a consequence of the loss of a localized and direct activity of DPPA2 or an indirect result of *Dppa2* misregulation. DPPA2 and DPPA4 are transcription factors with known DNA binding domains, to learn more about their function and potential role in epigenetic regulation, I performed DPPA2 CUT&RUNseq to annotate DPPA2 binding sites in ESC and EpiLC. Counting DPPA2 CUT&RUN fragments over all DMRs (+/- 4kb) reveals clear genomic occupancy of DPPA2 at DMRs (Figure 14A), suggesting a loss of a localized activity of DPPA2 at sites that gain DNA methylation. Additionally, 28338 DPPA2 binding sites in ESC and EpILC were identified using MACS2. Profiling the DPPA2 binding sites revealed significant enrichment at gene promoters ($P<0.001$) and 5' end of full-length Line elements (>5kb) ($P<0.001$) with a large number of binding sites overlapping genebodies (Figure 14B). This is in agreement with the previously noted localization of DMRs (Figure 13A). Moreover, DPPA2 correlates more with full-length LINE1 than truncated LINE1 elements (Figure 14C). Analysis of the DPPA2 bindings sites using DREME MOTIF analysis revealed that DPPA2 strongly binds CG rich motifs (Figure 14D) which is reflected by its preference to bind CGI promoters over non-CGI promoters (Figure 14E). DPPA2 does not undergo redistribution in EpiLC as DPPA2 peaks are highly correlated between ESC and EpiLC (Figure 14F) and the additional DMR seen in EpiLC (Figure 13A) are most likely a consequence of increased de novo methylation in EpilC. Lastly, DPPA2 strongly occupies the *Dazl* promoter, explaining why the focal activity of DPPA2 was picked up in the screen (Figure14H).

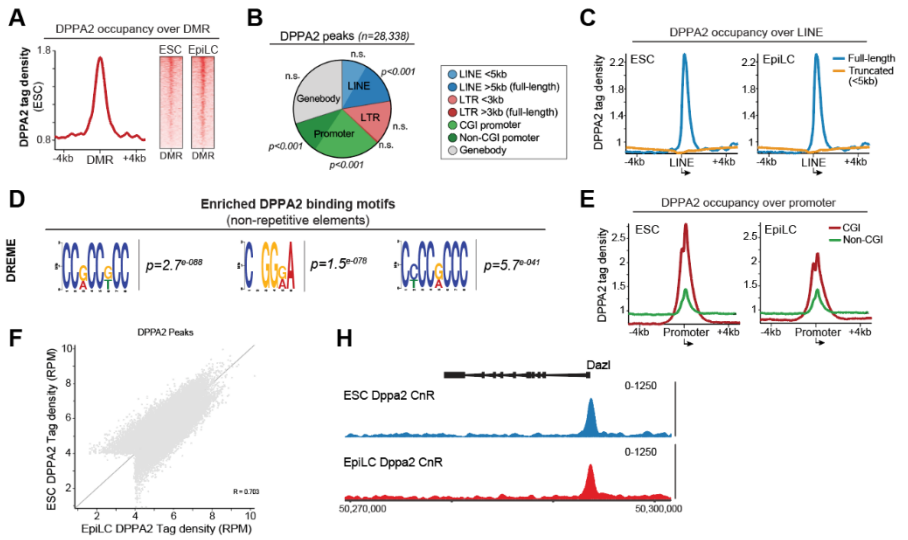


Figure 14. DPPA2 binds CG rich regions and gene and evolutionary young LINE1 promoters.

A). Distribution of DPPA2 occupancy (CUT&RUN-seq) over differentially methylated regions (DMRs) +/− 4kb in ESC and EpiLC. B) Pie plot demonstrating the partition of genomic features associated with DPPA2 binding. Student t test, adjusted p-value (Holm-Sidak). C) Distribution of DPPA2 binding at full-length (>5kb) or truncated LINE1 promoters, +/- 4kb. D) The top enriched DPPA2 binding motifs from DREME analysis of DPPA2 binding sites at non-repetitive elements. E) Distribution of DPPA2 binding at CGI and non-CGI promoters. F) Scatter plot of DPPA2 enrichment at all DPPA2 binding sites in ESC and EpiLC. H) Genome view of DPPA2 occupancy at the genomic sensor region used in eRGM (*Dazl*) in ESC and EpiLC.

In summary, *Dppa2* and *Dppa4* have a non-redundant role in safeguarding a subset of developmentally associated gene promoters and full-length LINE elements, from acquiring aberrant *de novo* DNA methylation. This gain of methylation is observed both in DNA hypomethylated (naïve ESC) and gain of methylation (EpiLC) phases because the *de novo* methylation machinery is active in both. Indeed, this also corresponds to the precise period where *Dppa2/4* are expressed, overall indicating a role for DPPA2/4 in counter acting the de novo DNA methylation machinery at specific loci.

4.13 Chromatin state at DPPA2 bound regions is altered upon *Dppa2* loss

DNA methylation interacts with different chromatin modifications in either correlative or anti correlative manner, therefore I decided to take a closer look

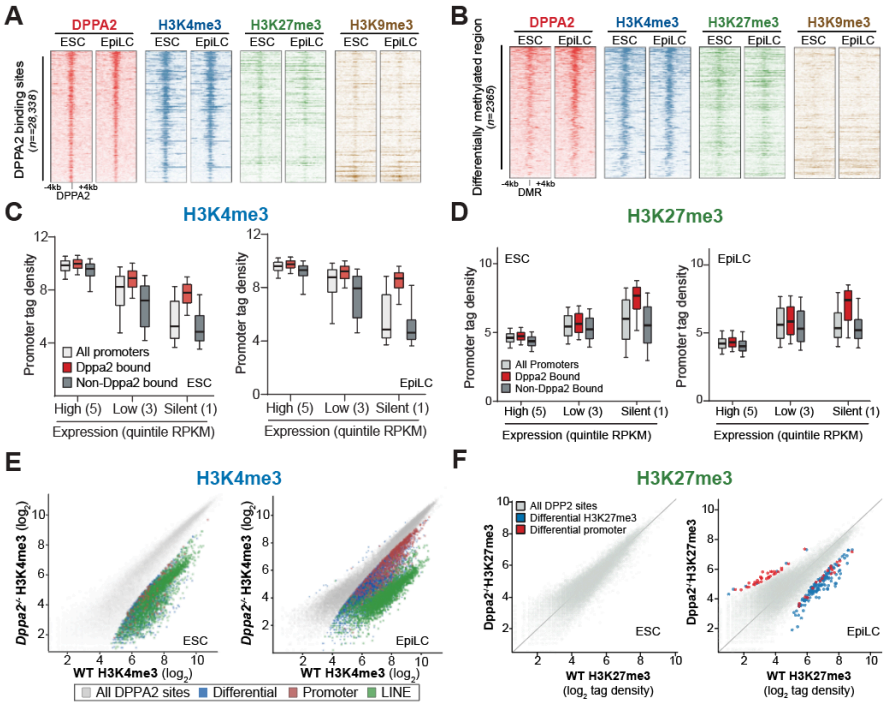


Figure 15. Analysis of the chromatin state upon loss of *Dppa2* or *Dppa4*

A-B) Density plots showing enrichment of DPPA2, H3K4me3, H3K27me3, and H3K9me3 centered on A) DPPA2 binding sites or DMRs B) +/- 4kb either side in WT ESC and EpiLC. Plots are ordered by DPPA2 binding enrichment. C-D) Boxplot showing C) H3K4me3 or D) H3K27me3 enrichment over all promoters, DPPA2-bound, and non-DPPA2-bound promoters, binned for expression quintile (RPKM). Box indicates the 25th, median and 75th percentiles, whiskers the 10th to the 90th percentiles. E-F) Scatter plot of E) H3K4me3 or F) H3K27me3 enrichment at DPPA2-binding sites in WT and *Dppa2*$^{-/-}$ ESC and EpiLC. Significant differentially H3K4me3 enriched sites overlapping promoters (red), 5´LINE (green) or neither (blue) are highlighted. Significance by LIMMA < 0.01.

at the accompanied chromatin changes at the sites gaining DNA methylation in *Dppa2* and *Dppa4* knockouts. Taking a functional genomic approach, I used CUT&RUN data from several histone modifications associated with promoters (H3K4me3 and H3K27me3) and heterochromatin (H3K9me3 and H3K27me3) and mapped over DPPA2 peaks (Figure 15A). DPPA2 bound chromatin presented a strong enrichment of H3K4me3 and weaker but notable enrichment of H3K27me3, establishing a bivalent state at few DPPA2 occupied regions, but H3K9me3 is largely depleted. Loci that acquire hypermethylation in *Dppa2/4* mutants (DMR) show similar relationships with the histone modifications as DPPA2 bound chromatin (Figure 15B), which is in line with the previously established correlation of DMRs and DPPA2 occupancy (Figure

14A). Analysis of the DPPA2 occupancy at gene promoters reveals that DPPA2 bound promoters are enriched for H3K4me3 and H3K27me3 irrespective of the corresponding gene expression state in both ESC and EpiLC (Figure 15C and D). This thus implies that DPPA2 binding is linked with direct targeting of H3K4me3, rather than DPPA2 bound promoters containing elevated H3K4me3, as enriched H3K4me3 at gene promoters is associated with increased transcription. Moreover, loss of DPPA2 resulted in depletion of H3K4me3 at a significant subset of DPPA2-bound sites in both ESC and EpiLC, with no loss at the remaining DPPA2 occupied sites. The loci losing H3K4me3 correspond to sites with intermediate H3K4me3 enrichment in WT and overlap with gene promoters and LINE1 elements (Figure 15E). However, there is only a modest reduction of H3K27me3 at DPPA2 sites in *Dppa2* and *Dppa4* mutants (Figure 15F), indicating a specific loss of H3K4me3 upon knockout of *Dppa2* and *Dppa4*. Taken together, this suggests a potential interplay between DPPA2 occupancy and H3K4me3.

4.14 Downregulation of developmental genes in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$

To further explore the effect of the epigenetic perturbation in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$, the transcriptome in ESC and during differentiation were assessed. The naïve pluripotency network (*Nanog, Klf2, Prdm14*) in *Dppa2* or *Dppa4* knockout 2i/LIF ESC is unaffected and the cells differentiate without problems to the formative EpiLC state with appropriate downregulation of the naïve markers and upregulation of the formative markers (*Fgf5*, *Wnt3*, *Dnmt3a*) (Figure 16A). Moreover, in accordance with the localized pattern of epimutations no difference in the expression of the de novo or the maintenance DNA methylation machinery was observed in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ (Figure 16A). Globally, the *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC exhibit a distinct but generally comparable gene expression signature to WT naïve ESC, with 269 and 245 genes being differentially expressed (DEGs) in *Dppa2* and *Dppa4* knockout respectively (Figure 16B and C). However, induction of EpiLC causes increasingly divergent transcriptomes as tracking with PCA demonstrates, with 802 and 611 genes being differentially expressed in *Dppa2* and *Dppa4* KOs (Figure 16B-D). Gene ontology analysis of the differentially expressed genes are specially related to developmental processes (*single-multicellular organism process* FDR=*0.000004, cell differentiation* FDR=*0.00012*) in both *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ (Figure 16E and F). Moreover, multiple DEGs fail to be activated in EpiLC (Figure 16A) in agreement with EpiLC harboring primed expression of developmental genes compared to ESC. Thus, *Dppa2* and *Dppa4* knockouts result in a failure to activate multiple genes involved in

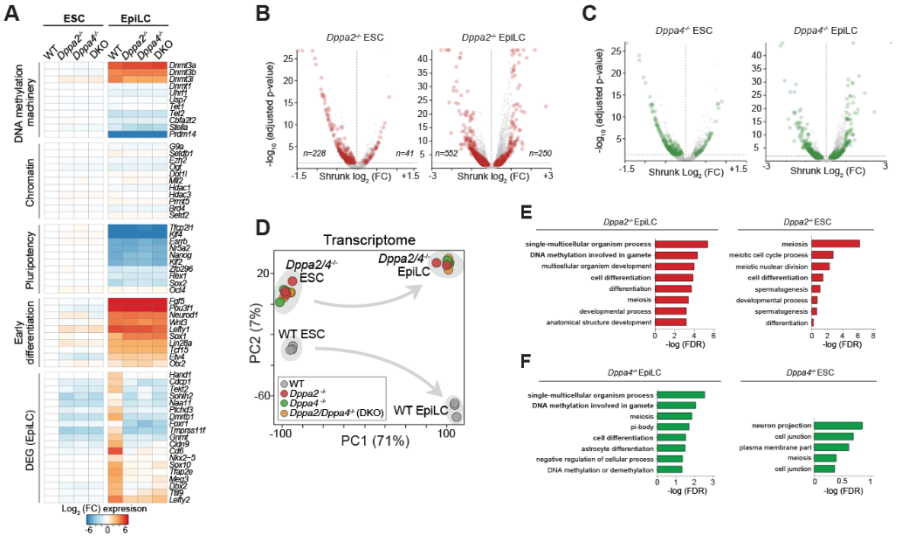


Figure 16. Transcriptional changes in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC.

A) Heatmap showing log fold gene expression changes normalised to WT ESC for selected gene clusters in WT, *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$ and DKO ESC and EpiLC. B-C) Volcano plot highlighting significant differentially expressed genes (DEG) in B) *Dppa2*$^{-/-}$ and C) *Dppa4*$^{-/-}$ ESC and EpiLC compared to WT. D) Principal component analysis of global transcriptomes of WT, *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$ and DKO ESC and EpiLC. E-F) Gene ontology (GO) of E) *Dppa2* and F) *Dppa4* KO DEGs in ESC and EpiLC.

lineage-specific functions rather than affecting pluripotency, early differentiation or epigenome gene regulatory networks.

4.15 *Dppa2* and *Dppa4* are required for permissive epigenetic state maintenance at developmental genes during pluripotency

Thus, DPPA2 bound loci are associated with H3K4me3 and *Dppa2/4* perturbation results in loss of H3K4me3 and gain of DNA methylation at the bound loci, furthermore *Dppa2* and *Dppa4* KO leads to downregulation of a set of developmental promoters. However, only a fraction of the DEGs in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC contain epimutated promoters indicating that at many genes the dysregulation is indirect. To further investigate the role of DPPA2 in epigenetic regulation at gene promoters, I analyzed the set of genes associated with the differentially methylated promoters (DMPs). Just as DMRs

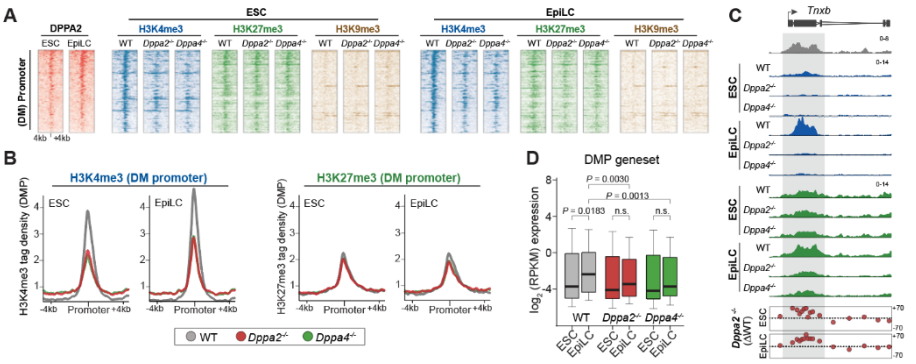


Figure 17. DPPA2 and DPPA4 are required for permissive epigenetic state maintenance at developmental genes in pluripotency.

A) Density plot revealing enrichment of H3K4me3, H3K27me and H3K9me3 centered on differentially methylated promoters (DMP) +/− 4kb in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC. Plots are ordered by DPPA2 binding enrichment. B) Trend plot showing the enrichment of chromatin marks (H3K4me3 and H3K27me3) over gene promoters associated with DMPs in *Dppa2* or *Dppa4* KO in ESC and EpiLC. C) Representative genome view of a developmental promoter showing DPPA2 occupancy, H3K4me3, H3K27me3 or H3K9me3 Cut&Run tracks and changes in DNA methylation in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC. D) Boxplot showing expression (RPKM) of genes linked with a differentially methylated promoter (DMPs) in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC. Box indicates the 25th, median and 75th percentiles, whiskers the 10th to the 90th percentiles. Data from expression (RNAseq) of 284 genes. Significance by one-tailed students t-test

and DPPA2 bound sites, the DMPs were strongly enriched for H3K4me3 in WT ESC and EpiLC with weaker H3K27me3 association (Figure 17A). Since DNA methylation is anti-correlated with H3K4me3 and H3K4me3 can directly inhibit de novo methylation, I asked if the gain of methylation seen at the DMPs in *Dppa2* and *Dapp4* mutants was combined with loss of H3K4me3. Indeed, a clear reduction of H3K4me3 was observed at almost all DMPs in ESC and EpiLC with noticeable but weak reduction of H3K27me3 upon Dppa2 or Dppa4 removal in EpiLC (Figure 17A and B). This is exemplified at the promoter of the developmental gene *Tnxb* bound by DPPA2 and in *Dppa2*$^{-/-}$ and Dppa4$^{-/-}$ there is a complete loss of H3K4me3 in both ESC and EpiLC and minor loss of H3K27m3 in EpiLC (Figure 17C). Since the genes of the corresponding DMPs are related to developmental processes I investigated how the observed loss of H3K4me3 and gain of DNA methylation affected their expression and activation. The DMP geneset includes many mesendoderm genes including *Hand1, Tnxb, Ttl9, Cldn9 and Gnmt* which are significantly upregulated in WT EpiLC (p=0.018) consistent with priming of developmental genes. These genes however fail to initiate primed expression in mutant EpiLC in either

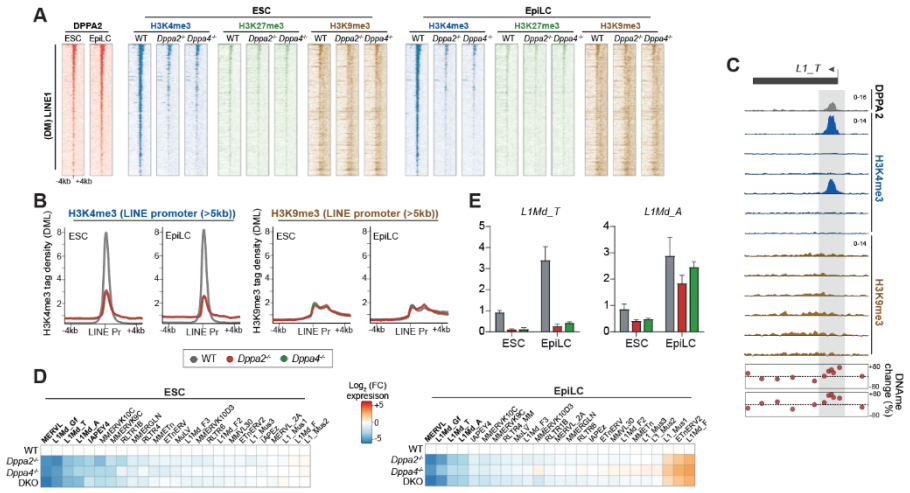


Figure 18. DPPA2 and DPPA4 are required for permissive epigenetic state maintenance at evolutionary young LINE1 during pluripotency.

A) Density plot revealing enrichment of H3K4me3, H3K27me and H3K9me3 centered on differentially methylated LINE1 (DML) +/− 4kb in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC. Plots are ordered by DPPA2 binding enrichment. B) Trend plot showing the enrichment of chromatin marks (H3K4me3 and H3K27me3) over differentially methylated full-length LINE1 in *Dppa2* or *Dppa4* KO in ESC and EpiLC. C) Representative genome view of a L1Md_T showing DPPA2 occupancy, H3K4me3, H3K27me3 or H3K9me3 Cut&Run tracks and changes in DNA methylation in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC. D) Heatmap showing log fold change of expression normalised to WT of all transposable elements in WT, *Dppa2* and *Dppa4* KO and DKO ESC and EpiLC. E) qRT-PCR quantification of L1Md_T and L1Md_A expression in ESC and EpiLC in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ normalised to WT ESC. Data represent mean ± s.d. (n=2 biologically independent experiments).

Dppa2$^{-/-}$ (p=0.29) or *Dppa4*$^{-/-}$ EpiLC (p=0.40) (Figure 17D) suggesting they have lost competence for expression with aberrant epigenetic state in ESC.

4.16 *Dppa2* and *Dppa4* are required for permissive epigenetic state maintenance at evolutionary young LINE1

The 5' end of young LINE1s displayed enriched DPPA2 occupancy (Figure 14C) and aberrant gain of DNA methylation in *Dppa2* and *Dppa4* mutants similar to a sub set of gene promoters. Analysing the subset of LINE1 with elevated DNA methylation (DMLs) in *Dppa2*$^{-/-}$ revealed a strong enrichment for H3K4me3 in WT with low H3K9me3 (Figure 18A), a mark often occupying transposon elements. Loss of *Dppa2* or *Dppa4* caused almost complete loss of H3K4me3 at all DMLs and was observed in both ESC and EpiLC (Figure 18A and B). This indicates a role for *Dppa2* and *Dppa4* in the maintenance of

a permissive epigenetic environment at a subset of evolutionary young LINE1s. This is exemplified at a LINE1_T element on chromosome 17, demonstrating the clear loss of H3K4m3 and gain of DNA methylation at the 5' LINE1 promoter in *Dppa2* and *Dppa4* KO (Figure 18C). Next, I investigated if the gain DNA methylation and loss of H3K4me3 corresponded to loss of LINE1 expression. Analysing the expression of a singular repeat element from RNAseq data is difficult because of the uncertainty of mapping the short reads and the generally low transcription of repeat elements. Therefore, the examination was done by analysing sub classes of LINE1s. A striking downregulation of full length (>5kb) LINE families (L1Md_T, L1Md_A and L1Md_Gf) was observed in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESC and EpiLC along with indirect downregulation of MERV as previously reported (Figure 18D). Using independent qRT-PCR, the repression of L1Md_T in *Dppa2* and *Dppa4* mutants was confirmed in both ESC and EpiLC (Figure 18E). *Dppa2* and *Dppa4* therefore maintain epigenetic competence at LINE1 in an analogous manner observed at developmental gene promoters, a system that may have been co-opted by LINE1 to avoid epigenetic silencing during pluripotency.

4.17 Developmental genes and evolutionary young LINE1 acquire repressive epigenetic memory upon loss of *Dppa2* or *Dppa4* during pluripotency.

To understand if the failure to upregulate the mesendoderm genes in *Dppa2/4* KO EpiLC was caused by a stable silencing or activation delay, I induced endoderm differentiation for 12 days. During the endoderm differentiation, *Dppa2* and *Dppa4* are both downregulated (Figure 19A). Therefore, any remaining repression is caused by stable epigenetic memory from earlier stages when lack of DPPA2 or DPPA4 resulted in failure to maintain permissive landscape. *Dppa2* and *Dppa4* knockout endoderm cells were morphologically equivalent to WT and successfully activated the master endoderm regulators like *Emb* and *Foxa1*, indicative of appropriate differentiation of the mutant cell lines (Figure 19B). However, the endoderm associated genes, *Gnmt, Nkx2-5, Col16a1 and Hand1* which encompassed aberrant epigenetic environment and suppression in ESC remained repressed during endoderm induction and exhibited a significant failure to be activated in mutants compared to WT (Figure 19C). Indeed, using pyrosequencing I confirmed that the ectopic promoter DNA methylation established in ESC at the same gene promoters was maintained through the 12-day endoderm differentiation (Figure 19E). Taken together, this suggests that the lack of DPPA2 and DPPA4 in earlier pluripotent phases leads to impaired competence for appropriate gene activation at later time points in development.

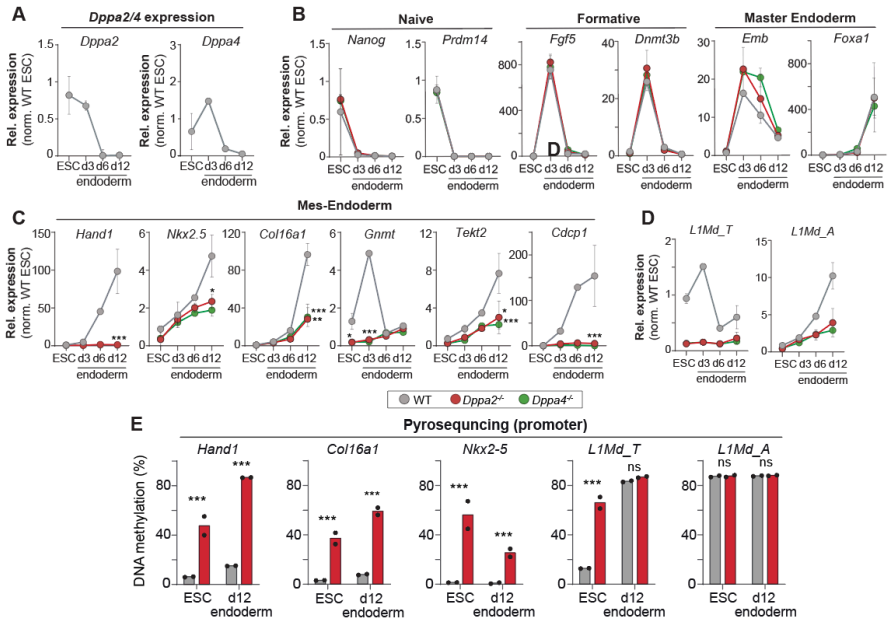


Figure 19. Developmental genes and LINE1 acquire transcriptionally repressive memory in *Dppa2/4*-mutants.

A-D) qRT-PCR quantification of expression of selected genes and LINE1 during ESC to endoderm differentiation in WT, *Dppa2*$^{-/-}$ *and Dppa4*$^{-/-}$ normalised to WT ESC. Data represent mean ± s.d. (n=2 biologically independent experiments) Significance by adjusted students t-test (Holm-Sidak), ns p>0.05, * p<0.05, ** p<0.01, *** p<0.001. **H.)** Bisulfite pyrosequencing quantification of DNA methylation at selected promoters and LINE1. (n=2 biologically independent experiments each of multiple CpG sites). Significance by adjusted students t-test (Holm-Sidak), ns p>0.05, * p<0.05, ** p<0.01, *** p<0.001.

4.18 H3K4me3 and DNA methylation interaction direct epigenetic memory at gene and LINE1 promters

Since H3K4me3 can directly prevent de novo DNA methylation the loss of H3K4me3 upon *Dppa2* KO could cause the observed gain of DNA methylation at the previous H3K4me3 occupied loci. In fact, analysis of all gene promoters and LINE1 promoters revealed a subpopulation negative correlation between progressive H3K4me3 loss and of DNAme gain upon *Dppa2* KO, demonstrating the negative correlation between H3K4me3 and DNA methylation (Figure 20A). Moreover, both depletion of promoter H3K4me3 and gain of DNA methylation are both significantly negatively correlated with gene expression (Figure 20B).

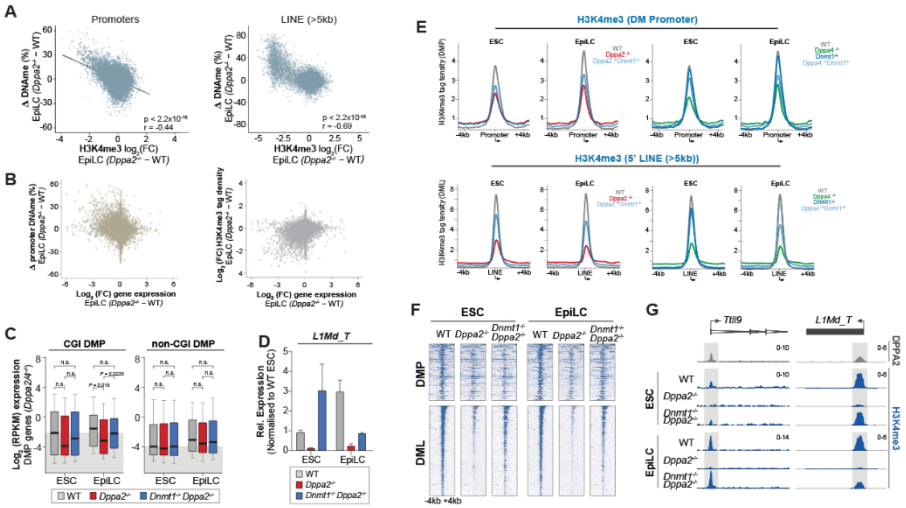


Figure 20. Functional interaction between DNA methylation, H3K4me3, and gene silencing.

A) Scatter plot showing correlated inter-relationship between changes in H3K4me3 and DNA methylation in *Dppa2* KO compared to WT in EpiLC at all gene promoters (left) and all full-length LINE1 (right). B) Scatter plot showing inter-relationships between changes in DNA methylation (Left) and H3K4me3 (Right) and gene expression in *Dppa2* KO compared to WT in EpiLC. **C)** Boxplot showing the expression of the genes associated with CGI or non-CGI differentially methylated promoters (DMPs) in WT, *Dppa2*$^{-/-}$ and the double KO; *Dnmt1*$^{-/-}$, *Dppa2*$^{-/-}$ cells. Box indicates the 25th, median and 75th percentiles, whiskers the 10th to the 90th percentiles, data from expression (RNAseq) of 125 (CGI DMP) and 159 (non-CGI DMP) genes. Significance by one-tailed students t-test. D) qRT-PCR quantifying expression of *L1Md_T* in WT, *Dppa2*$^{-/-}$ and double KO; *Dnmt1*$^{-/-}$, *Dppa2*$^{-/-}$ ESC and EpiLC, normalised to WT ESC. Data represent mean ± s.d. E) Trend plots showing the enrichment of H3K4me3 over gene promoters (upper panels) and over full-length LINE1 promoters (lower panels) that acquire DNAme in *Dppa2 KO*. Trends for WT, *Dppa2 KO* (or *Dppa4 KO*), *Dnmt1 KO* and *Dnmt1/Dppa2 KO* (*or Dppa4 KO*) in ESC and EpiLC are shown**.** F) Density plot showing enrichment of H3K4me3 centered on differentially methylated promoters (DMP) (upper) or LINEs (DML) (lower) +/− 4kb in WT, *Dppa2*$^{-/-}$ and compound-*Dnmt1*$^{-/-}$, *Dppa2*$^{-/-}$ ESC and EpiLC. G) Representative H3K4me3 CUT&RUN-seq genome tracks of a developmental promoter and LINE1 element of WT, *Dppa2*$^{-/-}$ and double KO; *Dnmt1*$^{-/-}$, *Dppa2*$^{-/-}$ ESC and EpiLC.

Finally, I wanted to determine if the established repressive epigenetic state in mutants, gain of DNA methylation and loss of H3K4me3, was instructive for gene silencing or a byproduct from lack of DPPA2. To study this, I deleted *Dnmt1* in *Dppa2*$^{-/-}$ to generate compound *Dnmt1* and *Dppa2* mutants that have completely lost DNA methylation. Analysis of the expression of the DMP geneset, that had aberrant methylation, revealed that loss of DNA methylation caused rescue of their activation in EpiLC (Figure 20C).

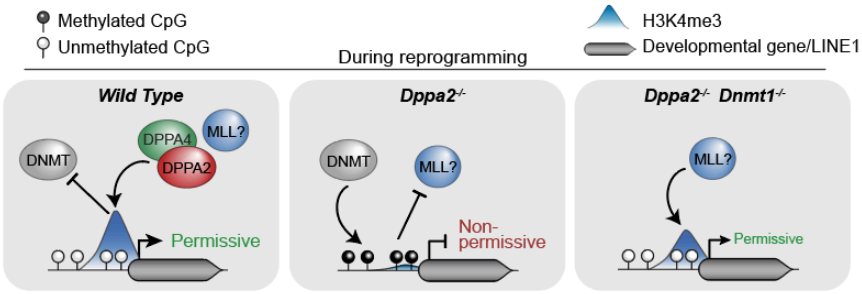


Figure 21. Proposed model of the role of DPPA2 and DPPA4 at Developmental genes and LINE1.

Proposed model of the relationship between H3K4me3 and DNA methylation at *Dppa2/4* targets in regulating expression competence at gene and LINE1 promoters.

This rescue was only significant at CpG dense promoters (CGI promoters). Moreover, a similar reactivation of L1Md_T was observed in the *Dnmt1* and *Dppa2* double mutants in both ESC and EpiLC (Figure 20D). Therefore, by removing DNA methylation a rescue of gene activation and L1 expression was seen at some genes and transposons indicating an instructive role for DNA methylation in maintaining their repression in *Dppa2*$^{-/-}$. Additionally, it demonstrates a role for DPPA2 in restricting DNA methylation, either by direct or indirect means, to maintain competence at developmental promoter and expression at LINE1s. As H3K4me3 can counteract DNA methylation and vice versa, I explored if the lack of DNA methylation also affected H3K4Me3. Using H3K4me3 Cut&Run-seq of the double *Dnmt1*, *Dppa2* mutant, a clear rescue of H3K4m3 was observed at a subset of promoters and most LINE1 elements (Figure 20E-G). These observations could possibly be explained by a model where absence of *Dppa2* or *Dppa4* leads to loss of H3K4me3 which in turn allows the de novo methylation mechanism to target the loci resulting in gain of DNA methylation and stable silencing. Removal of DNA methylation would then move the equilibrium towards H3K4me3 accumulation through DPPA2/4 independent mechanism (Figure 21).

5 Discussion

5.1 CRISPR-screen to identify regulators of DNA methylation

Using a CRISPR-Cas9 knock-out screen coupled with a real time single cell reporter of global DNA methylation, I have in addition to previously known regulators of DNA methylation, identified potential novel regulators of DNA methylation. Unlike other screens that have been performed to identify novel focal epigenetic regulators at transgenes or proviral elements (Blewitt et al., 2005; Fukuda et al., 2018; Yang et al., 2015) the readout used here is specially designed to be sensitive to global DNA methylation perturbations (Stelzer et al., 2015). Therefore, rather than candidates being validated post screening for their effect on DNA methylation, the aim here was to directly identify factors that regulate global DNA methylation. Indeed, the enhanced reporter of global methylation established here, proved to be practical to detect the global DNA methylation level in single cells as shown by the difference in GFP intensity in hypomethylated state (2i/LIF or DNMT1 KO) and hypermethylated state (titrated 2i/LIF and EpiLC). This system was used with CRISPR KO screening to identity both a) epigenetic silencers (spontaneous activation of eRGM in t2i/LIF) and b) reprogramming factors that are needed for hypomethylation (failure to activate eRGM in 2i/LIF).

5.1.1 DNA methylation maintenance candidates

The DNA methylation maintenance screen, designed to pick up factors with a role in maintaining DNA methylation and repression in mESC, identified *Dnmt1* and *Uhrf1*. *Dnmt1* and *Uhrf1* make up the DNA maintenance system and knock out of each gene results in near complete loss of DNA methylation due to failure to maintain methylated CpGs on both strands during DNA replication (E. Li et al., 1992b; Sharif et al., 2007). This demonstrates the strength of the screening approach in finding regulators of DNA methylation. mESC are undergoing constant de novo methylation with the expression and activity of *Dnmt3a* and *Dnmt3b* (T. Chen et al., 2003) and the methylation status of multiple foci depends to the balance between removal and addition of DNA methylation (Ginno et al., 2020). However, neither of the de novo methyltransferases nor the associated *Dnmt3l* were identified in the screen. This could be explained by redundancy of *Dnmt3a* and *Dnmt3b* in establishing DNA methylation at the reporter. Additionally, lack of gRNAs targeting *Dnmt3s* or other genes known to affect global DNA methylation in the selected cell populations could be a consequence of a saturated screen and therefore not

identifying weaker hits. Moreover, only selecting 1% of the population for sequencing resulted in a large fraction of gRNAs with zero reads similarly masking potential hits. Although not all the components of the DNA methylation machinery were identified, additional known epigenetic regulators were found. *Fam208a*, *Morc2a*, *Mpp8*, *Setdb1* and *Atf7ip* are all part of or associated with the HUSH complex which was recently discovered in a genomic screen utilizing a viral reporter (Robbez-Masson et al., 2018; Tchasovnikarova et al., 2015; Timms et al., 2016). The HUSH complex has been shown to be involved in heterochromatin spreading, resulting in elevated H3K9me3 and inducing epigenetic gene silencing at specific positions demonstrating position-effect variegation. This might indicate that the HUSH complex is regulating the DNA methylation on the reporter based on its integrated location in the genome. Moreover, multiple genes associated with non-canonical PRC1 (PRC1.6) complex were identified in the screen (*Max*, *Mga*, *L3mbtl2*). This complex has been suggested to be involved in maintaining silencing at CpG rich promoters of germ line genes like *Dazl* (Endoh et al., 2017; Stielow et al., 2018). In fact, knock out of the individual genes triggers derepression of germ cell genes in mESC (Maeda et al., 2013; Stielow et al., 2018). A recent model suggested PRC1.6 to be linked to DNA methylation at germ line promoters to ensure efficient silencing of germ line through H3K9me2 established by L3MBTL2 and GA9 interaction (Greenberg & Bourc'his, 2019).

In addition to known epigenetic regulators the screen also picked up subtler epigenetic candidates. One such hit, *Nudt21*, was recently reported to have an unexpected role in epigenetic regulation (Brumbaugh et al., 2018). *Nudt21* is an mRNA processing factor that regulates differential polyadenylation of chromatin regulators in cells acquiring pluripotency and disruption of *Nudt21* enhances reprogramming of somatic cells. *Nudt21* (*Cpsf5*) was identified in the epigenetic maintenance screen along with other genes related to mRNA regulation (*Cpsf6*, *Thoc3*, *Tra2b*), implying potentially related functions for these genes.

Although individual knock outs of almost all the candidates picked for validation proved to considerably affect the eRGM, KO of most candidate genes did not result in loss of global methylation as screened for. In fact, aside from *Dnmt1* only KO of factors targeting signaling pathways resulted in hypomethylation (*Socs3* and *Shoc2* (discussed below)). The failure to identify global regulators of DNA methylation could stem from the extensive epigenetic silencing pressure on the reporter through other means (PRC1.6, HUSH and (ATRX/DAXX chromatin remodeling), therefore saturating the screen with proteins that directly targeted the reporter.

5.1.2 DNA methylation reprogramming candidates

Likewise, I screened for factors with a role in DNA methylation reprogramming by selecting for cells that failed to reactivate eRGM during global DNA demethylation transition in mESC. Using this approach, known epigenome modulators, such as *Prdm14* (Leitch et al., 2013; Okashita et al., 2014; Yamaji et al., 2013) and its cofactor *Cbfa2t2* (Nady et al., 2015; Tu et al., 2016), as well as a cohort of novel candidates were identified. Similarly, to the epigenetic maintenance screen, failure to identify factors that have been indicated in safeguarding hypomethylated genome such as *Tet1*/*Tet2* and *Dppa3* (Mullholland et al., 2020) could stem from the saturation of enriched gRNAs. However, unlike the epigenetic silencer screen individual KO of most of the candidates resulted in gain of global methylation, highlighting the success of the approach to identify factors essential to maintain hypomethylation in 2i/LIF.

Furthermore, a large number of genes with roles in cell signaling were picked up by both screens supporting a major role for the signaling pathways in epigenetic regulation. Interestingly, these hits belonged to two main pathways, JAK/STAT LIF and MERK/ERK signaling. Moreover, the candidates are internally consistent between the two screens in their directionality in both pathways. Candidates for maintaining DNA methylation and repression are either activating in ERK/MERK (*Shoc2, Raf1, Ptpn11, Mapk1 and Grb2*) or antagonist to the LIF signaling pathway (*Socs3* and *Ptpn2*). On the opposite end, candidates in epigenetic reprograming screen attenuate MEK/ERK signaling (*Dusp6*) or activate the LIF signaling pathway (*Lifr*, *Jak1*, *Stat3*). Placing both pathways at the center of DNA methylation regulation in 2i/LIF ESC.

The MEK/ERK pathway, activated by FGF, directs ESC towards differentiation by activation of lineage specifying genes (Kunath et al., 2007), therefore repressing the pathway is essential to maintain naïve pluripotency (Hackett & Surani, 2014; Silva & Smith, 2008b). The role of the MERK/ERK pathway in pluripotency regulation and epigenetic regulation has been well established as the MERK inhibitor, PD promotes long term self-renewal and naïve pluripotency in ESC (Ying et al., 2008). Reducing the concentration of PD and therefore increasing ERK signaling causes increase in global DNA methylation (Choi, Huebner, et al., 2017; Yagi et al., 2017) and is demonstrated here by the ability of the eRGM to separate the titrated 2i/LIF and the full 2i/LIF conditions with hyper- and hypomethylated genome respectively. Predictably, multiple positive and negative regulators of MERK/ERK were identified in the two screens. One such candidate (*Dusp6*) was validated in this study by generating clonal KO mESC lines which resulted in globally hypermethylated DNA in 2i/Lif. Indeed, X-linked DUSP9, another dual phosphatase with role in

repressing MEK/ERK signaling has been shown to contribute to female-specific ESC hypomethylation (Choi, Clement, et al., 2017), indicating towards a comparable but non-redundant role of DUSP6 in thresholding MERK/ERK signaling more generally in pluripotent cells to promote epigenome erasure. Additionally, I also demonstrated the potential mechanism through which DUSP6 influences the methylome. *Dusp6* knock out amplifies ERK signaling which increases expression of the de novo methyltransferases *Dnmt3a/b/l* and eventually results in DNA hypermethylation. Moreover, KO of the *Shoc2* activator of ERK signaling caused loss of global DNA methylation. Interestingly *Med24* (rank=4) was recently found to be an effector of MEK/ERK signaling (Hamilton et al., 2019). Therefore, other candidates currently not directly linked to the signaling pathways might be effectors or regulators of MERK/ERK signaling.

Here the hits involved in the LIF pathway in both screens and their effect on global DNA methylation were validated. *Socs3*, a repressor of LIF signaling (R. F. Wu et al., 2019) was identified in the epigenetic maintenance screen and perturbation of *Socs3* caused loss of global DNA methylation. Conversely, loss of *Lifr* and *Jak1* resulted in increase in DNA methylation validating the role of LIF signaling on global DNA methylation. The influence of LIF on DNA methylation has been suggested before (Habibi et al., 2013). A study comparing the DNA methylation status of different ESC cultures demonstrated significantly elevated DNA methylation in 2i ESC cultured without LIF (Hackett et al., 2017). Thus, KO of the LIF pathway related candidates from the screens matched a previously observed phenotype in mESC, although how LIF signaling influences DNA methylation is unknown. LIF signaling ensures the activation of the core pluripotency circuitry by activating *Sox2* through KLF4 and *Nanog* through PI3K-Akt signaling and TBX3 (Niwa et al., 2009) and absence of LIF results in differentiation in Serum/LIF culture (Ying et al., 2003). Therefore, LIF signaling is linked to pluripotency and DNA hypomethylation.

Another cluster of hits was related to the methylation of RNA and an associated reader of RNA methylation. m^6A modification of RNA is put down by the WMM complex and can be recognized by m^6A binding reader proteins (YTHDF1-3) (Zaccara et al., 2019). Interestingly, when screening for reprogramming factors almost all known components of the WMM complex were identified (*Mettl3, Mettl14, Virma, Wtap, Zc3h13 and Cbll1*) (J. Liu et al., 2014; Wen et al., 2018). m^6A has been linked to mRNA splicing, localization, translation and degradation. More specifically, loss of m6A in ESC in ground state naïve conditions (2i/LIF) results in failure to exit pluripotency as pluripotent genes are not downregulated during differentiation and $Mettl3^{-/-}$ and $Mettl14^{-/-}$ have been described to maintain a hyper naïve pluripotency in meSC

(Geula et al., 2015). Conversely, KO of *Virma* resulted in a global increase in DNA methylation, generally linked to more advanced pluripotency states (Hackett et al., 2017; Leitch et al., 2013). Interestingly, METTL3 and YTHDF2 have been linked to the regulation of the mRNA of *Socs3* (R. Wu et al., 2019), a candidate factor from the epigenetic silencer screen and a negative regulator of LIF signaling. *Socs3* mRNA is degraded through m^6A modification resulting in amplified LIF signaling. This links m^6A to LIF signaling and potentially explaining the gain of DNA methylation observed upon loss of m^6A (*Virma* KO) and LIF signaling (*Jak1* and *Lifr* knock outs) and the opposite loss of DNA methylation upon activated LIF signaling (*Socs3* KO). Additionally, the CCR4-NOT has been shown to deadenylate mRNA after being recruited by m6A bound YTDHF2 (H. Du et al., 2016), demonstrating a link for the CCR4-NOT cluster identified in the epigenetic maintenance screen (*Cnot10* and *Rqcd1* (*Cnot9*)) to DNA methylation through m^6A regulation.

5.2 Functional analysis of DPPA2 and DPPA4

In my system, the eRGM contains *Dazl* promoter in an antisense orientation as a sensor for global DNA methylation. *Dazl* is primarily regulated by promoter DNA methylation which correlates well with global DNA methylation levels (Hackett et al., 2012). Thus, the methylation status of the *Dazl* promoter functions well as a sensor for global DNA methylation. Moreover, Dazl activation converges with pluripotency and it functions as a marker of complete epigenetic resetting in primordial germ cell development making it a relevant gene to study epigenetic reprogramming. However, one of the instructive biases of such system are factors that directly affect the sensor without altering the global environment under investigation. The drawbacks of the system have already been demonstrated in the epigenetic maintenance screen as multiple direct repressors were picked up in the screen. However, two of the top hits from our reprogramming screen, *Dppa2* and *Dppa4* showed more interesting behaviour. Because their perturbations did not cause increase in global DNA methylation, but rather focal gain of DNA methylation. Using EMseq (Vaisvila et al. 2019) to study the global DNA methylation at a basepair resolution, the regions that gained DNA methylation upon loss of *Dppa2* or *Dppa4* were mostly confined to a specific set of gene promoters, genebodies and young LINE1 transposons.

Interestingly, although DPPA2 displayed a genome wide binding profile, interacting with over 8000 promoters only a fraction of DPPA2 bound promoters displayed an aberrant epigenetic profile. This indicates that the vast majority of DPPA2 bound genes were in fact independent of DPPA2. Differences in chromatin landscape, transcription factors bound and transcriptional status of the DPPA2 bound promoters could rationalise the lack

of epigenetic dynamics upon *Dppa2/4* perturbation at these genes. Moreover, portion of the differentially expressed genes upon *Dppa2/4* KO do not establish any epigenetically perturbations and are not bound by DPPA2 indicating an indirect effect on their expression. This is, for example, the case for the Zscan gene cluster and MERVL transposons as the effect on their expression is a downstream effect of *Dux* downregulation and *Dppa2* KO does not affect them directly (De Iaco et al., 2019; M. Eckersley-Maslin et al., 2019). The promoters with disturbed epigenetic landscape upon loss of *Dppa2/4*, elevated DNA methylation and loss of H3K4me3 were predominantly regions with intermediate level of H3K4me3 and upstream of developmental genes with low expression in mESC. Although DPPA2 has a preference to bind highly expressed genes, the genes most sensitive to epigenetic perturbations upon *Dppa2/4* KO are developmental genes. These genes have low to no expression in early development and lack of *Dppa2/4* caused these genes to fail to be upregulated upon differentiation and their promoters were still methylated in differentiated cells. This potentially signifies a role for DPPA2 and DPPA4 in maintaining the permissive epigenetic state and guarding against ectopic de novo methylation activity at a specific set of promoters with low expression during pluripotency. Failure to maintain this permissive state causes aberrant epigenetic silencing at later stages during development as seen here during endodermal differentiation. In fact, *Dppa2* or *Dppa4* mutant mice develop normally but die perinatally due to failure to upregulate genes in lungs, where *Dppa2/4* are not expressed (Nakamura et al., 2011). DPPA2 and DPPA4 therefore maintain the competence of lineage associated genes (such as *Hand1*, *Gnmt*, *Cdcp1*) throughout early development for future activation.

As *Dppa2/4* are essential to maintain permissive epigenetic status of a set of gene and LINE1 promoters, what is the role of *Dppa2/4* in epigenetic regulation? Eckersley-maslin et al. demonstrated that DPPA2 and DPPA4 interact with the trxG complex, indicating a role for DPPA2/4 in establishing stable H3K4me3 signal (M. A. Eckersley-Maslin et al., 2020). Moreover, DPPA2 binding sites are highly enriched for H3K4me3 irrespective of the underlying transcriptional activity indicating that DPPA2 appears to directly target H3K4m3. H3K4me3 has previously been shown to restrict the recruitment of the de novo DNA methyltransferases (X. Guo et al., 2015; Ooi et al., 2007; Y. Zhang et al., 2010) enforcing the observed mutual exclusion of both marks. Therefore, the loss of H3K4me3 following *Dppa2/4* KO could enable the recruitment of the *de novo* DNA methyltransferases and ectopic gain of DNA methylation. By comparing the change in H3K4me3 and DNA methylation upon *Dppa2* KO, I observed a great anti correlation between the two epigenetic marks; the greater the loss of H3K4me3 at gene promoters or young LINE1 promoters the more gain of DNA methylation was observed.

However, loss of H3K4me3 does not lead to instinctive gain of DNA methylation. Loss of H3K4me3 at a subset of gene promoters by *Cpf1* (*Cxxc1*) depletion did not lead to aberrant DNA methylation (Clouaire et al., 2012), Indicating that the observed relationship between H3K4me3 and DNA methylation is context dependent. In fact, DNMT3A engineered to tolerate H3K4me3 resulted in aberrant gain of DNA methylation, specifically at developmental genes (Noh et al., 2015). This supports a model were DPPA2/4 dependent H3K4me3 protects against gain of DNA methylation that could lead to subsequent stable silencing at specific loci. Surprisingly however, removal of all (and ectopic) DNA methylation by *Dnmt1* knockout in *Dppa2*$^{-/-}$ cells partially restored both H3K4me3 and expression defects at genes and LINE1. This indicates an instructive role of DNA methylation in silencing and the importance of *Dppa2/4* to guard against de novo methylation. Moreover, it suggests that DNA methylation reciprocally limits H3K4me3 recruitment and without DNA methylation H3K4me3 can be recruited to promoters independently of DPPA2. Likewise, DNA methylation is anti-correlated with H3K27me3 as DNA methylation can counteract PRC2 and H3K27me3 recruitment (Atlasi & Stunnenberg, 2017; Reddington et al., 2013). The ectopic gain of DNA methylation at bivalent genes could therefore explain the slight reduction of H3K27me3 observed at some developmental genes (*Tnxb* etc). Overall this implies a balance between H3K4me3 and DNA methylation that is needed to maintain permissive epigenetic state at developmental genes and LINE1 elements regulated by DPPA2/4.

Remarkably, only full length and evolutionarily young LINE1 elements were directly bound by DPPA2 and in the same fashion as developmental genes relied upon DPPA2 and DPPA4 for their activity. KO of *Dppa2* or *Dppa4* resulted in loss of various TE namely young L1 elements and MERVL elements. However, DPPA2 does not bind MERVL or influence its epigenetic state but instead, as has been shown before (De Iaco et al., 2019; M. Eckersley-Maslin et al., 2019), this is a consequence of DPPA2 regulated *Dux* expression which in turns regulates MERVL (De Iaco et al., 2017b; Hendrickson et al., 2017a). DPPA2 directly occupies the promoters of full-length LINE1 elements and maintains H3K4me3, hypomethylation and expression. Therefore, this mirrors the epigenetic regulation observed at developmental genes implying an evolutionarily hi-jacking by LINE1 to counteract host directed epigenetic repression. This leads to the expression of LINE1 transposons when *Dppa2* and *Dppa4* are expressed which is applicable as the expression pattern of *Dppa2* and *Dppa4* aligns well the optimal period of transposon expression in early development (Kano et al., 2009). This presents a genomic conflict as *Dppa2/4* are essential for ensuring epigenetic competence of developmental genes but at the same time allow for the

expression of full length LINE1, therefore potentially threatening genomic integrity through retrotransposition events. However, recent evidence has demonstrated an altruistic role of LINE1 expression in early development (Jachowicz et al., 2017; Percharde et al., 2018). More specifically, LINE1 upregulation is essential during development and for repression of Dux and the 2-cell state (Percharde et al., 2018). KD of LINE1 leads to persistent 2C stage in ESC and developmental failure beyond the 2-cell stage in embryos. Paradoxically, Dux is upregulated by DPPA2 (De Iaco et al., 2019; M. Eckersley-Maslin et al., 2019) hinting towards a precise control of the 2-Cell stage by DPPA2; through direct *Dux* upregulation followed by LINE1 mediated silencing of *Dux* to exit the 2 C state. Although *Dux* is not essential for early development, embryos without *Dux* (Z. Chen & Zhang, 2019) and *Dppa2/4* KO embryos develop normally, indicating LINE1 mediated repression of activated *Dux* might be critical for early development.

5.3 Future directions

Here a germ line gene (*Dazl*), that is generally known to be sensitive to DNA methylation (Hackett et al., 2012), was used in anti-sense ordination as a sensor for eRGM to track the global DNA methylation level in single cells. As initially tried here using IAP as a sensor, further optimization of the system could be done by using a different kind of methylation sensor. Demonstrated here by the *Dppa2/4* story, such system could give insight into the focal DNA methylation regulation at a specific genomic feature used as a sensor such as exons (active or inactive genes), transposons (IAP, MERVL, LINE1 etc.) or reprogramming escapees (Hackett, Sengupta, et al., 2013).

BRD4 is a histone acetylation reader and a transcriptional co-activator and has a key role in mESC and in embryonic development (Houzelstein et al., 2002). Recent study showed that *Brd4* is essential for self-renewal in Serum/LIF mESC through co-activation at loci bound by the core pluripotency TFs (OCT4, NANOG and SOX2) (Finley et al., 2018). However, in naïve mESC, loss of *Brd4* is sustainable as long as TET1 or TET2 activity is maintained. As shown here, loss of *Brd4* in 2i/LIF ESC triggers hypermethylation, suggesting an essential role of *Tet1/2* in maintaining focal hypomethylation at promoters for co-activation of genes acting at BRD4 bound loci or in directly maintaining hypomethylation at the same loci. In fact, TET1 and BRD4 were shown to be recruited to overlapping binding sites (Finley et al., 2018). Although using Vitamin C to boost the TET1/2 activity in Serum/LIF was not enough to rescue the cells from *Brd4* dependency, this could be because of a possible formation of phase separated condensates in 2i/LIF cells at BRD4 bound pluripotency loci (M. Zhang et al., 2020). In 2i/LIF ESC the GSK3 inhibitor causes B-catenin stabilization and eventual recruitment of B-

catenin to pluripotency loci to form phase separated condensates with cohesin and the mediator complex. This induces expression by increasing DNA polymerase II initiation, therefore rescuing the cells from *Brd4* dependency (M. Zhang et al., 2020). Overall, this could implicate a role for TET1/2 in maintaining the formation of phase separation in 2i/LIF mESC caused by the increased global DNA methylation upon *Brd4* depletion. How loss of *Brd4* results in hypermethylation is still unknown, but it could be a secondary effect of the reduced strength of the pluripotency TF binding network. Further research into the interplay between BRD4 and DNA methylation and TET1/2 could give insight into the co-activation function of BRD4 at pluripotency loci.

Here, the ability of DPPA2/4 to maintain local specific hypomethylation and H3K4 trimethylation at LINE1 and developmental gene promoters was demonstrated. But how this is achieved is not completely understood, although one hypothesis is demonstrated here (Figure 21). DPPA2/4 interact with KMT2B (M. A. Eckersley-Maslin et al., 2020) which is responsible for maintaining H3K4me3 and active transcription at a number of developmental and bivalent genes (D. Hu et al., 2013). However, removal of MLL2 resulted in gain of DNA methylation at only a subset of MLL2 dependent genes (~40) (Douillet et al., 2020) possibly indicating that DPPA2/4 counter DNA methylation not only through H3K4me3 maintenance. Understanding the molecular mechanisms that are affected upon loss of H3K4me3 as well as the changes that allow transcription upon removal of DNA methylation remains an intriguing question; what changes in transcription factor binding and chromatin are happening during this switch? And are they the same for both developmental genes and LINE1? Exploring this further could also reveal why only a small subset of genes were affected by lack *of Dppa2/4*, although DPPA2 showed a wide gene promoter occupancy.

Dppa2 and *Dppa4* expression follows epigenetic reprogramming in early development and, as demonstrated here, play a role in ensuring epigenetic competence at developmental genes during reprogramming. Interestingly, *Dppa2/4* are upregulated again in PGCs during another wave of epigenetic reprogramming and extensive loss of global DNA methylation (Seisenberger et al., 2012). The role of Dppa2/4 during PGC maturation has not been studied and a lot of questions are yet to be answered. More specifically, the consequences of *Dppa2/4* perturbation on PGC development are unknown. A system to induce *Dppa2/4* removal at the induction of PGCs in vivo or during PGCLC differentiation would be needed to isolate the specific function of Dppa2/4 at this stage. DPPA2 binds germ line genes and could be essential to ensure their expression during PGC development. But the role of DPPA2/4 at developmental genes or LINE1 transposons is less obvious during PGC

development; these developmental genes are not expressed in that period, and extensive epigenetic reprogramming takes place during germ cell maturation resetting any permissive environment established by DPPA2/4.

6 Conclusions

Here, a powerful single cell reporter of global DNA methylation was set up to identify regulators of DNA methylation during epigenetic reprogramming, modelled in mESC, by means of unbiased CRISPR-screening. Using this approach, multiple potential regulators of DNA methylation in mESC were identified. These candidates might be exciting to follow up on to further explore their role in DNA methylation or epigenetic regulation during early development.

Furthermore, analysis of the role of *Dppa2* and *Dppa4* in regulating DNA methylation revealed their essential role in maintaining a focal permissive epigenetic state during pluripotency by preventing abnormal gain of DNA methylation and loss of H3K4me3 at developmental promoters. Indeed, loss of either gene in pluripotent cells resulted in stable repressive epigenetic silencing at DPPA2 bound promoters in differentiated cells that did not express either *Dppa2* or *Dppa4*, indicating an epigenetic memory of the repressive state. Therefore, by maintaining the permissive epigenetic state *Dppa2* and *Dppa4* contribute to the developmental competence of pluripotent cells. Importantly, *Dppa2* and *Dppa4* also establish a permissive epigenetic state at evolutionary young LINE1 elements indicating co-adaptation of the TEs to enable their expression during pluripotency.

Overall, this thesis adds to our knowledge about the mechanisms at play during early embryo epigenetic reprogramming, providing an excellent example of the role of focal regulators in ensuring developmental competency.

References

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N. R., Koseki, H., Sasaki, H., & Hajkova, P. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nature Cell Biology, 18*(2), 225-+

Andreu-Vieyra, C. V., Chen, R. H., Agno, J. E., Glaser, S., Anastassiadis, K., Stewart, A. F., & Matzuk, M. M. (2010). MLL2 Is Required in Oocytes for Bulk Histone 3 Lysine 4 Trimethylation and Transcriptional Silencing. *Plos Biology, 8*(8)

Aravin, A. A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science, 318*(5851), 761-764

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., & Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell, 31*(6), 785-799

Aravind, L., & Koonin, E. V. (2000). SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends in Biochemical Sciences, 25*(3), 112-114

Atlasi, Y., & Stunnenberg, H. G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. *Nature Reviews Genetics, 18*(11)

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science, 354*(6314), 909-912

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A., & Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature, 520*(7546), 243-247

Baubec, T., Ivanek, R., Lienert, F., & Schubeler, D. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell, 153*(2), 480-492

Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E., & Spadafora, C. (2006). Expression of LINE-1 retroposons is essential for murine preimplantation development. *Molecular Reproduction and Development, 73*(3), 279-287

Berns, K., Hijmans, E. M., Mullenders, J., Brummelkamp, T. R., Velds, A., Heimerikx, M., Kerkhoven, R. M., Madiredjo, M., Nijkamp, W., Weigelt, B., Agami, R., Ge, W., Cavet, G., Linsley, P. S., Beijersbergen, R. L., & Bernards, R. (2004). A large-scale RNAi

screen in human cells identifies new components of the p53 pathway. *Nature, 428*(6981), 431-437
Bernstein, B. E., Mikkelsen, T. S., Xie, X. H., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*(2), 315-326
Berrens, R. V., Andrews, S., Spensberger, D., Santos, F., Dean, W., Gould, P., Sharif, J., Olova, N., Chandra, T., Koseki, H., von Meyenn, F., & Reik, W. (2017). An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell, 21*(5), 694-+
Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H., & Klose, R. J. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell, 157*(6), 1445-1459
Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martinez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., & Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature, 500*(7461), 222-+
Blewitt, M. E., Vickaryous, N. K., Hemley, S. J., Ashe, A., Bruxner, T. J., Preis, J. I., Arkell, R., & Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A, 102*(21), 7629-7634
Borowiak, M., Maehr, R., Chen, S. B., Chen, A. E., Tang, W. P., Fox, J. L., Schreiber, S. L., & Melton, D. A. (2009). Small Molecules Efficiently Direct Endodermal Differentiation of Mouse and Human Embryonic Stem Cells. *Cell Stem Cell, 4*(4), 348-358
Bortvin, A., Eggan, K., Skaletsky, H., Akutsu, H., Berry, D. L., Yanagimachi, R., Page, D. C., & Jaenisch, R. (2003). Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development, 130*(8), 1673-1680
Bortvin, A., Goodheart, M., Liao, M., & Page, D. C. (2004). Dppa3 / Pgc7 / stella is a maternal factor and is not required for germ cell specification in mice. *BMC Dev Biol, 4*, 2
Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science, 317*(5845), 1760-1764
Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science, 294*(5551), 2536-2539
Boutros, M., & Ahringer, J. (2008). The art and design of genetic screens: RNA interference. *Nature Reviews Genetics, 9*(7), 554-566

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., & Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature, 441*(7091), 349-353

Boyes, J., & Bird, A. (1992). Repression of Genes by DNA Methylation Depends on Cpg Density and Promoter Strength - Evidence for Involvement of a Methyl-Cpg Binding-Protein. *Embo Journal, 11*(1), 327-333

Brinkman, A. B., Gu, H. C., Bartels, S. J. J., Zhang, Y. Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., & Stunnenberg, H. G. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Research, 22*(6), 1128-1138

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research, 42*(22)

Broske, A. M., Vockentanz, L., Kharazi, S., Huska, M. R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., Nerlov, C., Leutz, A., Andrade-Navarro, M. A., Jacobsen, S. E. W., & Rosenbauer, F. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature Genetics, 41*(11), 1207-1215

Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K. J., Ji, F., Schwarz, B. A., Kalocsay, M., Elledge, S. J., Chen, Y., Sadreyev, R. I., Gygi, S. P., Hu, G., Shi, Y., & Hochedlinger, K. (2018). Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. *Cell, 172*(1-2), 106-120 e121

Burton, A., & Torres-Padilla, M. E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol, 15*(11), 723-734

Chamberlain, S. J., Yee, D., & Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells, 26*(6), 1496-1505

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., & Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol, 23*(16), 5594-5605

Chen, Z., & Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat Genet, 51*(6), 947-951

Chiang, A., Oconnor, M. B., Paro, R., Simon, J., & Bender, W. (1995). Discrete Polycomb-Binding Sites in Each Parasegmental Domain of the Bithorax Complex. *Development, 121*(6), 1681-1689

Choi, J., Clement, K., Huebner, A. J., Webster, J., Rose, C. M., Brumbaugh, J., Walsh, R. M., Lee, S., Savol, A., Etchegaray, J. P., Gu, H., Boyle, P., Elling, U., Mostoslavsky, R., Sadreyev, R., Park, P. J., Gygi, S. P., Meissner, A., & Hochedlinger, K. (2017). DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells. *Cell Stem Cell, 20*(5), 706-719 e707

Choi, J., Huebner, A. J., Clement, K., Walsh, R. M., Savol, A., Lin, K. X., Gu, H. C., Di Stefano, B., Brumbaugh, J., Kim, S. Y., Sharif, J., Rose, C. M., Mohammad, A., Odajima, J., Charron, J., Shioda, T., Gnirke, A., Gygi, S., Koseki, H., Adreyev, R. I. S., Xiao, A., Meissner, A., & Hochedlinger, K. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature, 548*(7666), 219-+

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J. H., Skalnik, D., & Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & Development, 26*(15), 1714-1728

Dahl, J. A., Jung, I., Aanes, H., Greggains, G. D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A. Y., Preissl, S., Jermstad, I., Haugen, M. H., Suganthan, R., Bjoras, M., Hansen, K., Dalen, K. T., Fedorcsak, P., Ren, B., & Klungland, A. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature, 537*(7621), 548-552

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., Powell, B. E., Cheng, A. W., Faull, K. F., Lyko, F., & Jaenisch, R. (2014). Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. *Developmental Cell, 29*(1), 102-111

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., Gao, Q., Powell, B. E., Li, Z., Xu, M. J., Faull, K. F., Lyko, F., & Jaenisch, R. (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell, 24*(3), 310-323

Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., Gao, Q., Kim, J., Choi, S. W., Page, D. C., & Jaenisch, R. (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. *Cell Stem Cell, 9*(2), 166-175

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 Reports, 20*(5)

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017a). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics, 49*(6), 941-+

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017b). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet, 49*(6), 941-945

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H. G., & Stewart, A. F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development, 141*(3), 526-537

Deniz, O., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications (vol 20, pg 417, 2019). *Nature Reviews Genetics, 20*(7), 432-432

Dennis, K., Fan, T., Geiman, T., Yan, Q. S., & Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes & Development, 15*(22), 2940-2944

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem, 285*(34), 26114-26120

Di Croce, L., & Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nature Structural & Molecular Biology, 20*(10), 1147-1155

Doench, J. G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics, 19*(2), 67-80

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016a). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol, 34*(2), 184-191

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016b). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology, 34*(2), 184-+

Dorighi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., Still, C. D., Garcia, B. A., Adelman, K., & Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell, 66*(4), 568-+

Dornan, D., Wertz, I., Shimizu, H., Arnott, D., Frantz, G. D., Dowd, P., O'Rourke, K., Koeppen, H., & Dixit, V. M. (2004). The ubiquitin ligase COP1 is a critical negative regulator of p53. *Nature, 429*(6987), 86-92

Douillet, D., Sze, C. C., Ryan, C., Piunti, A., Shah, A. P., Ugarenko, M., Marshall, S. A., Rendleman, E. J., Zha, D. D., Helmin, K. A., Zhao, Z. B., Cao, K. X., Morgan, M. A., Singer, B. D., Bartom, E. T., Smith, E. R., & Shilatifard, A. (2020). Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nature Genetics, 52*(6), 615-+

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., & Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat Commun, 7*, 12626

Du, J., Chen, T. J., Zou, X., Xiong, B., & Lu, G. X. (2010). Dppa2 knockdown-induced differentiation and repressed proliferation of mouse embryonic stem cells. *Journal of Biochemistry, 147*(2), 265-271

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 & Development, 33*(3-4), 194-208

Eckersley-Maslin, M. A., Parry, A., Blotenburg, M., Krueger, C., Ito, Y., Franklin, V. N. R., Narita, M., D'Santos, C. S., & Reik, W. (2020). Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. *Nature Structural & Molecular Biology, 27*(8), 696-705

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature, 411*(6836), 494-498

Endoh, M., Endo, T. A., Shinga, J., Hayashi, K., Farcas, A., Ma, K. W., Ito, S., Sharif, J., Endoh, T., Onaga, N., Nakayama, M., Ishikura, T., Masui, O., Kessler, B. M., Suda, T., Ohara, O., Okuda, A., Klose, R., & Koseki, H. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *Elife, 6*

Engelen, E., Brandsma, J. H., Moen, M. J., Signorile, L., Dekkers, D. H. W., Demmers, J., Kockx, C. E. M., Ozgur, Z., van IJcken, W. F. J., van den Berg, D. L. C., & Poot, R. A. (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nature Communications, 6*

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., & Torres-Padilla, M. E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nature Structural & Molecular Biology, 20*(3), 332-338

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., Schroder, K., Cloonan, N., Steptoe, A. L., Lassmann, T., Waki, K., Hornig, N., Arakawa, T., Takahashi, H., Kawai, J., Forrest, A. R. R., Suzuki, H., Hayashizaki, Y., Hume, D. A., Orlando, V., Grimmond, S. M., & Carninci, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nature Genetics, 41*(5), 563-571

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm (vol 12, pg 565, 2011). *Nature Reviews Genetics, 12*(9)

Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y. L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R., & Reik, W. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell, 13*(3), 351-359

Finley, L. W. S., Vardhana, S. A., Carey, B. W., Alonso-Curbelo, D., Koche, R., Chen, Y., Wen, D., King, B., Radler, M. R., Rafii, S., Lowe, S. W., Allis, C. D., & Thompson, C. B. (2018). Pluripotency transcription factors and Tet1/2 maintain Brd4-independent stem cell identity. *Nat Cell Biol, 20*(5), 565-574

Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Diaz, N., Rowe, H. M., Ecco, G., Unzu, C., Planet, E., Lombardo, A., Mangeat, B., Wildhaber, B. E., Naldini, L., & Trono, D. (2014). Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Research, 24*(8), 1251-1259

Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A., & Dreyfuss, G. (2001). SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Molecular Cell, 7*(5), 1111-1117

Fuks, F., Burgers, W. A., Godin, N., Kasai, M., & Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *Embo Journal, 20*(10), 2536-2544

Fukuda, K., Okuda, A., Yusa, K., & Shinkai, Y. (2018). A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res, 28*(6), 846-858

Funaki, S., Nakamura, T., Nakatani, T., Umehara, H., Nakashima, H., & Nakano, T. (2014). Inhibition of maintenance DNA methylation by Stella. *Biochemical and Biophysical Research Communications, 453*(3), 455-460

Galli, U. M., Sauter, M., Lecher, B., Maurer, S., Herbst, H., Roemer, K., & Mueller-Lantzsch, N. (2005). Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. *Oncogene, 24*(19), 3223-3228

Galonska, C., Ziller, M. J., Karnik, R., & Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell, 17*(4), 462-470

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y. S., Ben-Haim, M. S., Eyal, E., Yunger, S., Pinto, Y., Jaitin, D. A., Viukov, S., Rais, Y., Krupalnik, V., Chomsky, E., Zerbib, M., Maza, I., Rechavi, Y., Massarwa, R., Hanna, S., Amit, I., Levanon, E. Y., Amariglio, N., Stern-Ginossar, N., Novershtern, N., Rechavi, G., & Hanna, J. H. (2015). Stem cells. m6A mRNA methylation facilitates

resolution of naive pluripotency toward differentiation. *Science, 347*(6225), 1002-1006

Ginno, P. A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A., Edenhofer, F., Smallwood, S. A., Krebs, A. R., & Schubeler, D. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat Commun, 11*(1), 2680

Glaser, S., Lubitz, S., Loveland, K. L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K., & Stewart, A. F. (2009). The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics & Chromatin, 2*

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K. R., Aasland, R., Anastassiadis, K., Ang, S. L., & Stewart, A. F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development, 133*(8), 1423-1432

Gokbuget, D., & Blelloch, R. (2019). Epigenetic control of transcriptional regulation in pluripotency and early differentiation. *Development, 146*(19)

Grabole, N., Tischler, J., Hackett, J. A., Kim, S., Tang, F. C., Leitch, H. G., Magnusdottir, E., & Surani, M. A. (2013). Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *Embo Reports, 14*(7), 629-637

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol, 20*(10), 590-607

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L. Y., He, X. Y., Jin, S. G., Iqbal, K., Shi, Y. J. G., Deng, Z. X., Szabo, P. E., Pfeifer, G. P., Li, J. S., & Xu, G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature, 477*(7366), 606-U136

Guo, F., Li, X. L., Liang, D., Li, T., Zhu, P., Guo, H. S., Wu, X. L., Wen, L., Gu, T. P., Hu, B. Q., Walsh, C. P., Li, J. S., Tang, F. C., & Xu, G. L. (2014). Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell, 15*(4), 447-458

Guo, X., Wang, L., Li, J., Ding, Z. Y., Xiao, J. X., Yin, X. T., He, S., Shi, P., Dong, L. P., Li, G. H., Tian, C. L., Wang, J. W., Cong, Y., & Xu, Y. H. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature, 517*(7536), 640-U281

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., Benedetti, R., Altucci, L., Jansen, J. H., Walter, J., Gut, I. G., Marks, H., & Stunnenberg, H. G. (2013). Whole-Genome Bisulfite

Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell, 13*(3), 360-369
Hackett, J. A., Dietmann, S., Murakami, K., Down, T. A., Leitch, H. G., & Surani, M. A. (2013). Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency. *Stem Cell Reports, 1*(6), 518-531
Hackett, J. A., Huang, Y., Gunesdogan, U., Gretarsson, K. A., Kobayashi, T., & Surani, M. A. (2018). Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat Commun, 9*(1), 4292
Hackett, J. A., Kobayashi, T., Dietmann, S., & Surani, M. A. (2017). Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. *Stem Cell Reports, 8*(6), 1645-1658
Hackett, J. A., Reddington, J. P., Nestor, C. E., Dunican, D. S., Branco, M. R., Reichmann, J., Reik, W., Surani, M. A., Adams, I. R., & Meehan, R. R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development, 139*(19), 3623-3632
Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., & Surani, M. A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science, 339*(6118), 448-452
Hackett, J. A., & Surani, M. A. (2014). Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell, 15*(4), 416-430
Hamilton, W. B., Mosesson, Y., Monteiro, R. S., Emdal, K. B., Knudsen, T. E., Francavilla, C., Barkai, N., Olsen, J. V., & Brickman, J. M. (2019). Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature, 575*(7782), 355-360
Hashimoto, H., Liu, Y. W., Upadhyay, A. K., Chang, Y. Q., Howerton, S. B., Vertino, P. M., Zhang, X., & Cheng, X. D. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research, 40*(11), 4841-4849
Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L. E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J. A., Ecker, J. R., & Ren, B. (2010). Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. *Cell Stem Cell, 6*(5), 479-491
Hayashi, K., de Sousa Lopes, S. M. C., Tang, F., Lao, K., & Surani, M. A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell, 3*(4), 391-401
He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q. Y., Ding, J. P., Jia, Y. Y., Chen, Z. C., Li, L., Sun, Y., Li, X. X., Dai, Q., Song, C. X., Zhang, K. L., He, C., & Xu, G. L. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science, 333*(6047), 1303-1307

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y. T., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C. X., Ching, K. A., Wang, W., Weng, Z. P., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics, 39*(3), 311-318

Hendrich, B., & Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology, 18*(11), 6538-6547

Hendrickson, P. G., Dorais, J. A., Grow, E. J., Whiddon, J. L., Lim, J. W., Wike, C. L., Weaver, B. D., Pflueger, C., Emery, B. R., Wilcox, A. L., Nix, D. A., Peterson, C. M., Tapscott, S. J., Carrell, D. T., & Cairns, B. R. (2017a). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet, 49*(6), 925-934

Hendrickson, P. G., Dorais, J. A., Grow, E. J., Whiddon, J. L., Lim, J. W., Wike, C. L., Weaver, B. D., Pflueger, C., Emery, B. R., Wilcox, A. L., Nix, D. A., Peterson, C. M., Tapscott, S. J., Carrell, D. T., & Cairns, B. R. (2017b). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nature Genetics, 49*(6), 925-+

Hernandez, C., Wang, Z., Ramazanov, B., Tang, Y., Mehta, S., Dambrot, C., Lee, Y. W., Tessema, K., Kumar, I., Astudillo, M., Neubert, T. A., Guo, S., & Ivanova, N. B. (2018). Dppa2/4 Facilitate Epigenetic Remodeling during Reprogramming to Pluripotency. *Cell Stem Cell, 23*(3), 396-411 e398

Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A. I., Ranish, J., & Crabtree, G. R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences of the United States of America, 106*(13), 5181-5186

Houzelstein, D., Bullock, S. L., Lynch, D. E., Grigorieva, E. F., Wilson, V. A., & Beddington, R. S. (2002). Growth and early postimplantation defects in mice deficient for the bromodomain-containing protein Brd4. *Mol Cell Biol, 22*(11), 3794-3802

Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R., & Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature Structural & Molecular Biology, 20*(9), 1093-1097

Hu, D. Q., Gao, X., Morgan, M. A., Herz, H. M., Smith, E. R., & Shilatifard, A. (2013). The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology, 33*(23), 4745-4754

Huang, Y., Fang, J., Bedford, M. T., Zhang, Y., & Xu, R. M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science, 312*(5774), 748-751

Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylicz, J. J., Marioni, J. C., Hackett, J. A., & Surani, M. A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *Elife, 6*

Hutnick, L. K., Huang, X. H., Loo, T. C., Ma, Z. C., & Fan, G. P. (2010). Repression of Retrotransposal Elements in Mouse Embryonic Stem Cells Is Primarily Mediated by a DNA Methylation-independent Mechanism. *Journal of Biological Chemistry, 285*(27), 21082-21091

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet, 6*(9), e1001134

Imbeault, M., Helleboid, P. Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature, 543*(7646), 550-+

Inoue, A., Jiang, L., Lu, F. L., Suzuki, T., & Zhang, Y. (2017). Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature, 547*(7664), 419-+

Iqbal, K., Jin, S. G., Pfeifer, G. P., & Szabo, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America, 108*(9), 3642-3647

Ishiyama, S., Nishiyama, A., Saeki, Y., Moritsugu, K., Morimoto, D., Yamaguchi, L., Arai, N., Matsumura, R., Kawakami, T., Mishima, Y., Hojo, H., Shimamura, S., Ishikawa, F., Tajima, S., Tanaka, K., Ariyoshi, M., Shirakawa, M., Ikeguchi, M., Kidera, A., Suetake, I., Arita, K., & Nakanishi, M. (2017). Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. *Mol Cell, 68*(2), 350-360 e357

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature, 466*(7310), 1129-U1151

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science, 333*(6047), 1300-1303

Jachowicz, J. W., Bing, X. Y., Pontabry, J., Boskovic, A., Rando, O. J., & Torres-Padilla, M. E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics, 49*(10), 1502-+

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., & Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Molecular and Cellular Biology, 24*(20), 8862-8871

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, T., Lee, P., Wilson, C., Lander, E., & Jaenisch, R. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genetics, 27*(1), 31-39

Jacobs, F. M. J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A. D., Katzman, S., Paten, B., Salama, S. R., & Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature, 516*(7530), 242-+

Jambhekar, A., Dhall, A., & Shi, Y. (2019). Roles and regulation of histone methylation in animal development. *Nature Reviews Molecular Cell Biology, 20*(10), 625-641

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017a). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols, 12*(4), 828-863

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017b). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc, 12*(4), 828-863

Kacem, S., & Feil, R. (2009). Chromatin mechanisms in genomic imprinting. *Mammalian Genome, 20*(9-10), 544-556

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., & Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *Embo Journal, 32*(3), 340-353

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature, 429*(6994), 900-903

Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., & Kazazian, H. H., Jr. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev, 23*(11), 1303-1312

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., & Lorincz, M. C. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell, 8*(6), 676-687

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C. A., Mostoslavsky, G., Lahesmaa, R., Orkin, S. H., Rodig, S. J., Daley, G. Q., & Rao, A. (2011). Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage

Specification in Mouse Embryonic Stem Cells. *Cell Stem Cell, 8*(2), 200-213
Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C. H., Ilicic, T., Henriksson, J., Natarajan, K. N., Tuck, A. C., Gao, X. F., Buhler, M., Liu, P. T., Marioni, J. C., & Teichmann, S. A. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell, 17*(4), 471-485
Kunarso, G., Chia, N. Y., Jeyakani, J., Hwang, C., Lu, X. Y., Chan, Y. S., Ng, H. H., & Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics, 42*(7), 631-U111
Kunath, T., Saba-El-Leil, M. K., 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*(16), 2895-2902
Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T. W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., & Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development, 22*(7), 908-917
Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B. B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K. M., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R., & Young, R. A. (2006). Control of developmental regulator's by polycomb in human embryonic stem cells. *Cell, 125*(2), 301-313
Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A., Surani, M. A., & Hajkova, P. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural & Molecular Biology, 20*(3), 311-316
Lesch, B. J., Silber, S. J., McCarrey, J. R., & Page, D. C. (2016). Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nature Genetics, 48*(8), 888-+
Li, C., Scott, D. A., Hatch, E., Tian, X., & Mansour, S. L. (2007). Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development. *Development, 134*(1), 167-176
Li, E., Bestor, T. H., & Jaenisch, R. (1992a). Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality. *Cell, 69*(6), 915-926
Li, E., Bestor, T. H., & Jaenisch, R. (1992b). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell, 69*(6), 915-926

Li, H. J., Liefke, R., Jiang, J. Y., Kurland, J. V., Tian, W., Deng, P. J., Zhang, W. D., He, Q., Patel, D. J., Bulyk, M. L., Shi, Y., & Wang, Z. X. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature, 549*(7671), 287-+

Li, W., Xu, H., Xiao, T. F., Cong, L., Love, M. I., Zhang, F., Irizarry, R. A., Liu, J. S., Brown, M., & Liu, X. S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology, 15*(12)

Li, X. J., Ito, M., Zhou, F., Youngson, N., Zuo, X. P., Leder, P., & Ferguson-Smith, A. C. (2008). A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell, 15*(4), 547-557

Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., Wang, M., Duan, X., Tan, J., Zheng, Y., Zhang, P., Fan, G., Wong, J., Xu, G. L., Wang, Z., Wang, H., Gao, S., & Zhu, B. (2018). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature, 564*(7734), 136-140

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., & Schubeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genetics, 43*(11), 1091-U1078

Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., & Ferguson-Smith, A. C. F. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nature Genetics, 35*(1), 97-102

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., & He, C. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol, 10*(2), 93-95

Liu, Y. W., Toh, H., Sasaki, H., Zhang, X., & Cheng, X. D. (2012). An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes & Development, 26*(21), 2374-2379

Lu, X. Y., Sachs, F., Ramsay, L., Jacques, P. E., Goke, J., Bourque, G., & Ng, H. H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature Structural & Molecular Biology, 21*(4), 423-U168

Lynch, V. J., Leclerc, R. D., May, G., & Wagner, G. P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics, 43*(11), 1154-U1158

Madan, B., Madan, V., Weber, O., Tropel, P., Blum, C., Kieffer, E., Viville, S., & Fehling, H. J. (2009). The Pluripotency-Associated Gene Dppa4 Is Dispensable for Embryonic Stem Cell Identity and Germ Cell Development but Essential for Embryogenesis. *Molecular and Cellular Biology, 29*(11), 3186-3203

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. *Nature Communications, 4*

Maldonado-Saldivia, J., Van den Bergen, J., Krouskos, M., Gilchrist, M., Lee, C., Li, R., Sinclair, A. H., Surani, M. A., & Western, P. S. (2007). Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells, 25*(1), 19-28

Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., Pirrotta, V., Reinberg, D., & Gamblin, S. J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature, 461*(7265), 762-U711

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A., & Stunnenberg, H. G. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell, 149*(3), 590-604

Mas, G., Blanco, E., Ballare, C., Sanso, M., Spill, Y. G., Hu, D. Q., Aoi, Y., Le Dily, F., Shilatifard, A., Marti-Renom, M. A., & Di Croce, L. (2018). Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nature Genetics, 50*(10), 1452-+

Masaki, H., Nishida, T., Kitajima, S., Asahina, K., & Teraoka, H. (2007). Developmental pluripotency-associated 4 (DPPA4) localized in active chromatin inhibits mouse embryonic stem cell differentiation into a primitive ectoderm lineage. *Journal of Biological Chemistry, 282*(45), 33034-33042

Masaki, H., Nishida, T., Sakasai, R., & Teraoka, H. (2010). DPPA4 modulates chromatin structure via association with DNA and core histone H3 in mouse embryonic stem cells. *Genes to Cells, 15*(4), 327-337

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature, 464*(7290), 927-931

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M., & Bernstein, B. E. (2010). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *Plos Genetics, 6*(12)

Messerschmidt, D. M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., & Knowles, B. B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. *Science, 335*(6075), 1499-1502

Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & Development, 28*(8), 812-828

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., & Yamanaka, S. (2003). The

homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell, 113*(5), 631-642
Myant, K., Termanis, A., Sundaram, A. Y. M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J. I., & Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Research, 21*(1), 83-94
Nady, N., Gupta, A., Ma, Z., Swigut, T., Koide, A., Koide, S., & Wysocka, J. (2015). ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife, 4*
Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., Tanaka, S., Shiota, K., & Nakano, T. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology, 9*(1), 64-U81
Nakamura, T., Liu, Y. J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., & Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature, 486*(7403), 415-+
Nakamura, T., Nakagawa, M., Ichisaka, T., Shiota, A., & Yamanaka, S. (2011). Essential Roles of ECAT15-2/Dppa2 in Functional Lung Development. *Molecular and Cellular Biology, 31*(21), 4366-4378
Nakashima, H., Kimura, T., Kaga, Y., Nakatani, T., Seki, Y., Nakamura, T., & Nakano, T. (2013). Effects of Dppa3 on DNA Methylation Dynamics During Primordial Germ Cell Development in Mice. *Biology of Reproduction, 88*(5)
Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature, 543*(7643), 72-77
Nichols, J., & Smith, A. (2009). Naive and Primed Pluripotent States. *Cell Stem Cell, 4*(6), 487-492
Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F., Koseki, H., & Nakanishi, M. (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature, 502*(7470), 249-253
Niwa, H., Ogawa, K., Shimosato, D., & Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature, 460*(7251), 118-122
Noh, K. M., Wang, H. B., Kim, H. R., Wenderski, W., Fang, F., Li, C. H., Dewell, S., Hughes, S. H., Melnick, A. M., Patel, D. J., Li, H. T., & Allis, C. D. (2015). Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. *Molecular Cell, 59*(1), 89-103
O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A., & Jenuwein, T. (2001). The Polycomb-group gene Ezh2 is required for

early mouse development. *Molecular and Cellular Biology, 21*(13), 4330-4336
Oda, M., Kumaki, Y., Shigeta, M., Jakt, L. M., Matsuoka, C., Yamagiwa, A., Niwa, H., & Okano, M. (2013). DNA Methylation Restricts Lineage-specific Functions of Transcription Factor Gata4 during Embryonic Stem Cell Differentiation. *Plos Genetics, 9*(6)
Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell, 99*(3), 247-257
Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet, 19*(3), 219-220
Okashita, N., Kumaki, Y., Ebi, K., Nishi, M., Okamoto, Y., Nakayama, M., Hashimoto, S., Nakamura, T., Sugasawa, K., Kojima, N., Takada, T., Okano, M., & Seki, Y. (2014). PRDM14 promotes active DNA demethylation through the Ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development, 141*(2), 269-280
Oksuz, O., Narendra, V., Lee, C. H., Descostes, N., Leroy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P. P., Garcia, B. A., Skok, J. A., & Reinberg, D. (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Molecular Cell, 70*(6), 1149-+
Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature, 448*(7154), 714-717
Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., & Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *Embo Reports, 10*(11), 1235-1241
Payer, B., Saitou, M., Barton, S. C., Thresher, R., Dixon, J. P. C., Zahn, D., Colledge, W. H., Carlton, M. B. L., Nakano, T., & Surani, M. A. (2003). stella is a maternal effect gene required for normal early development in mice. *Current Biology, 13*(23), 2110-2117
Percharde, M., Lin, C. J., Yin, Y. F., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X. H., & Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell, 174*(2), 391-+
Piunti, A., & Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science, 352*(6290)
Popp, C., Dean, W., Feng, S. H., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., & Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature, 463*(7284), 1101-U1126
Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V., Grimaldi, G., Riccio, A., &

Trono, D. (2011). In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. *Molecular Cell, 44*(3), 361-372

Reddington, J. P., Perricone, S. M., Nestor, C. E., Reichmann, J., Youngson, N. A., Suzuki, M., Reinhardt, D., Dunican, D. S., Prendergast, J. G., Mjoseng, H., Ramsahoye, B. H., Whitelaw, E., Greally, J. M., Adams, I. R., Bickmore, W. A., & Meehan, R. R. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biology, 14*(3)

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature, 447*(7143), 425-432

Ren, R., Horton, J. R., Zhang, X., Blumenthal, R. M., & Cheng, X. D. (2018). Detecting and interpreting DNA methylation marks. *Current Opinion in Structural Biology, 53*, 88-99

Rickels, R., Herz, H. M., Sze, C. C., Cao, K. X., Morgan, M. A., Collings, C. K., Gause, M., Takahashi, Y. H., Wang, L., Rendleman, E. J., Marshall, S. A., Krueger, A., Bartom, E. T., Piunti, A., Smith, E. R., Abshiru, N. A., Kelleher, N. L., Dorsett, D., & Shilatifard, A. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nature Genetics, 49*(11), 1647-+

Riising, E. M., Comet, I., Leblanc, B., Wu, X. D., Johansen, J. V., & Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Molecular Cell, 55*(3), 347-360

Robbez-Masson, L., Tie, C. H. C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I. A., Timms, R. T., Herrero, J., Lehner, P. J., & Rowe, H. M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res, 28*(6), 836-845

Rodriguez-Terrones, D., & Torres-Padilla, M. E. (2018). Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends in Genetics, 34*(10), 806-820

Rothbart, S. B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A. I., Barsyte-Lovejoy, D., Martinez, J. Y., Bedford, M. T., Fuchs, S. M., Arrowsmith, C. H., & Strahl, B. D. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nature Structural & Molecular Biology, 19*(11), 1155-1160

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature, 463*(7278), 237-240

Sasai, N., Nakao, M., & Defossez, P. A. (2010). Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Research, 38*(15), 5015-5022

Sato, M., Kimura, T. U., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M., & Nakano, T. (2002). Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mechanisms of Development, 113*(1), 91-94

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Goncalves, A., Kutter, C., Brown, G. D., Marshall, A., Flicek, P., & Odom, D. T. (2012). Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages (vol 148, pg 335, 2012). *Cell, 148*(4), 832-832

Schubeler, D. (2015). Function and information content of DNA methylation. *Nature, 517*(7534), 321-326

Schwammle, V., Sidoli, S., Ruminowicz, C., Wu, X. D., Lee, C. F., Helin, K., & Jensen, O. N. (2016). Systems Level Analysis of Histone H3 Post-translational Modifications (PTMs) Reveals Features of PTM Crosstalk in Chromatin Regulation. *Molecular & Cellular Proteomics, 15*(8), 2715-2729

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., & Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell, 48*(6), 849-862

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., & Zhang, F. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science, 343*(6166), 84-87

Sharif, J., Endo, T. A., Nakayama, M., Karimi, M. M., Shimada, M., Katsuyama, K., Goyal, P., Brind'Amour, J., Sun, M. A., Sun, Z. X., Ishikura, T., Mizutani-Koseki, Y., Ohara, O., Shinkai, Y., Nakanishi, M., Xie, H. H., Lorincz, M. C., & Koseki, H. (2016). Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. *Cell Stem Cell, 19*(1), 81-94

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature, 450*(7171), 908-912

Shen, L., Inoue, A., He, J., Liu, Y. T., Lu, F. L., & Zhang, Y. (2014). Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell, 15*(4), 459-470

Shi, X. B., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G.,

Shi, Y., Cote, J., Chua, K. F., & Gozani, O. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature, 442*(7098), 96-99
Shukla, R., Upton, K. R., Munoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., Brennan, P. M., Baillie, J. K., Collino, A., Ghisletti, S., Sinha, S., Iannelli, F., Radaelli, E., Dos Santos, A., Rapoud, D., Guettier, C., Samuel, D., Natoli, G., Carninci, P., Ciccarelli, F. D., Garcia-Perez, J. L., Faivre, J., & Faulkner, G. J. (2013). Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma. *Cell, 153*(1), 101-111
Silva, J., & Smith, A. (2008a). Capturing pluripotency. *Cell, 132*(4), 532-536
Silva, J., & Smith, A. (2008b). Capturing pluripotency. *Cell, 132*(4), 532-536
Sim, Y. J., Kim, M. S., Nayfeh, A., Yun, Y. J., Kim, S. J., Park, K. T., Kim, C. H., & Kim, K. S. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. *Stem Cell Reports, 8*(5), 1312-1328
Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols, 13*(5), 1006-1019
Smallwood, S. A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S. R., & Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genetics, 43*(8), 811-U126
Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H. C., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature, 484*(7394), 339-U374
Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet, 14*(3), 204-220
Song, J., Rechkoblit, O., Bestor, T. H., & Patel, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science, 331*(6020), 1036-1040
Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schubeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature, 480*(7378), 490-495
Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S., & Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell, 163*(1), 218-229
Stewart, K. R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S., Smallwood, S. A., Chen, T., & Kelsey, G. (2015). Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev, 29*(23), 2449-2462
Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., & Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical

Polycomb repressive complex PRC1.6. *PLoS Genet, 14*(1), e1007193

Streubel, G., Watson, A., Jammula, S. G., Scelfo, A., Fitzpatrick, D. J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G. L., Dillon, E., Wynne, K., Pasini, D., Krogan, N. J., Bracken, A. P., & Cagney, G. (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Molecular Cell, 70*(2), 371-+

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., & Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem, 279*(26), 27816-27823

Sze, C. C., Cao, K. X., Collings, C. K., Marshall, S. A., Rendleman, E. J., Ozark, P. A., Chen, F. X., Morgan, M. A., Wang, L., & Shilatifard, A. (2017). Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes & Development, 31*(17), 1732-1737

Tahiliani, M., Koh, K. P., Shen, Y. H., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science, 324*(5929), 930-935

Takahashi, N., Coluccio, A., Thorball, C. W., Planet, E., Shi, H., Offner, S., Turelli, P., Imbeault, M., Ferguson-Smith, A. C., & Trono, D. (2019). ZNF445 is a primary regulator of genomic imprinting. *Genes & Development, 33*(1-2), 49-54

Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., & Tajima, S. (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proceedings of the National Academy of Sciences of the United States of America, 108*(22), 9055-9059

Tchasovnikarova, I. A., Timms, R. T., Matheson, N. J., Wals, K., Antrobus, R., Gottgens, B., Dougan, G., Dawson, M. A., & Lehner, P. J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science, 348*(6242), 1481-1485

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R. W., Deaton, A., Andrews, R., James, K. D., Turner, D. J., Illingworth, R., & Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature, 464*(7291), 1082-U1162

Timms, R. T., Tchasovnikarova, I. A., Antrobus, R., Dougan, G., & Lehner, P. J. (2016). ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. *Cell Reports, 17*(3), 653-659

Tu, S., Narendra, V., Yamaji, M., Vidal, S. E., Rojas, L. A., Wang, X. S., Kim, S. Y., Garcia, B. A., Tuschl, T., Stadtfeld, M., & Reinberg, D. (2016). Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature, 534*(7607), 387-+

Turner, B. M. (2007). Defining an epigenetic code. *Nature Cell Biology, 9*(1), 2-6

van Mierlo, G., Dirks, R. A. M., De Clerck, L., Brinkman, A. B., Huth, M., Kloet, S. L., Saksouk, N., Kroeze, L. I., Willems, S., Farlik, M., Bock, C., Jansen, J. H., Deforce, D., Vermeulen, M., Dejardin, J., Dhaenens, M., & Marks, H. (2019a). Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell, 24*(1), 123-+

van Mierlo, G., Dirks, R. A. M., De Clerck, L., Brinkman, A. B., Huth, M., Kloet, S. L., Saksouk, N., Kroeze, L. I., Willems, S., Farlik, M., Bock, C., Jansen, J. H., Deforce, D., Vermeulen, M., Dejardin, J., Dhaenens, M., & Marks, H. (2019b). Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. *Cell Stem Cell, 24*(1), 123-137 e128

Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit-Mehouas, S., Arnaud, P., Tomizawa, S., Andrews, S., & Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biology, 16*

Voigt, P., Tee, W. W., & Reinberg, D. (2013). A double take on bivalent promoters. *Genes & Development, 27*(12), 1318-1338

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z. Q., Kroeze, L. I., Nesterova, T. B., Jansen, J. H., Xie, H. H., He, C., Reik, W., & Stunnenberg, H. G. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells (vol 62, pg 848, 2016). *Molecular Cell, 62*(6), 983-983

Voncken, J. W., Roelen, B. A. J., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J., & van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proceedings of the National Academy of Sciences of the United States of America, 100*(5), 2468-2473

Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics, 20*(2), 116-117

Walter, M., Teissandier, A., Perez-Palacios, R., & Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife, 5*

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X.,

& Liu, J. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell, 157*(4), 979-991

Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., Burgess, S. M., Brachmann, R. K., & Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Sciences of the United States of America, 104*(47), 18613-18618

Weber, A. R., Krawczyk, C., Robertson, A. B., Kusnierczyk, A., Vagbo, C. B., Schuermann, D., Klungland, A., & Schar, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications, 7*

Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., Horth, C., McGuire, J. T., Xu, X., Nikbakht, H., Lemiesz, A. E., Marchione, D. M., Marunde, M. R., Meiners, M. J., Cheek, M. A., Keogh, M. C., Bareke, E., Djedid, A., Harutyunyan, A. S., Jabado, N., Garcia, B. A., Li, H., Allis, C. D., Majewski, J., & Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature, 573*(7773), 281-286

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., Jiao, F., Liu, H., Yang, P., Tan, L., Lan, F., Shi, Y. G., He, C., Shi, Y., & Diao, J. (2018). Zc3h13 Regulates Nuclear RNA m(6)A Methylation and Mouse Embryonic Stem Cell Self-Renewal. *Mol Cell, 69*(6), 1028-1038 e1026

Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., & Shilatifard, A. (2008). Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Molecular and Cellular Biology, 28*(24), 7337-7344

Wu, R., Liu, Y., Zhao, Y., Bi, Z., Yao, Y., Liu, Q., Wang, F., Wang, Y., & Wang, X. (2019). m(6)A methylation controls pluripotency of porcine induced pluripotent stem cells by targeting SOCS3/JAK2/STAT3 pathway in a YTHDF1/YTHDF2-orchestrated manner. *Cell Death Dis, 10*(3), 171

Wu, R. F., Liu, Y. H., Zhao, Y. L., Bi, Z., Yao, Y. X., Liu, Q., Wang, F. Q., Wang, Y. Z., & Wang, X. X. (2019). m(6)A methylation controls pluripotency of porcine induced pluripotent stem cells by targeting SOCS3/JAK2/STAT3 pathway in a YTHDF1/YTHDF2-orchestrated manner. *Cell Death & Disease, 10*

Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., & Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature, 389*(6652), 745-749

Xiang, Y., Zhang, Y., Xu, Q., Zhou, C., Liu, B., Du, Z., Zhang, K., Zhang, B., Wang, X., Gayen, S., Liu, L., Wang, Y., Li, Y., Wang, Q., Kalantry, S., Li, L., & Xie, W. (2020). Epigenomic analysis of gastrulation identifies

a unique chromatin state for primed pluripotency. *Nat Genet, 52*(1), 95-105

Xiang, Y. L., Zhang, Y., Xu, Q. H., Zhou, C., Liu, B. F., Du, Z. H., Zhang, K., Zhang, B. J., Wang, X. X., Gayen, S., Liu, L., Wang, Y., Li, Y. Y., Wang, Q. J., Kalantry, S., Li, L., & Xie, W. (2020). Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nature Genetics, 52*(1), 95-+

Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A. B., Zhang, Y., Zhang, B., Yu, G., Xia, W., Du, Z., Huang, C., Ma, J., Zheng, H., Li, Y., Liu, C., Walker, C. L., Jonasch, E., Lefebvre, L., Wu, M., Lorincz, M. C., Li, W., Li, L., & Xie, W. (2019). SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat Genet, 51*(5), 844-856

Xu, Q. H., & Xie, W. (2018). Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. *Trends in Cell Biology, 28*(3), 237-253

Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T., & Yamada, Y. (2017). Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature, 548*(7666), 224-+

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., & Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nature Genetics, 40*(8), 1016-1022

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., & Saitou, M. (2013). PRDM14 Ensures Naive Pluripotency through Dual Regulation of Signaling and Epigenetic Pathways in Mouse Embryonic Stem Cells. *Cell Stem Cell, 12*(3), 368-382

Yang, B. X., El Farran, C. A., Guo, H. C., Yu, T., Fang, H. T., Wang, H. F., Schlesinger, S., Seah, Y. F., Goh, G. Y., Neo, S. P., Li, Y., Lorincz, M. C., Tergaonkar, V., Lim, T. M., Chen, L., Gunaratne, J., Collins, J. J., Goff, S. P., Daley, G. Q., Li, H., Bard, F. A., & Loh, Y. H. (2015). Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell, 163*(1), 230-245

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schubeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science, 356*(6337)

Ying, Q. L., Nichols, J., Chambers, I., & Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell, 115*(3), 281-292

Ying, Q. L., 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*(7194), 519-U515

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics, 13*(8), 335-340

Zaccara, S., Ries, R. J., & Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol, 20*(10), 608-624

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., Liu, W., Kou, X., Zhao, Y., He, W., Li, C., Chen, B., Li, Y., Wang, Q., Ma, J., Yin, Q., Kee, K., Meng, A., Gao, S., Xu, F., Na, J., & Xie, W. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature, 537*(7621), 553-557

Zhang, M., Lai, Y. W., Krupalnik, V., Guo, P. C., Guo, X. P., Zhou, J. G., Xu, Y., Yu, Z. J., Liu, L. Q., Jiang, A., Li, W. J., Ma, G., Li, N., Fu, X. L., Lv, Y., Jiang, M. L., Tariq, M., Kanwal, S., Liu, H., Xu, X. T., Zhang, H., Huang, Y. H., Wang, L. L., Chen, S. H., Babarinde, I. A., Luo, Z. W., Wang, D. Y., Zhou, T. T., Ward, C., He, M. H., Ibanez, D. P., Li, Y. P., Zhou, J. J., Yuan, J., Feng, Y. Y., Arumugam, K., Di Vicino, U., Bao, X. C., Wu, G. M., Schambach, A., Wang, H. T., Sun, H., Gao, F., Qin, B. M., Hutchins, A. P., Doble, B. W., Hartmann, C., Cosma, M. P., Qin, Y., Xu, G. L., Chen, R. S., Volpe, G., Chen, L., Hanna, J. H., & Esteban, M. A. (2020). beta-Catenin safeguards the ground state of mouse pluripotency by strengthening the robustness of the transcriptional apparatus. *Science Advances, 6*(29)

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., & Jeltsch, A. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res, 38*(13), 4246-4253

Zheng, H., Huang, B., Zhang, B. J., Xiang, Y. L., Du, Z. H., Xu, Q. H., Li, Y. Y., Wang, Q. J., Ma, J., Peng, X., Xu, F., & Xie, W. (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell, 63*(6), 1066-1079

Zhu, H. M., Geiman, T. M., Xi, S. C., Jiang, Q., Schmidtmann, A., Chen, T. P., Li, E., & Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. *Embo Journal, 25*(2), 335-345

Original publications

Paper I

ARTICLES
https://doi.org/10.1038/s41594-020-0445-1

nature structural & molecular biology

Dppa2 and *Dppa4* counteract de novo methylation to establish a permissive epigenome for development

Kristjan H. Gretarsson and Jamie A. Hackett ✉

Early mammalian development entails genome-wide epigenome remodeling, including DNA methylation erasure and reacquisition, which facilitates developmental competence. To uncover the mechanisms that orchestrate DNA methylation dynamics, we coupled a single-cell ratiometric DNA methylation reporter with unbiased CRISPR screening in murine embryonic stem cells (ESCs). We identify key genes and regulatory pathways that drive global DNA hypomethylation, and characterize roles for *Cop1* and *Dusp6*. We also identify *Dppa2* and *Dppa4* as essential safeguards of focal epigenetic states. In their absence, developmental genes and evolutionarily young LINE1 elements, which are specifically bound by DPPA2, lose H3K4me3 and gain ectopic de novo DNA methylation in pluripotent cells. Consequently, lineage-associated genes and LINE1 acquire a repressive epigenetic memory, which renders them incompetent for activation during future lineage specification. *Dppa2/4* thereby sculpt the pluripotent epigenome by facilitating H3K4me3 and bivalency to counteract de novo methylation, a function co-opted by evolutionarily young LINE1 to evade epigenetic decommissioning.

Mammalian fertilization is accompanied by widespread epigenetic remodeling of inherited genomes, including global DNA demethylation and reorganization of chromatin landscapes[1–4]. This epigenetic resetting equalizes the distinct parental epigenomes and also correlates with the emergence of naive pluripotency, implying that epigenome remodeling is central to the establishment of developmental 'competence'. Such competence confers the capacity of the genome to transcriptionally respond to future inductive signals for multiple lineages. This is particularly critical for lineage-associated genes that must be transiently repressed during pluripotent phases whilst remaining competent (primed) for robust activation in subsets of forthcoming cell fates[5,6]. Indeed, the importance of a permissive epigenome is supported by observations of impaired or reduced developmental competence after somatic cell nuclear transfer (SCNT) or in induced pluripotent stem cells, which are susceptible to incomplete epigenetic resetting[7,8]. Investigating the complex mechanisms that underpin epigenome (re)programming is therefore an important focus towards understanding developmental potency.

Several lines of evidence indicate that resetting DNA methylation (DNAme) during development is mediated by parallel mechanisms[9]. Amongst these, repression of the maintenance DNA methylation machinery is central and appears to occur through post-translational regulation of UHRF1 (refs. [10,11]), at least in part via STELLA activity[12–14]. This is further supported by PRDM14, which suppresses de novo methylases and is necessary for DNA hypomethylation in naive pluripotent cells[15,16]. In parallel, replication-independent DNAme erasure occurs on both the maternal and paternal genome[1]. Counterintuitively, de novo methylation remains active throughout epigenetic reprogramming but is offset, in part, via TET proteins[17]. These collective mechanisms contribute towards resetting the epigenome, but also present an opportunity for transposable elements (TE), such as LINE1, to mobilize due to epigenetic derestriction. Such LINE1 activation has been linked with key developmental events[18–20], but could also represent a hazard to the genome if left unrestrained[21,22]. Epigenetic (re)programming therefore probably strikes a balance between genome-wide resetting to a competent state for development, and targeted regulation. Nevertheless, a complete understanding of the mechanisms that crosstalk to remodel the epigenome, how they interact to balance focal and global effects, and the full repertoire of genes involved, is lacking.

Here we coupled a single-cell ratiometric reporter of cellular DNA methylation status with CRISPR screening to unbiasedly identify the gene networks that underpin DNAme remodeling. In doing so we identify upstream regulators of global DNAme erasure in pluripotent cells. We also identify *Dppa2* and *Dppa4* as key genes that safeguard against focal de novo DNA methylation and epigenetic silencing at lineage-associated genes by integrating chromatin states, and consequently confer developmental competence. Remarkably, full-length LINE1 elements appear to have exapted this *Dppa2/4* function to escape epigenetic surveillance and enable competence for precocious activation, potentially highlighting an evolving genomic conflict.

Results

Single-cell monitoring of DNA demethylation. To identify regulators of epigenetic remodeling, we exploited the reporter for genomic DNA methylation (RGM)[23], which tracks the dynamic DNA methylation state of single cells with green fluorescent protein (GFP). We optimized the system for CRISPR screening in two ways. First, we replaced the original *Snrpn* imprinted promoter for the core *Kcnq1ot1* imprinted promoter, which enhanced the dynamic range of reporter activity (*e*RGM), enabling better separation of hypo- and hypermethylated cells (Extended Data Fig. 1a). Second, we converted the readout to a ratiometric measure by introducing an additional *Ef1α*-mCherry that is not sensitive to DNA methylation (Fig. 1a). This enables a single-cell ratiometric score (*e*RGM(GFP):*Ef1α*-mCherry) that normalizes for confounding effects on a reporter in a screen (for example, disruption of

European Molecular Biology Laboratory (EMBL), Epigenetics & Neurobiology site, Rome, Italy. ✉e-mail: jamie.hackett@embl.it

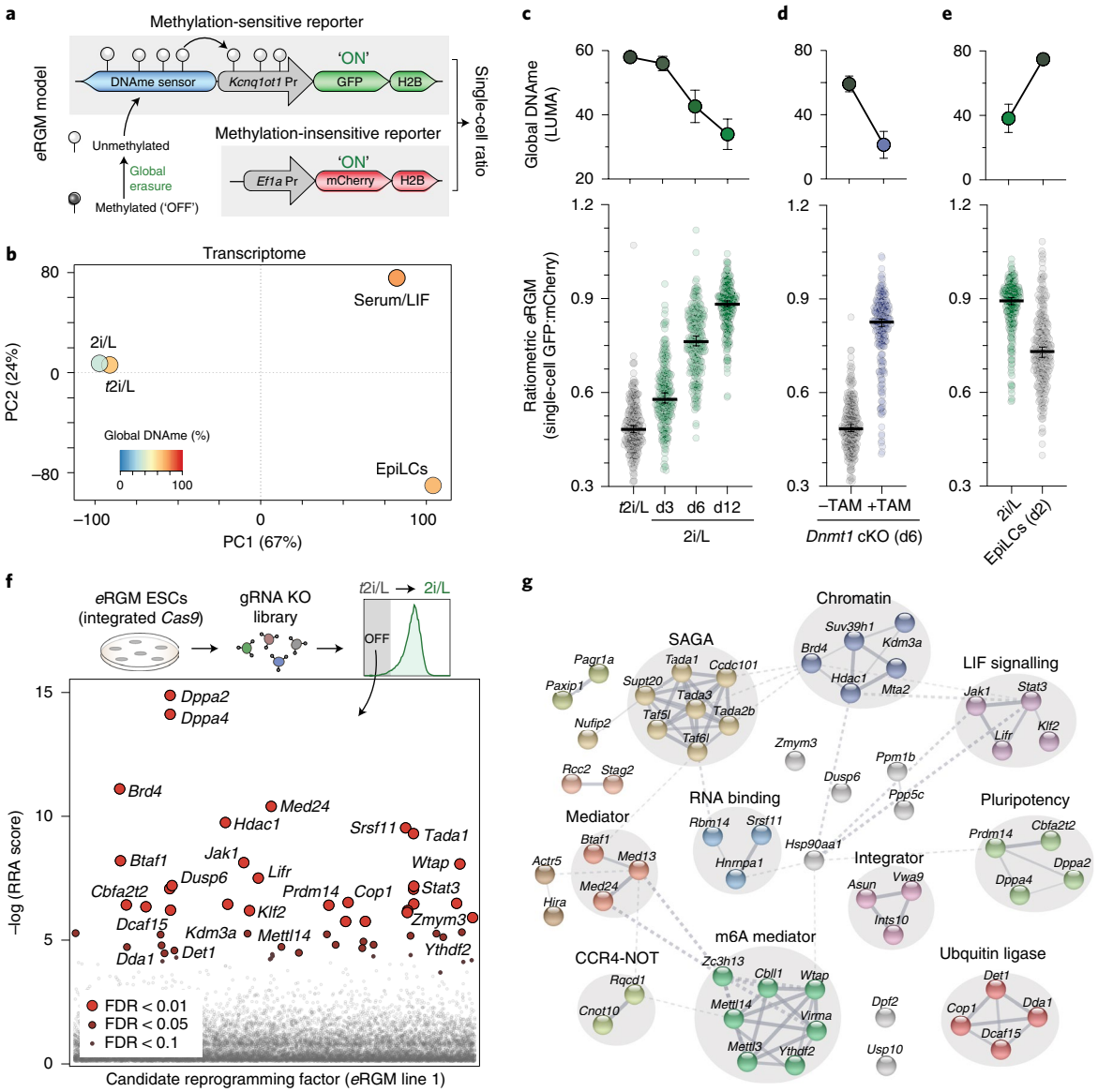


Fig. 1 | Developmental model and ratiometric reporter for DNA demethylation CRISPR screening. a, Schematic of the ratiometric *e*RGM real-time DNAme reporter. GFP is OFF in hypermethylated cells but expressed following hypomethylation, driven by a methylation-dependent imprinted promoter downstream of a DNAme sensor. mCherry remains active, establishing a single-cell ratio score. **b**, PCA of the transcriptomes (RNA-seq) of S/L (Serum/Lif)-, *t*2i/L- and 2i/L-cultured ESCs and EpiLCs, shaded by global DNA methylation level as determined by LUMA. **c**–**e**, Representative single-cell ratiometric scores of eRGM ($n = 250$ cells) during transition from *t*2i/L to 2i/L (**c**), after TAM-induced *Dnmt1* KO in *t*2i/L (**d**) or after EpiLC differentiation (**e**). Bars indicate median with 95% confidence intervals. Upper panels show corresponding changes in global DNA methylation, shown as mean ± s.d. of duplicate independent experiments. **f**, Significance scores (RRA) (see Methods for details) of CRISPR knockout (KO) screen candidates required for eRGM (LINE1) activation after DNA demethylation transition from *t*2i/L to 2i/L. **g**, STRING clustering of significant candidates (FDR < 0.05) required for eRGM activation from independent reprogramming screens.

translation factors) or inherent cell–cell variance (for example, cell cycle stage).

To test ratiometric *e*RGM, we developed a model of developmentally induced DNA demethylation. Here, ESCs are maintained in a titrated 2i/L (*t*2i/L) condition (see Methods) to promote high global levels of DNA methylation (range, 64–58%), and are then transitioned to 2i/L status to induce global demethylation (range, 30–44%; $P = 0.0002$) (Fig. 1b). Importantly, global DNA demethylation after switching from *t*2i/L to 2i/L occurs without changes in cell identity, as judged by the transcriptome, which is in contrast to the switch from conventional serum/leukemia inhibitory factor (LIF) to 2i/L that constitutes a major transcriptional shift[24–26] (Fig. 1b and Extended Data Fig. 1b). Moreover, the induced DNA hypomethylation pattern is well correlated with developmentally imposed DNA demethylation in vivo (Extended Data Fig. 1c). Thus, the *t*2i/L→2i/L model specifically captures an authentic global epigenetic transition, including global DNA demethylation, without changes in cell identity that could confound a screen for epigenome regulators.

We next examined the capacity for detection of DNA demethylation events in single cells by generating independent ESC lines carrying the ratiometric *e*RGM system. In *t*2i/L, *e*RGM was silenced in >95% of cells, consistent with high global DNA methylation. In contrast, *e*RGM exhibited a progressive activation concomitant with induced DNA methylation erasure in 2i/L, leading to *e*RGM activation in 12% of single cells after 3 days (d), 67% after 6 d and in >95%

of cells following complete DNA hypomethylation at 12 d (Fig. 1c). Independent *e*RGM lines exhibited consistent response to induced hypomethylation (Extended Data Fig. 1d). Notably, *Ef1α*-mCherry did not alter expression during this transition, enabling its use as a ratiometric normalizer (Extended Data Fig. 1d). To further confirm that *e*RGM directly reports cellular DNA methylation status, we used ESCs wherein tamoxifen (TAM) drives Cre-mediated deletion of *Dnmt1* (cDKO) and, consequently, global DNA demethylation occurs independent of culture conditions[27]. Following TAM exposure, we observed a strong and progressive activation of *e*RGM amongst single cells concomitant with cDKO-induced DNA hypomethylation (Fig. 1d).

Finally, we tested whether *e*RGM can also respond reciprocally to acquisition of DNA methylation by inducing differentiation of hypomethylated ESCs (in 2i/L) into hypermethylated epiblast-like cells (EpiLCs; global DNAme 33→75%). Here the reporter initiated rapid silencing in parallel with induction of DNA hypermethylation (Fig. 1e). We conclude that the enhanced ratiometric reporter of genomic DNAme (*e*RGM) represents a single-cell readout for dynamic transitions of cellular DNA methylation status.

A CRISPR screen for regulators of dynamic DNA methylation. To identify factors critical for DNA methylation resetting, we generated independent ESC lines carrying ratiometric *e*RGM and *spCas9* and introduced into them a CRISPR knockout guide RNA library[28]. To validate the strategy, we isolated ESCs that activated *e*RGM under hypermethylated (*t*2i/L) conditions, which is predicted to identify factors necessary to maintain DNA methylation and/or epigenetic silencing. This revealed that top hits comprised the key machinery for maintenance of DNA methylation, including *Dnmt1* (rank 5, false discovery rate (FDR) = 0.00049) and *Uhrf1* (rank 48, FDR = 0.066), unbiasedly confirming *e*RGM sensitivity to DNA hypomethylation (Extended Data Fig. 1e). We also identified chromatin-mediated silencers including *Setdb1* (rank 51, FDR = 0.073) and the HUSH complex (*Mphosph8*, rank 6; *Morc2a*, rank 9; *Fam208a*, rank 13). These data support *e*RGM specificity for the detection of developmental epigenome regulators, including those of cellular DNA methylation status.

We next aimed to identify factors contributing to resetting the epigenome at focal or global scales. We induced global DNA demethylation and isolated individual ESCs that failed to ratiometrically activate *e*RGM, indicative of a failure to undergo epigenetic resetting (Fig. 1f). Importantly, this population was highly enriched for knockout of *Prdm14* (rank 16, FDR = 0.0006), the key regulator known to instruct global DNA demethylation[15], as well as its heterodimeric cofactor *Cbfa2t2a* (rank 20, FDR = 0.0006)[29,30], supporting the sensitivity of the strategy for identification of reprogramming factors (Fig. 1f). Moreover, screens of independent *e*RGM ESC lines identified highly correlated ($P = 0.01$, Spearman relative ranking algorithm (RRA)) candidates (Extended Data Fig. 2a), suggesting that this system is robust. We therefore intersected significant hits (FDR < 0.05, fold-change > 3) from independent screens to define 56 core candidate genes linked with resetting the epigenome (Supplementary Table 1).

Gene ontology analysis suggested that candidates are enriched for nuclear localization (FDR = 3.2^{-16}), with roles in 'nucleic acid metabolism' (FDR = 1.1^{-7}) and 'histone modification' (FDR = 4.3^{-10}) (Extended Data Fig. 2b), consistent with epigenome regulation. Markov clustering in STRING revealed that they assembled into 13 clusters enriched for functional interactions (protein–protein interaction (PPI), $P < 1.0^{-16}$), implying that candidate genes link into common molecular pathways (Fig. 1g). Amongst these, a cluster corresponding to the LIF–JAK/STAT3 axis emerged, supporting the role previously observed for LIF signaling in promoting DNA hypomethylation in naive ESCs[26,31]. Moreover, a cluster linked with pluripotency included the most significant hits *Dppa2* and *Dppa4*, whilst another included chromatin regulators including *Brd4* and *Kdm3a*. We also observed multiple candidate factors involved in m6A RNA methylation (*Virma*, *Ythdf2*, *Mettl14*, *Zc3h13*, *Mettl3*, *Cbll1*), ubiquitin ligases (*Cop1*, *Det1*, *Dda1*, *Dcaf15*) and phosphatases (*Dusp6*, *Ppm1b*, *Ppp5c*) (Fig. 1g), which could potentially indirectly influence epigenetic status through acting as upstream regulators.

To validate the CRISPR screen hits, we generated knockout ESC populations for 24 selected candidates and transitioned them to hypomethylated conditions. Strikingly, knockout of each candidate resulted in a degree of impaired *e*RGM activation, implying altered epigenome remodeling in their absence (Fig. 2a). This effect was robust, since we generated additional knockouts in an independent *e*RGM line with similar outcomes (Extended Data Fig. 2c). Interestingly, the response kinetics of *e*RGM during transition to 2i/L varied amongst candidate knockout. For example, *Jak1*, *Dppa2*, *Dppa4* and *Brd4* mutants failed to activate *e*RGM per se, indicating a general block. In contrast, other candidate knockouts including *Dusp6*, *Kdm3a*, *Nufip2* and *Cop1* exhibited late-onset heterogeneous activation amongst single cells, implying delayed demethylation dynamics and reduced robustness in their absence. (Fig. 2b and Extended Data Fig. 2d). These validations suggest that candidate factors influence both the kinetics and absolute response of *e*RGM.

We next used luminometric CpG methylation assay (LUMA) to quantitatively assess global DNA methylation levels. Consistent with *e*RGM, we found that knockout of 20 of 24 candidate factors resulted in impaired global DNA demethylation across independent mutant lines (Fig. 2c). Amongst these is the known epigenetic regulator *Prdm14*, which maintained 58–64% global DNAme relative to hypomethylated wild-type (WT) control (39%), as well as *Cbfa2t2* (52–54%). Novel candidates that exhibited substantially elevated DNAme following knockout and transition to 2i/L include the phosphatase *Dusp6* (56–60%), the tyrosine kinase *Jak1* (65–70%), the epigenetic regulator *Brd4* (59–59%) and the E3 ubiquitin ligase *Cop1* (54–56%). These data suggest that our screen is sufficient to identify critical components of gene regulatory networks that contribute to driving complete DNA demethylation in naive ESCs.

***Dusp6* and *Cop1* promote global DNA hypomethylation.** To further investigate the role of candidates *Dusp6* and *Cop1* in epigenetic transitions, we generated independent clonal knockout ESC lines. DUSP6 is a phosphatase that acts downstream of MEK to attenuate the ERK signal cascade, whilst COP1 mediates ubiquitination and proteasomal degradation of target proteins[32,33]. We used enzymatic methyl-sequencing (EM-seq)[34], an enhanced equivalent of bisulfite-sequencing (BS-seq), to chart the global DNA methylome in WT, *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ naive ESCs, which confirmed that mutant lines remain hypermethylated in in 2i/L (*Dusp6*$^{-/-}$ 67%, *Cop1*$^{-/-}$ 58%) (Fig. 2d). Notably, elevated DNAme is distributed equivalently across genomic features including promoters, repeats and intergenic regions, indicating a general impairment of DNA demethylation rather than failure in locus-specific resetting (Fig. 2d,e).

Mechanistically, both *Cop1*$^{-/-}$ and *Dusp6*$^{-/-}$ ESCs exhibited transcriptional upregulation of the de novo methylation machinery (*Dnmt3a*, *Dnmt3b*, *Dnmt3L*), whilst *Dusp6*$^{-/-}$ cells additionally downregulate *Stella*, which together may contribute to impaired global DNA demethylation (Fig. 2f). Consistent with elevated DNAme in mutants, we observed strong repression of DNA methylation-dependent (germline) genes whilst naive genes were largely unaffected, implying no underlying change to pluripotency networks (Fig. 2f). However, we did observe inappropriate expression of some early developmental genes in *Cop1*$^{-/-}$ and *Dusp6*$^{-/-}$ ESCs, and their transcriptomes clustered separately by principle component analysis (PCA), which may partly reflect their disrupted epigenetic state (Fig. 2g). Taken together, our screen identifies and validates genes and pathways involved in the promotion of genome-scale DNA methylation transitions.

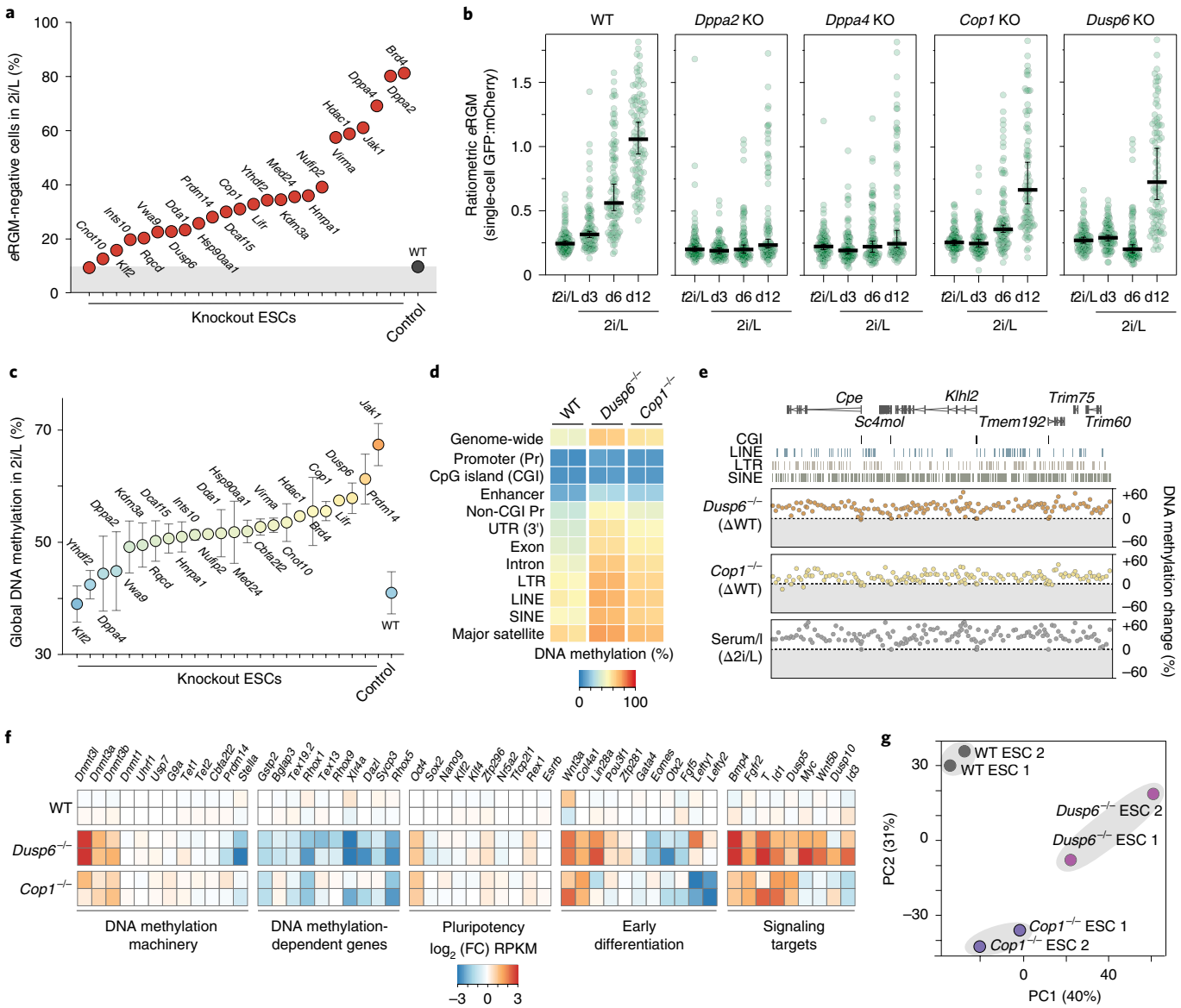


Fig. 2 | Validation of candidate DNAme reprogramming factors reveals a role for *Dusp6* and *Cop1*. **a**, Percentage of single cells with eRGM remaining 'OFF' in hypomethylated 2i/L following knockout (KO) of indicated candidate genes ($n = 24$) or control (non-targeting gRNA). **b**, Dynamics of ratiometric eRGM activation amongst single cells ($n = 250$) during transition from *t*2i/L to 2i/L in WT, *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$, *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ lines. Bars indicate median with 95% confidence interval. **c**, LUMA quantification of global DNA methylation levels in naive ESCs with indicated KO of candidate genes. Data represent mean ± s.d. of duplicate experiments with independent KO lines. **d**, Heatmap showing DNA methylation levels genome-wide and at indicated genomic features in WT, *Dusp6*$^{-/-}$and *Cop1*$^{-/-}$ naive ESCs. **e**, Representative genome view of DNA hypermethylation in *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ ESCs in 2i/L. Serum/l, serum/leukemia. Each data point represents percentage absolute change in DNA methylation over a 100-CpG tile relative to WT in 2i/L. **f**, Heatmap of $\log_2$ fold-change in gene expression for selected pathways in *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ ESCs normalized to mean WT. **g**, PCA of global transcriptomes in WT, *Dusp6*$^{-/-}$ and *Cop1*$^{-/-}$ ESCs. Source data for **c** are available online.

***Dppa2/4* protect against aberrant de novo DNA methylation.** Our screen is designed to identify both global and focal epigenome regulators. Amongst the *e*RGM screen hits, the paralogs *Dppa2* and *Dppa4* (hereafter, *Dppa2/4*) consistently ranked in the top five, suggesting a role in modulation of epigenome dynamics. However, in contrast to other candidates (for example, *Dusp6*, *Cop1* and *Jak1*), deletion of *Dppa2* or *Dppa4* did not affect genome-scale DNA demethylation (Fig. 2c). This could imply that, rather than a global influence, *Dppa2/4* modulate the methylation landscape at specific genomic features during pluripotent phases, when they are specifically expressed (Fig. 3a).

To explore the molecular function of *Dppa2/4* we generated multiple *Dppa2*$^{-/-}$ or *Dppa4*$^{-/-}$ ESC lines, and used EM-seq to chart their DNA methylome. Consistent with LUMA, *Dppa2/4* knockouts did not impact global DNA hypomethylation in naive ESCs (Fig. 3b). However, using sliding 50-CpG genomic tiles, we identified 1,662 differentially methylated regions (DMRs) (logistic regression adjusted $P < 0.05$ and binomial test $P < 0.01$). Remarkably, 1,605 (>96%) of these DMRs correspond to loci with elevated DNA methylation, whilst only 57 (3.4%) exhibit reduced DNAme, suggesting a general acquisition of focal hypermethylation in *Dppa2*$^{-/-}$ ESCs (Fig. 3c). *Dppa4*$^{-/-}$ ESCs exhibit a highly correlated pattern of hypermethylated DMRs (Fig. 3c), probably because disruption of one protein led to reciprocal destabilization of the other (Extended Data Fig. 3a). Strikingly, DNA hypermethylation was apparent at

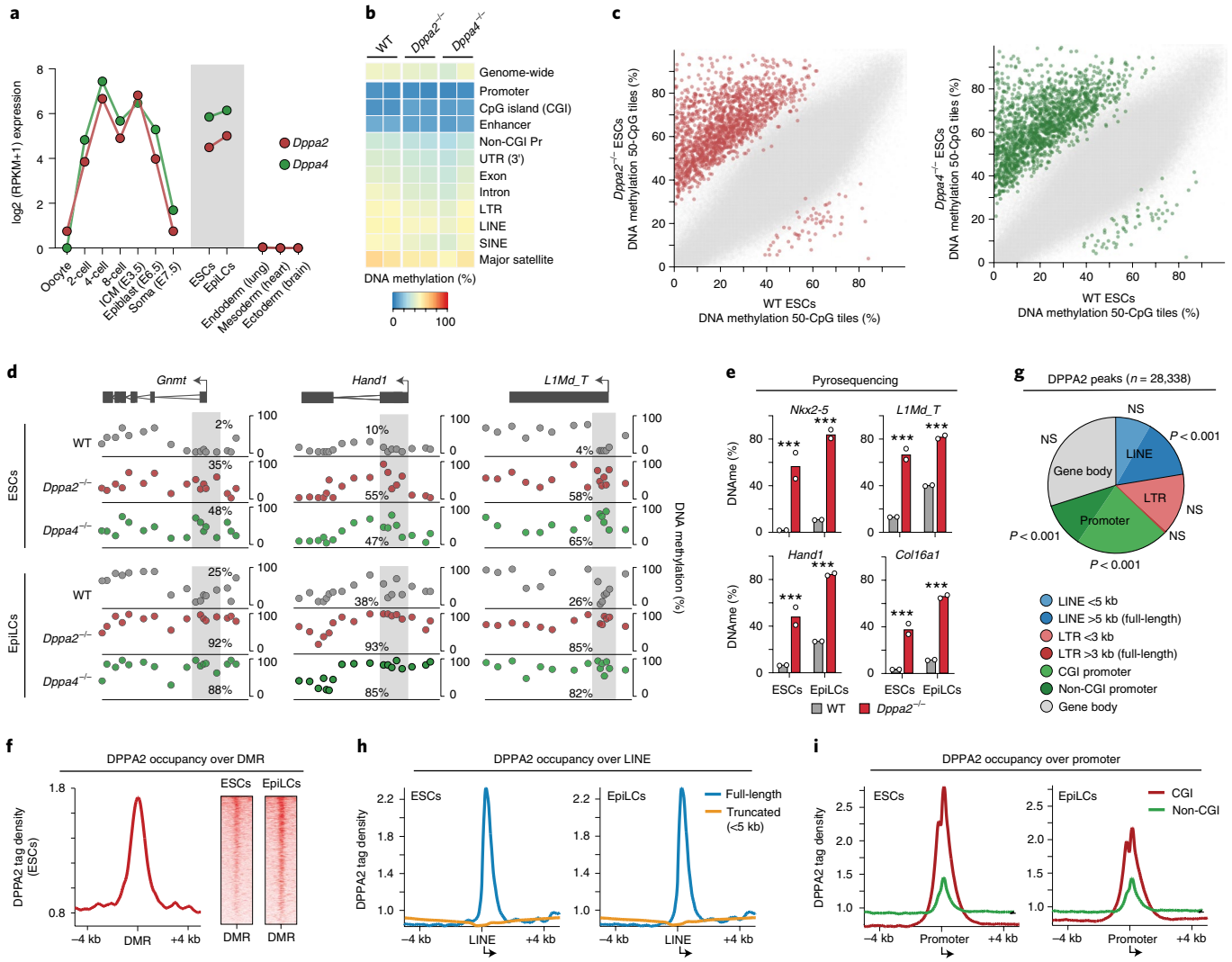


Fig. 3 | *Dppa2/4* occupancy promotes DNA hypomethylation in pluripotent cells. a, Line plot of *Dppa2/4* expression dynamics during early development, in pluripotent ESCs and EpiLCs, and in representative somatic tissues. **b**, Heatmap showing DNA methylation levels genome-wide and at indicated genomic features in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ naive ESCs. **c**, Scatter plot of DNA methylation across genome-wide sliding 50-CpG tiles in WT against *Dppa2*$^{-/-}$ or *Dppa4*$^{-/-}$ ESCs. Significant DMR (logistic regression adjusted $P < 0.05$ and binomial test $P < 0.01$) are shown in red (*Dppa2*) or green (*Dppa4*). **d**, Genome views showing DNA methylation distribution at genes and at LINE1 in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESCs and EpiLCs. Each data point corresponds to average methylation at sliding tiles of 15 or 20 CpG. **e**, Bisulfite pyrosequencing quantification of DNA methylation at selected gene and *L1Md_T* promoters, derived from two experiments of independent cell lines each analyzing multiple CpG sites ($n = 5$–10). Student's *t*-test, adjusted *P* values (Holm–Šidák), ***$P < 0.001$. **f**, Distribution of DPPA2 occupancy (CUT&RUN-seq) over DMRs ± 4 kb. **g**, Pie plot showing the distribution of genomic features associated with DPPA2 binding. Enrichment *P* values by one-sample *t*-test against triplicate simulations. NS, not significant. **h**,**i**, DPPA2 binding at full-length or truncated LINE1 promoters (**h**) and CpG island (CGI) and non-CGI promoters ± 4 kb (**i**). Source data for **e** are available online.

many gene promoters that usually remain strictly unmethylated at all developmental stages (Fig. 3d). This indicates that, rather than impaired DNA demethylation per se in *Dppa2/4* mutants, there is aberrant de novo methylation activity that could establish DNA methylation 'epimutations'.

To determine whether such epimutations persist during differentiation, we profiled EpiLCs, which correspond to a formative state that has undergone genomic remethylation. We observed that the hypermethylated sites established in *Dppa2*$^{-/-}$ or *Dppa4*$^{-/-}$ ESCs are retained in EpiLCs, whilst additional loci also acquire aberrant de novo methylation, including promoters (Extended Data Fig. 3b). Indeed, direct analysis identified 354 differentially methylated promoters (DMPs) in *Dppa2/4* mutants. Gene ontology analysis revealed these DMPs are enriched specifically for developmental processes (multicellular organism development FDR = 0.0053; developmental process FDR = 0.024; anatomical structure development FDR = 0.01) (Extended Data Fig. 3c). For example, *Hand1*, *Tnxb* and *Gnmt1* are key nodes for lineage-restricted cells and usually remain demethylated in all tissues, but acquire significant promoter hypermethylation (>90%) in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESCs and EpiLCs (Fig. 3d and Extended Data Fig. 3d). Intriguingly, in addition to developmental gene promoters, we observed that DMRs are enriched specifically for the 5' end of full-length (>5 kb) LINE1 elements (indicative of evolutionarily young LINEs), but not for truncated LINEs and long terminal repeat (LTR) elements (Extended Data Fig. 3b). Direct analysis identified 1,131 differentially methylated LINE1 (DML), of which >80% are *L1Md_T* (Fig. 3d). We used pyrosequencing to verify that LINE (*L1Md_T*), as well as the promoters of *Hand1*, *Nkx2-5* and *Col16a1*, acquire DNA methylation in *Dppa2*$^{-/-}$ naive ESCs (Fig. 3e), confirming a disrupted epigenomic landscape.

Because DMRs are focal rather than global, we next asked whether they reflect localized DPPA2/4 activity. We performed CUT&RUN for DPPA2 binding in WT cells and observed that genomic occupancy is strikingly increased over sites that become hypermethylated DMRs in *Dppa2/4*$^{-/-}$ cells, suggesting that DPPA2 may act proximally to sculpt the DNA methylome (Fig. 3f). Indeed, DPPA2-binding peaks ($n=28{,}338$) are significantly enriched specifically over gene promoters ($P<0.001$) and the 5' end of full-length LINE1 elements (>5 kb) ($P<0.001$) (Fig. 3g), consistent with DMR associations. Overall, promoters and LINE account for >65% of DPPA2 genomic occupancy. The latter enrichment is specific for full-length LINE, since DPPA2 is not enriched at truncated LINEs (Fig. 3h). Amongst promoters, DPPA2 exhibits a preference for CpG-dense loci, which is reflected by its GC-rich binding motifs and preference for CpG island (CGI) promoters (Fig. 3i and Extended Data Fig. 3e). Moreover DPPA2-binding profiles are highly correlated between ESCs and EpiLCs, implying that the additional DMR in the latter reflects their higher de novo activity rather than DPPA2 redistribution (Extended Data Fig. 3f). Notably, DPPA2 binding was observed at the sensor region used for *e*RGM (Extended Data Fig. 3g), explaining why a focal DNAme modulator was identified in the screen. In summary, *Dppa2* and *Dppa4* have a nonredundant role in protecting a target subset of developmentally associated promoters and full-length LINE elements from acquiring de novo DNA hypermethylation during naive and formative pluripotency phases, when *Dppa2/4* are specifically expressed.

DPPA2/4 binding establishes a permissive chromatin state. To understand the broader chromatin features associated with DPPA2 binding, and susceptibility to hypermethylation, we used CUT&RUN to profile H3K4me3, H3K27me3 and H3K9me3. DPPA2 occupancy correlates with strong H3K4me3 enrichment in WT cells, across all binding sites (Extended Data Fig. 4a). H3K27me3 is also enriched at a subset of DPPA2-bound sites, establishing bivalent states, but H3K9me3 is largely depleted. Strikingly, H3K4me3 enrichment at DPPA2-bound promoters occurs irrespective of expression state in both ESCs and EpiLCs (Fig. 4a), implying that DPPA2 may directly target H3K4me3, rather than H3K4me3 reflecting expression status of DPPA2-bound sites. Importantly, hypermethylated loci in *Dppa2/4* mutants (DMR) correspond to genomic regions that are H3K4me3-enriched and DPPA2-bound in WT (Extended Data Fig. 4b). Taken together this suggests a potential connection between DPPA2 occupancy, H3K4me3 and DNA methylation status.

To investigate this further, we assayed H3K4me3, H3K27me3 and H3K9me3 in *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ cells. Remarkably, deletion of *Dppa2* or *Dppa4* resulted in a dramatic loss of H3K4me3 across a significant subset of DPPA2-bound sites in both ESCs and EpiLCs, whilst the remaining sites were apparently unaffected (Fig. 4b). The subset of DPPA2 sites that lost H3K4me3 were enriched for full-length LINE1 elements and promoters (Fig. 4b). Moreover, the effect on H3K4me3 was specific, since there was no significant change of H3K9me3 in mutants whilst H3K27me3 was reduced at some loci such as *Txnb* but relatively unaffected at most (Fig. 4c,d and Extended Data Fig. 4c–e). In general, loci that lost H3K4me3 and gained DNAme in *Dppa2/4* mutants are associated with specific absolute levels of H3K4me3 enrichment in WT cells—intermediate for promoters and (relatively) high for full-length LINE1 (Fig. 4e). Because H3K4me3 can directly impair de novo DNA methylation[35,36], the dramatic depletion of H3K4me3 in *Dppa2/4*-mutants may enable aberrant DNA hypermethylation. In support of this, DMPs and DML correspond to loci that exhibit reduced H3K4me3 (Fig. 4c,d). Furthermore, by investigating all promoters and LINE1, we observed a marked negative correlation ($P<2.2\times10^{-16}$) between progressive H3K4me3 loss and DNAme gain following *Dppa2* knockout (Fig. 4f).

In summary, abrogation of *Dppa2/4* is linked with depletion in H3K4me3 at a specific subset of DPPA2 target loci, which directly correlates with acquisition of aberrant DNA hypermethylation. *Dppa2/4* could therefore integrate chromatin states to safeguard the pluripotent epigenome, particularly at developmentally associated genes and LINE1 elements.

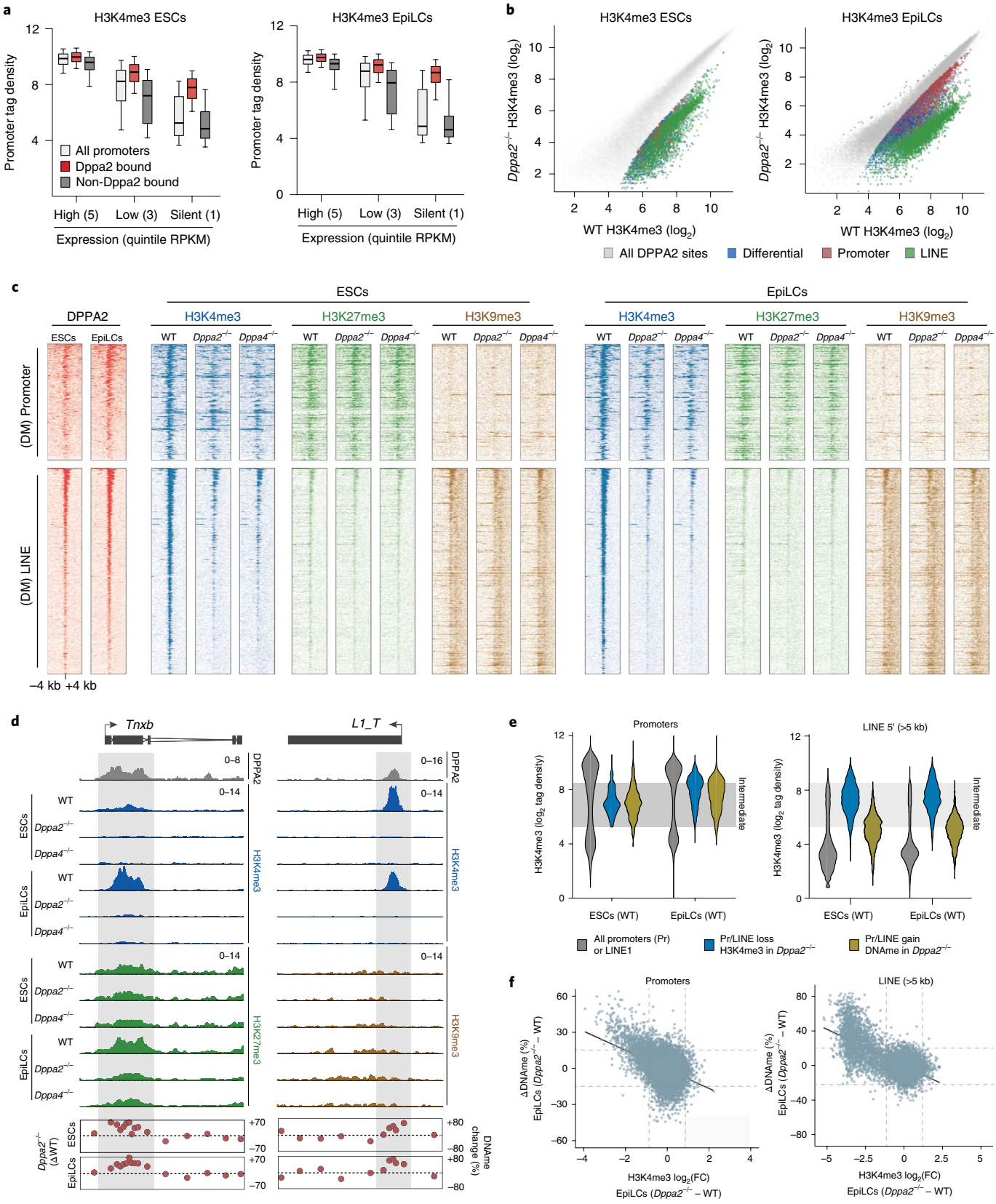
***Dppa2/4* ensure a competent epigenome for developmental expression.** To understand the relevance of *Dppa2/4*-mediated epigenome surveillance for developmental competence, we assessed the transcriptome. *Dppa2/4*$^{-/-}$ ESCs expressed an unperturbed naive pluripotency network and underwent apparently normal exit from pluripotency, since naive markers (*Nanog*, *Klf2* and *Prdm14*) were appropriately downregulated whilst formative markers (*Fgf5*, *Wnt3*, *Dnmt3a*) were upregulated as expected (Fig. 5a). Moreover, there was no difference in expression of the DNA methylation machinery or chromatin-modifying genes which, taken together, implies that *Dppa2/4* do not have an overarching influence on naive, early differentiation or epigenome gene regulatory networks (Fig. 5a).

Globally, *Dppa2/4*$^{-/-}$ naive ESCs exhibit a gene expression signature distinct but broadly comparable to WT, with 269 and 245 differentially expressed genes (DEG) in *Dppa2* and *Dppa4* knockout, respectively, primarily downregulated (85 and 82%, respectively) (Fig. 5b and Extended Data Fig. 5a). Induction of EpiLCs leads to a more divergent transcriptome as judged by PCA (Fig. 5c), with 801 and 611 DEG, again preferentially downregulated. Significantly, gene ontology indicated that these DEG in EpiLCs specifically relate to developmental processes (single multicellular organism process FDR = 0.000004, cell differentiation FDR = 0.00012) (Extended Data Fig. 5b), which reflects a general failure to activate genes involved in lineage-specific functions, particularly mesendoderm regulators. For example, *Hand 1*, *Cldn9*, *Tnxb* and others all failed to initiate primed expression in mutant EpiLCs (Fig. 5a and Extended Data Fig. 5c). This could be linked with the ectopic promoter DNA methylation acquired in the preceding ESC state. Indeed, the collective DMP geneset ($n=354$), which comprises many of the same mesendoderm genes including *Hand1*, *Tnxb*, *Ttl9*, *Cldn9* and *Gnmt*, is significantly upregulated in WT EpiLCs ($P=0.018$) consistent with priming of developmental genes but, strikingly, fails to initiate activation in either *Dppa2*$^{-/-}$ ($P=0.29$) or *Dppa4*$^{-/-}$ EpiLCs ($P=0.40$), suggesting they have lost competence for expression (Fig. 5d).

To understand whether the impaired expression of mesendoderm genes in EpiLCs represents a delay in their activation or stable

Fig. 4 | *Dppa2* and *Dppa4* establish chromatin states to safeguard against de novo methylation. a, Boxplot showing H3K4me3 enrichment over all DPPA2-bound and non-DPPA2-bound promoters, binned for expression quintile (RPKM). Box indicates the 25th, median and 75th percentiles, whiskers the 10th to 90th percentiles. **b**, Scatter plot of H3K4me3 enrichment at DPPA2-binding sites in WT and *Dppa2*$^{-/-}$ ESCs and EpiLCs. Significant differentially H3K4me3-enriched sites overlapping promoters (red), 5' LINE (green) or neither (blue) are highlighted. Significance by linear models for microarray data (LIMMA), $P<0.01$. **c**, Density plot of H3K4me3, H3K27me3 and H3K9me3 enrichment centered on DMP (upper) or LINE (DML) (lower) ± 4 kb in WT, *Dppa2*$^{-/-}$ *and Dppa4*$^{-/-}$ ESCs and EpiLCs. All plots are ordered equivalently by DPPA2-binding enrichment (left). **d**, Representative genome view of a developmental promoter and a LINE1 with tracks for DPPA2 occupancy, H3K4me3, H3K27me3 or H3K9me3, and changes in DNA methylation in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESCs and EpiLCs. **e**, Violin plots detailing distribution of H3K4me3 in WT cells at all promoters (left) or LINE1 (right), and those susceptible to H3K4me3 loss or DNAme gain following *Dppa2* knockout. **f**, Scatter plot showing correlated interrelationship between changes in H3K4me3 and DNA methylation following *Dppa2* knockout at all gene promoters (left) and all full-length LINE1 (right). Source data for **a**,**e** are available online.

silencing, we induced endoderm differentiation for 12 d. *Dppa2*$^{-/-}$ cells appeared morphologically equivalent to WT and activated master endoderm regulators, including *Emb* and *Foxa1*, with comparable dynamics, indicating no general impairment in differentiation (Fig. 5e). However, endoderm-associated genes, including *Gnmt*, *Nkx2-5*, *Col16a1* and *Hand1*, exhibited a highly significant failure to activate in mutants, even after 12 d of endoderm induction, implying an absolute blockade in their response (Fig. 5f). Importantly,



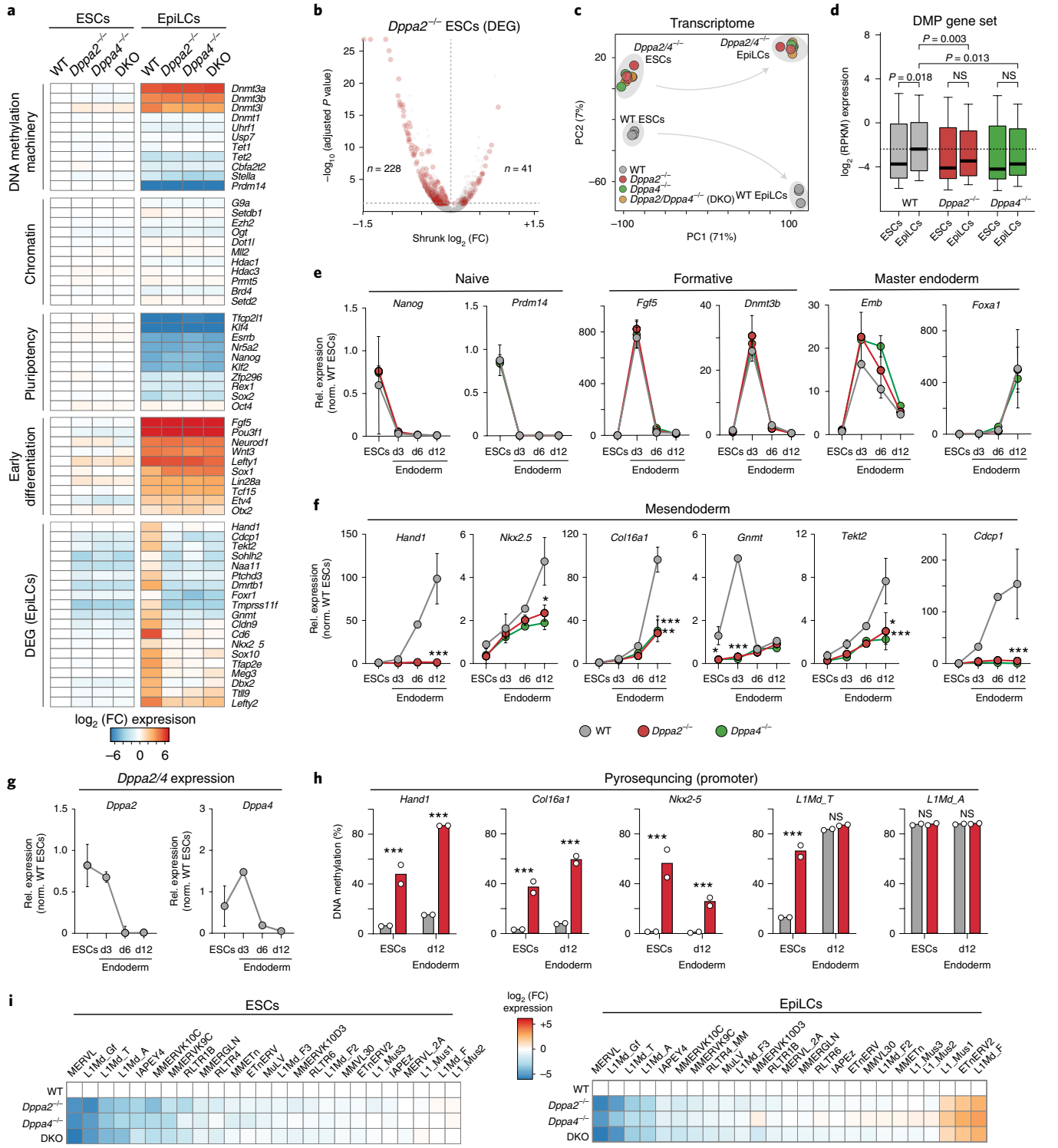


Fig. 5 | Developmental genes and LINE1 acquire transcriptional silencing memory in *Dppa2/4* mutants. a, Heatmap showing fold-change gene expression normalized to WT ESCs for selected pathways in *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$ and compound double KO (DKO) ESCs and EpiLCs. **b**, Volcano plot highlighting significant differentially expressed genes (DESeq2 FDR < 0.05) from *Dppa2*$^{-/-}$ ESCs compared to WT ESCs. **c**, Principal component analysis of global transcriptomes of WT, *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$ and DKO ESCs and EpiLCs. **d**, Boxplot showing activity of genes associated with DMPs in WT, *Dppa2*$^{-/-}$ and *Dppa4*$^{-/-}$ ESCs and EpiLCs. Box indicates the 25th, median and 75th percentiles, and whiskers the 10th to 90th percentiles. Data derived from expression of DMP geneset in $n = 3$ independent cell lines. Dashed line indicates median expression in WT EpiLCs. Significance by one-tailed Student's *t*-test. **e**–**g**, RT-qPCR quantification of expression of selected genes in ESCs to endoderm differentiation in WT, *Dppa2*$^{-/-}$ *and Dppa4*$^{-/-}$ normalized to WT ESCs. Data represent mean ± s.d. of two experiments using independent cell lines. Significance by adjusted Student's *t*-test (Holm–Šidák) $^{*}P > 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. **h**, Bisulfite pyrosequencing quantification of DNA methylation at selected promoters and LINE1 of multiple CpG sites ($n = 5$–10). Data represent mean ±s.d. of two experiments on independent cell lines. Significance by adjusted Student's *t*-test (Holm–Šidák), $^{***}P < 0.001$. **i**, Heatmap showing log fold-change (FC) expression normalized to WT of all full-length LINE1 (>5 kb) and LTR (>3 kb) elements above threshold in WT, *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$ and DKO ESCs and EpiLCs. Source data for **d**–**h** are available online.

Dppa2 is rapidly downregulated after 3 d of endoderm differentiation, but impaired gene upregulation manifests at later timepoints, suggesting a memory of previous DPPA2 activity (Fig. 5g). Indeed, pyrosequencing revealed that ectopic promoter DNA methylation established in ESCs propagates through to day-12 endoderm (Fig. 5h). Together, this indicates that the absence of *Dppa2/4* in pluripotent phases leads to impaired competence for gene activation during later differentiation. Importantly, cell fate transition per se appears unperturbed, but rather specific genes within the developmental program are rendered stably epigenetically silenced.

We next asked whether repetitive element activation is also affected by *Dppa2/4*, since many evolutionary young LINE1 become hypermethylated and lose H3K4me3 in their absence (Figs. 3d and 4c). We observed a particularly striking downregulation of full-length (>5 kb) LINE families (*L1Md_T*, *L1Md_A* and *L1Md_Gf*) in *Dppa2/4* mutant ESCs and EpiLCs (Fig. 5i and Extended Data Fig. 5d). We confirmed this using independent quantitative PCR with reverse transcription (RT–qPCR), which showed that disruption of *Dppa2* in ESCs and EpiLCs leads to extensive repression of *L1Md_T* (Extended Data Fig. 5e) whilst *MERVL* is also strongly repressed, consistent with recent reports[37]. *IAP* and other LTR elements were largely unaffected. These data suggest that the same *Dppa2/4*-dependent system that maintains epigenetic competence at developmental promoters may have been co-opted by LINE1 elements to evade epigenetic silencing in pluripotent phases.

H3K4me3 and DNAme interact to confer functional epigenetic memory. Last, we investigated whether induced DNA methylation and H3K4me3 loss is functionally instructive for subsequent gene-silencing memory. We noted that depletion of promoter H3K4me3 and gain of DNAme are both correlated with gene repression in *Dppa2*$^{-/-}$ cells (Extended Data Fig. 6a). Moreover, altered DNAme and H3K4me3 are also directly anticorrelated (Fig. 4f), implying a hierarchy of robust molecular changes following *Dppa2/4* abrogation. To determine whether acquired DNA methylation (and H3K4me3 loss) could instruct gene repression, we deleted *Dnmt1* in *Dppa2*$^{-/-}$ ESCs to generate compound mutants (*Dnmt1*$^{-/-}$*Dppa2*$^{-/-}$) that are hypomethylated and predicted to erase ectopic DNAme at developmental promoters and LINE1. Remarkably, analysis of the DMP geneset that acquired aberrant promoter methylation and silencing in *Dppa2*$^{-/-}$ EpiLCs revealed that additional deletion of *Dnmt1* partially rescues their activation block in EpiLCs. This effect is significant ($P=0.024$) among genes with CGI promoters, but not among non-CGI promoters ($P=0.23$) (Fig. 6a). Moreover, we observed reactivation of *L1Md_T* elements in *Dnmt1*$^{-/-}$*Dppa2*$^{-/-}$ ESCs and EpiLCs (Fig. 6b). These data imply that ectopic DNAme in *Dppa2*$^{-/-}$ cells is instructive, at least at some CpG-dense genes and LINE1, and directly impairs their response to inductive activating signals.

We next investigated whether the depletion of H3K4me3 in *Dppa2* mutant cells is also affected in *Dnmt1*$^{-/-}$*Dppa2*$^{-/-}$mutants and, surprisingly, observed reinstatement of H3K4me3 at a subset of promoters and most LINE1 elements (Fig. 6c,d and Extended Data Fig. 6b). This indicates a potentially complex interplay whereby absence of *Dppa2/4* leads to loss of H3K4me3, enabling aberrant DNA methylation, but that subsequent removal of ectopic DNAme tips the balance back, allowing H3K4me3 to reaccumulate through *Dppa2/4*-independent mechanisms (Extended Data Fig. 6c). More generally, we show that altered DNAme and H3K4me3 in the absence of *Dppa2/4* can propagate through differentiation to manifest as instructive gene silencing at future developmental stages, long after the epimutation is established. *Dppa2/4* therefore act as a safeguarding system during dynamic epigenome remodeling phases to ensure epigenetic competence for impending multilineage development.

Discussion

Here, we have established a ratiometric reporter of DNA methylation (*e*RGM) that is enhanced to enable unbiased CRISPR screening. By coupling this with an ESC model of developmental DNAme reprogramming, we identify and validate epigenome modulators that influence global DNA demethylation events and also focal DNAme states (*Dppa2* and *Dppa4*). The global regulators relate to diverse pathways such as m6A RNA methylation (for example, *Mettl3*, *Virma* and *Ytdhf2*), LIF signaling (for example, *Lifr*, *Jak1* and *Stat3*) and E3 ubiquitin ligases, which presumably exert influence through acting as upstream regulators as recently reported for *Nudt21* (ref. [38]). Amongst these, we show that the phosphatase *Dusp6* is necessary for completion of global DNA demethylation in naive ESCs. DUSP6 functions to attenuate MEK/ERK signaling[32], which is linked with DNA methylation[26,39], suggesting a probable connection. Indeed X-linked DUSP9 contributes to female-specific ESC hypomethylation by influencing MEK/ERK[40], and DUSP6 could play a comparable, but nonredundant, role in thresholding MEK/ERK more generally in pluripotent cells to promote epigenome erasure. Interestingly another screen hit, *Med24* (rank 3), also impacts MEK/ERK signaling[41]. Mechanistically, *Dusp6* may function via modulation of the de novo methylation machinery and/or *Stella*. The mechanism through which the E3 ubiquitin ligase *Cop1* modulates global epigenetic state is less clear, but could relate to regulation of the stability of proteins involved in maintaining DNAme, such as UHRF1 (ref. [11]).

In addition to global regulators we identify *Dppa2/4*, which we show guards against ectopic de novo methylation activity at key genomic sites during phases of both DNAme erasure (in naive ESCs) and remethylation (EpiLCs). Previous studies have shown that *Dppa2/4* overexpression enhances induced pluripotent stem cell generation, and they are linked with facilitating the two-cell program via modulation of *Dux*, suggesting broad functional roles[37,42,43]. Nevertheless, *Dppa2/4* mutant mice undergo normal embryogenesis but die perinatally due to aberrant gene repression in lung, where *Dppa2/4* are not expressed[44], implying that the phenotypically relevant activity of *Dppa2/4* is ensuring that lineage-associated genes are appropriately primed during early development. We dissect this molecularly by demonstrating that the absence of *Dppa2* or *Dppa4* leads to a marked loss of H3K4me3 and parallel acquisition of de novo DNA methylation at developmental genes and LINE1 elements, which propagates to manifest as epigenetic silencing in lineage-restricted cells. The equivalence of DPPA2 and DPPA4 probably reflects that they reciprocally stabilize each other (Extended Data Fig. 3), whilst the unusual association of DPPA2 outside classical pluripotency networks could underpin its distinct role[45].

Mechanistically, several lines of evidence suggest that DPPA2 targets H3K4me3. First, DPPA2 genomic binding sites are highly H3K4me3 enriched, irrespective of underlying transcription. Second, H3K4me3 is lost at a subset of sites following *Dppa2* deletion; and third, DPPA2 is reported to interact with the H3K4me methylase MLL2 (ref. [46]). Because H3K4me3 restricts the recruitment of de novo DNA methyltransferases[35,36], the depletion of H3K4me3 at these loci may enable access for ectopic DNA methylation to follow[47]. Indeed, we observe a striking correlation between the degree of H3K4me3 loss and DNAme gain in *Dppa2/4* mutants. Consistently, a *Dnmt3a* engineered to tolerate H3K4me3 enables aberrant de novo methylation at developmental genes[48], supporting a model whereby *Dppa2*-dependent H3K4me3 protects against DNAme. Such a system is probably necessary due to widespread de novo methylation activity throughout developmental (re)programming phases[17], which targets specific genomic compartments such as TE, but must also be restrained[49]. Notably, by counteracting de novo methylation, DPPA2/4 may also facilitate H3K27me3 accumulation and bivalency because a subset of loci exhibit H3K27me3

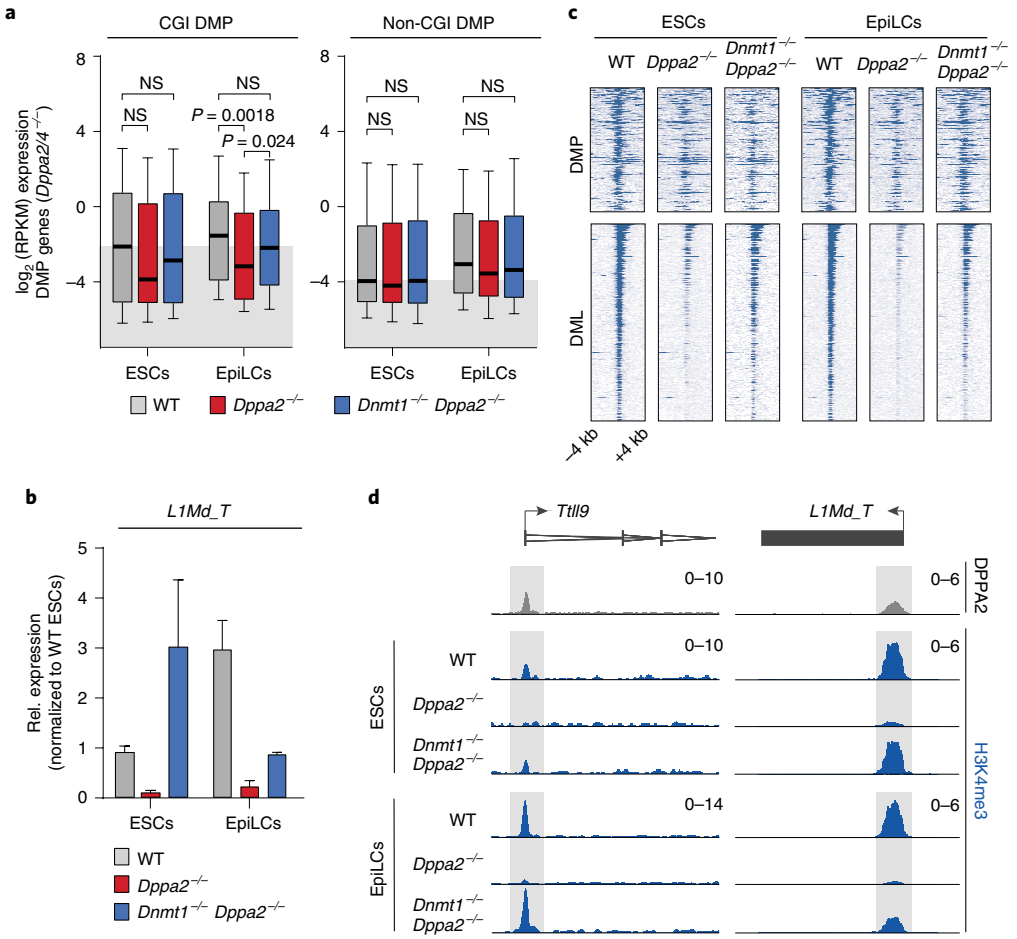


Fig. 6 | Functional interaction between DNAme, H3K4me3 and gene silencing. a, Boxplot showing expression of genes associated with CGI or non-CGI DMPs in WT, *Dppa2*$^{-/-}$ and rescue *Dnmt1*$^{-/-}$*Dppa2*$^{-/-}$ cells. The box indicates the 25th, median and 75th percentiles, whiskers the 10th to 90th percentiles; data derived from expression (RNA-seq) of 125 (CGI DMP) and 159 (non-CGI DMP) genes based on two experiments using independent cell lines. Significance by one-tailed Students *t*-test. **b**, RT–qPCR quantification of expression of *L1Md_T* in WT, *Dppa2*$^{-/-}$ and *Dnmt1*$^{-/-}$*Dppa2*$^{-/-}$ ESCs and EpiLCs, normalized to WT ESCs. Data represent mean ± s.d of two experiments using independent cell lines. **c**, Density plot showing enrichment of H3K4me3 centered on DMP (upper) or LINEs (DML, (lower)) ± 4 kb in WT, *Dppa2*$^{-/-}$ and *Dnmt1*$^{-/-}$*Dppa2*$^{-/-}$ ESCs and EpiLCs. **d**, Representative H3K4me3 genome tracks of a developmental promoter and LINE1 (*L1Md_T*). Source data for **a**,**b** are available online.

depletion in *Dppa2/4* mutants. This could reflect either direct loss of targeting by DPPA2/4 or inhibition of PRC2 activity by acquired DNAme[50–52]. Functionally, the acquisition of ectopic DNAme and loss of H3K4me3 appear to be instructive for epigenetic silencing following differentiation, at least at some CpG-dense promoters. Indeed, erasure of acquired promoter DNAme partially rescues expression defects and also reinstates H3K4me3. This suggests a switch-like interdependency whereby DPPA2-dependent H3K4me3 impairs DNAme, but acquired DNAme reciprocally prevents H3K4me3 accumulation, potentially underpinning stable transcriptional memory at target genes.

In addition to maintaining epigenetic competence at developmental genes, we observe that evolutionarily young LINE1 elements rely directly on *Dppa2* for H3K4me3 and their activity. This may reflect a strategy of successful LINE1 elements that have acquired/co-opted DPPA2-binding sites at their 5' ends to protect against host-directed epigenetic silencing. This would in turn enable expression of full-length LINE1s during early development when *Dppa2/4* are expressed, which aligns well with the optimal period for retrotransposition[53]. This scenario would represent a genomic conflict, whereby *Dppa2/4* activity is critical for epigenetic competence of lineage-associated genes, and therefore essential for viability, but also renders full-length LINE1s transcriptionally competent—a potential threat to genome integrity. An alternative scenario is that *Dppa2/4*-mediated activation of LINE1 reflects an important developmental role for the host genome. For example, LINE1 activation has been linked with establishment of zygotic chromatin accessibility, X-chromosome inactivation and regulation of the two-cell program[18–20]. Consequently, *Dppa2/4*-dependent LINE1 activation could represent exaptation by host systems to exploit LINE1 functionality at critical developmental points. In any case, *Dppa2/4* are placed at the center of an epigenetic competence circuit in pluripotent cells that facilitates expression of both LINE1 and developmental genes.

In summary, we characterize upstream gene networks that influence global DNA methylation erasure, and additionally uncover a complementary pathway that protects against the counterforce of uncontrolled de novo methylation during (re)programming to ensure developmental competence.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of

author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-020-0445-1.

Received: 4 March 2020; Accepted: 30 April 2020;
Published online: 22 June 2020

References

1. Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
2. Smith, Z. D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
3. Xia, W. et al. Resetting histone modifications during human parental-to-zygotic transition. *Science* **365**, 353–360 (2019).
4. Cantone, I. & Fisher, A. G. Epigenetic programming and reprogramming during development. *Nat. Struct. Mol. Biol.* **20**, 282–289 (2013).
5. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
6. Xiang, Y. et al. Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nat. Genet.* **52**, 95–105 (2019).
7. Smith, Z. D., Sindhu, C. & Meissner, A. Molecular features of cellular reprogramming and development. *Nat. Rev. Mol. Cell Biol.* **17**, 139–154 (2016).
8. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
9. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
10. Maenohara, S. et al. Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLoS Genet.* **13**, e1007042 (2017).
11. von Meyenn, F. et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* **62**, 848–861 (2016).
12. Funaki, S. et al. Inhibition of maintenance DNA methylation by Stella. *Biochem. Biophys. Res. Commun.* **453**, 455–460 (2014).
13. Li, Y. et al. Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* **564**, 136–140 (2018).
14. Mulholland, C. B. et al. Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals. Preprint at *bioRxiv* https://doi.org/10.1101/321604 (2020).
15. Leitch, H. G. et al. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
16. Hackett, J. A. et al. Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Rep.* **1**, 518–531 (2013).
17. Amouroux, R. et al. *De novo* DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
18. Jachowicz, J. W. et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).
19. Chow, J. C. et al. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**, 956–969 (2010).
20. Percharde, M. et al. A LINE1-Nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405 (2018).
21. Garcia-Perez, J. L., Widmann, T. J. & Adams, I. R. The impact of transposable elements on mammalian development. *Development* **143**, 4101–4114 (2016).
22. Rodriguez-Terrones, D. & Torres-Padilla, M. E. Nimble and ready to mingle: transposon outbursts of early development. *Trends Genet.* **34**, 806–820 (2018).
23. Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S. & Jaenisch, R. Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell* **163**, 218–229 (2015).
24. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
25. Walter, M., Teissandier, A., Perez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5**, e11418 (2016).
26. Hackett, J. A., Kobayashi, T., Dietmann, S. & Surani, M. A. Activation of lineage regulators and transposable elements across a pluripotent spectrum. *Stem Cell Rep.* **8**, 1645–1658 (2017).
27. Sharif, J. et al. Activation of endogenous retroviruses in *Dnmt1*$^{-/-}$ ESCs involves disruption of SETDB1-mediated repression by NP95 binding to hemimethylated DNA. *Cell Stem Cell* **19**, 81–94 (2016).
28. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
29. Nady, N. et al. ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* **4**, e10150 (2015).
30. Tu, S. et al. Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature* **534**, 387–390 (2016).
31. Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
32. Li, C., Scott, D. A., Hatch, E., Tian, X. & Mansour, S. L. Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development. *Development* **134**, 167–176 (2007).
33. Dornan, D. et al. The ubiquitin ligase COP1 is a critical negative regulator of p53. *Nature* **429**, 86–92 (2004).
34. Vaisvila, R. et al. EM-seq: detection of DNA methylation at single base resolution from picograms of DNA. Preprint at *bioRxiv* https://doi.org/10.1101/2019.12.20.884692 (2019).
35. Ooi, S. K. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
36. Guo, X. et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640–644 (2015).
37. De Iaco, A., Coudray, A., Duc, J. & Trono, D. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep.* **20**, e47382 (2019).
38. Brumbaugh, J. et al. Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* **172**, 106–120 (2018).
39. Choi, J. et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).
40. Choi, J. et al. DUSP9 modulates DNA hypomethylation in female mouse pluripotent stem cells. *Cell Stem Cell* **20**, 706–719 (2017).
41. Hamilton, W. B. et al. Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature* **575**, 355–360 (2019).
42. Eckersley-Maslin, M. et al. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).
43. Hernandez, C. et al. Dppa2/4 facilitate epigenetic remodeling during reprogramming to pluripotency. *Cell Stem Cell.* **23**, 396–411 (2018).
44. Nakamura, T., Nakagawa, M., Ichisaka, T., Shiota, A. & Yamanaka, S. Essential roles of ECAT15-2/Dppa2 in functional lung development. *Mol. Cell. Biol.* **31**, 4366–4378 (2011).
45. Engelen, E. et al. Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* **6**, 7155 (2015).
46. Eckersley-Maslin, M. A. et al. Dppa2/4 target chromatin bivalency enabling multi-lineage commitment. Preprint at *bioRxiv* https://doi.org/10.1101/832873 (2019).
47. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16**, 519–532 (2015).
48. Noh, K. M. et al. Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol. Cell* **59**, 89–103 (2015).
49. Boulard, M., Edwards, J. R. & Bestor, T. H. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* **47**, 479–485 (2015).
50. Reddington, J. P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).
51. Lindroth, A. M. et al. Antagonism between DNA and H3K27 methylation at the imprinted *Rasgrf1* locus. *PLoS Genet.* **4**, e1000145 (2008).
52. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
53. Kano, H. et al. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23**, 1303–1312 (2009).

Methods

Cell culture and differentiation. Murine ESC lines were either derived freshly (mixed 129/B6, XY) or obtained from H. Koseki (*Dnmt1*^{flox})[27]. All ESCs were routinely maintained and manipulated on gelatin in titrated *t*2i/L culture medium (NDIFF 227 supplemented with PD0325901 (200 nM), CHIR99021 (3 μM), LIF (1,000 U ml^{-1}), fetal bovine serum (FBS, 1%) and penicillin/streptomycin, and filtered through a 0.22 -M filter) in a humidified CO_2 incubator at 37 °C. *t*2i/L maintains genomic stability[39] and DNA hypermethylation[26] (see Fig. 1). ESCs were passaged every 2 d or at subconfluence by dissociation with TrpLE, and periodically checked for *Mycoplasma* contamination by ultrasensitive qPCR assay (Eurofins). To induce DNA hypomethylation, ESCs were transitioned into full 2i/L culture media[54] (NDIFF 227 supplemented with PD0325901 (1 μM), CHIR99021 (3 μM), LIF (1,000 U ml^{-1}), FBS (1%) and penicillin/streptomycin, filtered through a 0.22-μm filter) on gelatin for up to 14 d. To induce EpiLCs, 3×10^4 naive ESCs cm^{-2} were seeded on fibronectin-coated wells in EpiLC medium (NDIFF 227 supplemented with knockout serum replacement (1%), activin-A (20 ng ml^{-1}), basic fibroblast growth factor (12.5 ng ml^{-1}) and penicillin/streptomycin) for 44 h. To induce endodermal differentiation, 6×10^3 naive ESC cm^{-2} was seeded on gelatin-coated wells in 2i/L medium. After overnight culture, cells were washed three times with PBS and endodermal medium[55] was introduced (RPMI supplemented with L-glutamine (2 mM), FBS (0.2%), inducer of definitive endoderm 1 (5 μM) and penicillin/streptomycin). Endodermal medium was replaced every 2 d.

***e*RGM constructs.** A H2B-GFP-SV40pA cassette (Addgene, no. 11680) was cloned downstream of the mouse core imprinted *Kcnq1ot1* promoter (mm10, Chr7:143,296,371–143,296,745) into a piggyBac backbone vector using InFusion assembly (pPB-Kcnq1ot1-H2B-GFP). A genomic 'DNAme sensor' region derived from the mouse *Dazl* locus (mm10, Chr17:50,293,285–50,294,435) was amplified and inserted in the antisense orientation upstream of the *Kcnq1ot1* promoter, using infusion assembly to generate pPB-as*Dazl*sensor-*Kcnq1ot1*-H2B::GFP, which exhibits methylation-sensitive activity that tracks global DNAme levels. It also reports on specific regulators of focal DNAme that operate at *Dazl*. The antisense orientation of *Dazl* ensures that it bears no promoter activity on the reporter per se, but simply instructs the DNAme status of the adjacent *Kcnq1ot1* imprinted promoter in line with global levels. To establish a ratiometric system, an additional construct was generated by cloning the methylation-insensitive *EF1a* promoter upstream of an mCherry::H2B cassette into a piggyBac vector to generate pPB-EF1a-H2B::mCherry. Correct assembly and sequences was confirmed by tiled Sanger sequencing, and vectors were amplified and purified by endotoxin-free midi-preparations.

Generation of *e*RGM ESC lines. Embryonic stem cell lines carrying floxed *Dnmt1* alleles[27] and WT ESCs were transfected with pPB-as*Dazl*sensor-*Kcnq1ot1*-H2B::GFP (in silico DNA methylated with *M.SssI*), pPB-*EF1a*-H2B::mCherry, pPB-*spCas9*-Hygro[56] and PBase using Lipofectamine 3000. Transfected cells were selected for *spCas9* integration in titrated 2i/L using Hygromycin (250 μg ml^{-1}) for 5 d, and clonally derived cell lines were subsequently isolated and expanded. Clonal ESC lines were tested to confirm single-copy integrations by qPCR on genomic DNA, and their response to DNA demethylation was confirmed by evaluation of *e*RGM (GFP and mCherry) expression using flow cytometry after culture for 7 d in either *t*2i/L (hypermethylated) or 2i/L (hypermethylated). Further confirmation of *e*RGM response was determined by the addition of TAM (800 nM) for 6 d in *t*2i/L to induce conditional *Dnmt1* knockout and DNA hypomethylation. Clonal ESC lines exhibiting the best dynamic range of *e*RGM response were selected, and two independent lines (*e*RGM nos. 1 and 2) were used for CRISPR screening.

CRISPR screen. Lentiviral particles carrying the Brie gRNA library[28] were produced by transfection of Lenti-X HEK 293T with *pPax2* plasmid, *pMD2.G* plasmid and the *Brie* library plasmid, with Lipofectamine 3000 in a BSL2 tissue culture facility. Lentivirus-containing supernatant (medium) was harvested at 48 and 72 h after transfection, and clarified by filtering through a 0.22-μm, low-protein-binding unit. Viral particles were concentrated using a Lenti-X concentrator, according to the manufacturer's instructions, and resuspended in NDIFF 227. Lentiviral activity and efficiency were determined by transduction of ESCs across a titration curve, and assay of cell survival following puromycin selection for virally encoded integration of a resistance cassette. To generate knockout library cell lines, 7×10^7 ESCs of eRGM nos. 1 and 2 cultured in *t*2i/L were transduced with a predetermined number of lentiviral particles carrying the Brie genome-wide CRISPR knockout single guide RNA library ($n = 78{,}637$)[28] to ensure ~45% infection efficiency (>400-fold guide RNA coverage). Transduced cells were selected for with puromycin (1.2 μg ml^{-1}) for 7 d in *t*2i/L. The minimum population of cells was maintained at $>3.2 \times 10^7$ during passaging (>400-fold coverage) to ensure maintenance of library coverage. To initiate the screen knockout library, *e*RGM cell lines were transitioned into 2i/L for 12 d to drive extensive DNA hypomethylation. At 12 d, GFP-negative cells (defined as the lowest 1% of GFP expression) that also maintained normal mCherry expression—together indicative of incomplete epigenetic resetting—were purified by flow cytometry (291,248 and 237,121 for *e*RGM nos. 1 and 2, respectively). We additionally collected total unsorted cells ($>3 \times 10^7$) from both *t*2i/L and 2i/L, and GFP-positive cells (top 1%) from *t*2i/L (indicative of loss of epigenetic silencing), as controls. Genomic DNA was isolated from purified populations using the Quick-DNA microprep plus kit (Zymo Research, no. D3020) or a DNeasy blood and tissue kit (Qiagen, no. 69504). Integrated gRNAs from each population were amplified from genomic DNA using custom primers with the P7 flow cell overhangs: 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGG AGTTCAGACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCAC TGT-3' (8 base-pair (bp) barcode) and P5 overhang: 5'-AATGATACGGCGACCA CCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTG TGGAAAGGACGAAACACCG-3' using Q5 Hot Start High-Fidelity polymerase (NEB, no. M0494S) for 21–24 cycles. sgRNA amplicons were purified with SPRI beads (Beckman Coulter, no. B23318) following the manufacturer's instructions, and double-stranded DNA was quantified with Qubit III. Amplicon libraries were multiplexed and SE50 sequenced with a Nextseq500 Illumina system.

Gene editing in ESCs. To generate clonal knockout lines with CRISPR/Cas9, ESCs carrying *e*RGM and maintained in *t*2i/L were transiently transfected with a vector carrying a gRNA cassette targeted against a critical coding exon of the gene of interest (including *Dppa2*, *Dppa4*, *Dusp6* and *Cop1*; see Supplementary Table 1 for the full list of gRNA used), and selected with puromycin (1.2 μg ml^{-1}) for 60 h. Transfected cells were subsequently seeded at low density (1,000 cells per 9.6 cm^2) for single clone isolation. After clonal expansion, successful homozygous knockout lines (carrying frame-shifting indels) were confirmed by Sanger sequencing using the tracking of indels by decomposition tool[57], by immunoblot and via functional assays. To generate population-scale knockout of multiple candidate factors ($n = 24$), gRNAs targeting the gene(s) of interest were cloned into a piggyBac vector containing the enhanced gRNA cassette[58]. This was cotransfected with PBase into independent *e*RGM lines carrying *spCas9* activity using Lipofectamine 3000, following the manufacturer's recommendations. ESCs were selected for successful integration of gRNAs for 7 d with puromycin (1,2 μg ml^{-1}), which drives iterative targeting of the gene of interest until indel formation is induced. We assayed the population for successful knockout by immunoblot and flow cytometry, and typically observed that >95% of individual cells within each population carried homozygous functional knockout.

Flow cytomtery. Cells were gently dissociated into single-cell suspension using TrpLE and resuspended in PBS+1% FBS (fluorescent activated cell sorting (FACS) medium) and filtered. FACS was performed using a FACS Aria III (Becton Dickinson) and FACS Diva software. For flow analysis, samples were run on Attune NxT (Thermo Fisher Scientific). Data were analyzed using FlowJo v.10.5.3 (Tree Star).

LUMA. LUMA was used to measure global DNA (CpG) methylation levels. Briefly 200–500 ng of purified genomic DNA was split equally and subjected to two parallel 4-h restriction digests at 37 °C: digestion A, *HpaII/EcoRI*; digestion B; *MspI/EcoRI*, in which *EcoR1* is included as an internal reference. An equal volume of annealing buffer was added, and samples were loaded into a PyroMark Q24 Advanced pyrosequencer to quantitate the protruding ends from each digestion, using the dispensation order GTGTGTCACACAGTGTGT. Percentage DNA methylation was calculated by comparing the *EcoRI* normalized *HpaII* signal intensity ratio to the normalized *MspI* signal intensity ratio using the formula

$$HpaII \text{ ratio} : \frac{\text{Dispensation } 7+13}{\left(\frac{\text{Dispensation } 8+14}{2}\right)}$$

$$MspI \text{ ratio} : \frac{\text{Dispensation } 7+13}{\left(\frac{\text{Dispensation } 8+14}{2}\right)}$$

$$\%\text{ DNA methylation} : 100\% x \left(1 - \frac{HpaII}{MspI}\right)$$

Immunoblot. Cellular protein was extracted using RIPA buffer (Sigma, no. R0278) with protease inhibitors (Roche, no. 4693159001) at 4 °C for 30 min. After centrifugation at full speed, cell lysis supernatant was collected and Bolt LDS sample buffer (Thermo Fisher Scientific, no. B0007) and Bolt reducing agent (Thermo Fisher Scientific, no. B0004) were added to the samples. These were heated at 70 °C for 10 min and loaded onto 4–12% Bis-Tris gel (Thermo Fisher Scientific, no. NW04125BOX). Proteins were separated by 150-V electrophoresis for 30 min and blotted onto a polyvinylidene difluoride membrane using the iBlot Dry 2 blotting system. The membrane was blocked in 5% milk/PBS for 1 h at room temperature, followed by incubation with primary antibody (dilution 1:500–1:1,000; see Nature Research Reporting Summary for antibodies used) with 5% milk/PBS at 4 °C overnight and agitation. After washing twice with PBS/0.1% Tween, the membrane was incubated for 1 h at room temperature with a horseradish peroxidase-linked secondary antibody diluted 1:10,000 in 5% milk/PBS. The membrane was washed three times with PBS 0.1% Tween, and Pierce ECL immunoblot plus solution (Thermo Fisher Scientific, no. 32132) was added to the membrane for 5 min before imaging using the ChemiDoc XRS+ system (Bio-Rad).

RT–qPCR. Total RNA was isolated with RNeasy (Qiagen) and used to synthesize complementary DNA with a mixture of random hexamers and reverse

transcriptase, following DNAase treatment (TAKARA PrimeScript RT Reagent Kit with gDNA Eraser). Diluted cDNA was used in triplicate quantitative PCR reactions with pretested gene-specific primers and qPCRbio SYgreen Blue Mix using a QuantStudio 5 (Applied Biosystems) thermal cycler. Results were analyzed using 2–ΔΔCt (relative quantitation) with quantstudio software and normalization to housekeeping gene *Rplp0*. Statistical significance was determined using Students *t*-test with Holm–Šidák correction, with alpha = 0.05 to identify differences in gene expression from two biological replicates using Prism.

Pyrosequencing. Genomic DNA (50–300 ng) was sodium bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo Research, no. D5005), eluting with 10 µl of H_2O. Bisulfite-converted DNA (1 µl) was used as template to amplify target genomic regions using specific primers (one biotinylated) with the PyroMark PCR kit (Qiagen, no. 978703), following the recommended conditions (annealing temperature, 56 °C). PCR reaction (10 µl) was used for pyrosequencing, with PyroMark Q24 advanced reagents (Qiagen, no. 970902), a sequencing primer, and run on a PyroMark Q24 Advanced pyrosequencer (Qiagen) with target-specific dispensation orders. Statistical significance was determined using Students *t*-test with Holm–Šidák correction, with alpha = 0.05 to identify differences in DNA methylation using two independent biological replicate lines, each consisting of multiple CpG site DNAme calls.

CUT&RUN-seq. To investigate protein–DNA interactions and histone modification locations, we used the recently developed Cut&Run protocol[59]. Dissociated single cells were pelleted at 600*g* for 3 min and washed twice with Wash Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM Spermidine and Protease Inhibitor tablet (Roche)) at room tempersture, then resuspended in 1 ml of wash buffer. Ten microlitersof concanavalin A-coated magnetic beads (Bangs Laboratories, no. BP531), prewashed and resuspended in Binding buffer (20 mM Hepes-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$), were added to the cells with rotation for 10 min at room temperature. The bead-bound cells were isolated on a magnetic stand to remove the supernatant, and 300 µl of antibody buffer (wash buffer plus 0.02% digitonin and 2 mM EDTA) with 0.5 µg of antibody was added to the beads and incubated with rotation at 4 °C overnight. The following day, cell–bead complexes were washed with 1 ml of cold Dig-Wash buffer (wash buffer plus 0.02% digitonin) using a magnetic stand, then resuspended in 300 µl of cold Dig-Wash buffer. Purified protein-A::MNase (pA-MNase) fusion was added to a final concentration of 700 ng ml^{-1} and the samples were rotated for 1 h at 4 °C. Samples were washed twice in 1 ml of cold Dig-Wash buffer and resuspended in 50 µl of Dig-Wash buffer by gentle flicking. The samples were placed in iced water to precool them to 0 °C. To initiate pA-MNase digestion, 2 µl of 100-mM $CaCl_2$ was added and the samples were flicked to mix and returned to iced water. After either 30 min (histone modifications) or 5 min (DPPA2) pA-MNase digestion, 50 µl of 2XSTOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% digitonin, 250 µg of Rnase A, 250 µg of glycogen) was added and the samples thoroughly mixed. Samples were incubated at 37 °C for 10 min to release CUT&RUN fragments from the insoluble nuclear chromatin, followed by centrifugation at 16,000*g* for 5 min at 4 °C. The supernatants were transferred to new tubes and the cell–bead complexes discarded. Subsequently, 2 µl of 10% SDS and 2.5 µl of Proteinase K was added and the samples were incubated for 10 min at 70 °C. DNA fragments were purified and double-size selected from the suspension using SPRIselect beads (Beckman Coulter, no. B23318) following the manufacturer's protocol for double selection (0.5× beads/DNA ratio followed by 1.3), and eluted with 30 µl of 0.1× TE. CUT&RUN dsDNA samples were quantified with a Qubit III, and 5–10 ng used as input for library preparation with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, no. E7645S). Libraries were prepared using the following PCR program: 98 °C for 30 s, 98 °C for 10 s, 65 °C for 10 s and 65 °C for 5 min; steps 2 and 3 were repeated for 12 cycles. Library samples were paired-end sequenced on the NextSeq Illumina sequencing system (PE40).

EM-seq. Purified genomic DNA was isolated from cells using the Zymo microprep DNA kit. To generate high-quality base resolution DNA methylation libraries, we used the NEBnext Enzymatic Methyl-seq (EM-seq) kit following the manufacturer's instructions. Briefly, genomic DNA was sheared to 300 bp then end-repaired and A-tailed. Repaired DNA was then ligated to EM-seq adapters, and genomic 5mC and 5hmC was oxidized to 5caC to protect methylated sites against deamination. Subsequently, APOBEC was used to deaminate unmethylated cytosines to uracils whilst oxidized forms of 5mC/5hmC were not deaminated. This generated a DNA conversion system identical to bislfite conversion, but with higher yields and lower duplication. The library was amplified using Q5 polymerase, and independent libraries multiplexed. These were sequenced on an Illumina Nextseq for single-end 75 (SE75).

RNA-seq. Total RNA was collected from fresh cells using the Qiagen RNeasy kit, following the manufacturer's guidelines. Total RNA was quantitated using a Quibit III and quality checked with a Bioanalyser 2100 (Agilent) to ensure RNA integrity number >8.5. Messenger RNA was enriched using the NEBNext Poly(A) mRNA magnetic isolation module and prepped into stranded libraries using the NEBnext Ultra II directional RNA library prep kit, following all manufacturers' guidelines. Amplified libraries were multiplexed and sequenced on NextSeq (SE80 or PE40).

Bioinformatics analysis. *CRISPR screen.* Raw sequence reads were trimmed and quality control checked using cutadapt (v.1.15) (cutadapt -g TTGTGGAAAGGA CGAAACACCG) and FastQC[60]. Counting and statistical analysis of sgRNA frequency was performed with the tool Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout (MAGeCK, v.0.5.9)[61]. Normalized gRNA counts (MAGeCK -count -norm-method total) from sorted and unsorted control samples were compared using the -test command (default settings) in MAGeCK, which identifies significantly enriched/depleted gRNAs between samples. This is used to determine the RRA score, which identifies enriched/depleted genes based on change in the distribution frequency of multiple independent gRNAs targeting the same gene. To identify final candidates, we applied a FDR threshold of <0.05 and a fold-change gRNA frequency threshold of >3 (to select larger effect-size candidates) and intersected lists from two independent screens, each performed in independent cell lines.

RNA-seq. Raw reads were quality trimmed using TrimGalore (0.4.3.1, -phred33–quality 20–stringency 1 -e 0.1–length 20). These were mapped to the mouse mm10 (GRCm38) genome assembly using RNA Star (2.5.2b-0, default parameters except for–outFilterMultimapNmax 1000), and reads with a mapping quality (MAPQ) score <20 were discarded to ensure that only unique-mapping, high-quality alignments were used for analysis of gene expression. The data were quantified using the RNA-seq quantification pipeline for directional libaries in seqmonk software to generate $\log_2$ reads per million (RPM) or gene length-adjusted (reads per kilobase million, RPKM) gene expression values. Where appropriate, the samples were normalized using the match distribution quantitation method. DEG were determined using the DESeq2 package (v.1.24.0), inputting raw mapping counts and applying a multiple-testing adjusted P (FDR) < 0.05 significance threshold. An additional fold-change filter of >2 was applied to generate final DEG. Differences in gene expression of DMPs were tested with a one-tailed Students *t*-test. For TE analysis, reads with MAPQ score <20 were allowed to enable multimapping reads, but taking the primary alignment only. Repeat locations for the mm10 (GRCm38) genome were extracted from repeatmasker, and instances overlapping or residing within 2 kb of an annotated gene were excluded to prevent mixed signal derived from genic and/or TE transcription, which could confound results. Mapped reads were quantitated over these high-confidence TE, which were categorized as either full-length (>5 kb for LINE, >3 kb for LTR) or truncated (<5 kb for LINE, <3 kb for LTR), enabling unbiased assessment of TE expression from WT and independent replicate mutant lines. Summed counts for all instances of each class of repeat were calculated. These were then corrected for both total length of TE class and the sequencing depth of individual libraries to generate $\log_2$ RPM expression values as previously described[62].

EM-seq. Raw fastq sequences were quality- and adapter-trimmed using TrimGalore (0.4.3.1), and reads aligned to mm10 using Bismark (0.20.0), discarding the first 8 bp from the 5' end and the last 2 bp from the 3' of single-end reads. Cytosine methylation status was extracted from mapped reads using the Bismark methylation extractor tool. Genome-wide methylation calls were analyzed using Seqmonk software (1.44.0) with biological independent replicate datasets for each condition. To identify DMR, the genome was first binned into sliding tiles containing 50 consecutive CpGs and their methylation status determined using the DNA methylation pipeline. DMRs were identified by running read depth-sensitive logistic regression (P < 0.05) and binomial test (P < 0.01) statistical filters, taking only regions scored as significant in both, with a minimum of ten reads. Differentially methylated promoters were identified by quantitating DNAme over tiles spanning ±1 kb from RefSeq gene transcriptional start sites, and intersecting significant hits from both logistic regression (P < 0.05) and binomial test (P < 0.01) statistical filters. Differentially methylated full-length LINE1 (DML) were identified using tiles over the 5' end of LINE1 elements >5 kb (±500 bp) using repeatmasker annotations and intersecting logistic regression (P < 0.05) and binomial test (P < 0.01) statistical filters, with minimal reads of ten per probe. To generate final DMR, DMP and DML datasets, significant hits from ESC and EpiLC *Dppa2*$^{-/-}$ were collated.

CUT&RUN-seq. Raw Fastq sequences were quality- and adapter-trimmed with TrimGalore (0.4.3.1, -phred33–quality 20–stringency 1 -e 0.1–length 20) and aligned to the mouse mm10 genome using Bowtie2 (2.3.4.2, -I 50 -X 800–fr -N 0 -L 22 -i 'S,1,1.15'–n-ceil 'L,0,0.15'–dpad 15–gbar 4–end-to-end–score-min 'L,-0.6,-0.6'). Mapped sequences with MAPQ score <20 were discarded. DPPA2-binding peaks were identified using MACS2 ($P < 1 \times 10^{-5}$, 200-nt fragments) with IgG CUT&RUN as input control. The enrichment of overlap between DPPA2-binding sites and genomic features was determined by one-sample *t*-test against triplicate randomized genomic probesets of distribution size similar to DPPA2-binding sites. Histone modification CUT&RUN was analyzed by quantification of normalized reads over genomic features (for example, DPPA2-binding sites, DMPs, DML, DMRs) to generate density, peak and trend datasets using seqmonk software. For analysis of repetitive elements (for example, LINE) we allowed multimapping (MAPQ < 20), taking only the primary alignment.

Gene Ontology analysis. Gene Ontology analyses were performed using ensemble gene ID in the tool DAVID (v.6.8)[63] for differentially expressed genes (WT versus *Dppa2* knockout and *Dppa4* knockout in ESCs and EpiLCs), genes associated with DMPs and screen candidates. FDR values for selected Gene Ontology terms from BP_all are displayed.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequencing data derived from RNA-seq, EM-seq, Cut&Run-seq and CRISPR screening have been deposited in the Gene Expression Omnibus database under accession code GSE146863. Source data for Figs. 2c, 3e, 4a,e, 5d–h and 6a,b are available online.

References

54. Ying, Q. L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
55. Borowiak, M. et al. Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell* **4**, 348–358 (2009).
56. Hackett, J. A. et al. Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.* **9**, 4292 (2018).
57. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
58. Chen, B. H. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
59. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
61. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
62. Berrens, R. V. et al. An endosiRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell* **21**, 694–703 (2017).
63. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

Acknowledgements

We thank C. Policarpi for advice on CUT&RUN and all members of the Hackett laboratory for experimental support. We thank M. Boulard and P. Avner for critical reading of the manuscript and E. Magnusdottir and A. Aulehla for their contributions to thesis advisory committee discussions. We thank H. Koseki and J. Sharif for kindly sharing floxed *Dnmt1* ESCs. We thank European Molecular Biology Laboratory (EMBL) core facilities and, in particular, those of genetic and viral engineering (J. Sawitzke) and flow cytometry (C. Chaddick), for key experimental assistance. This study was funded by an EMBL program grant to J.A.H.

Author contributions

K.H.G. performed experiments and bioinformatics analysis and contributed to the manuscript. J.A.H. designed and supervised the study, performed bioinformatics analyses and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

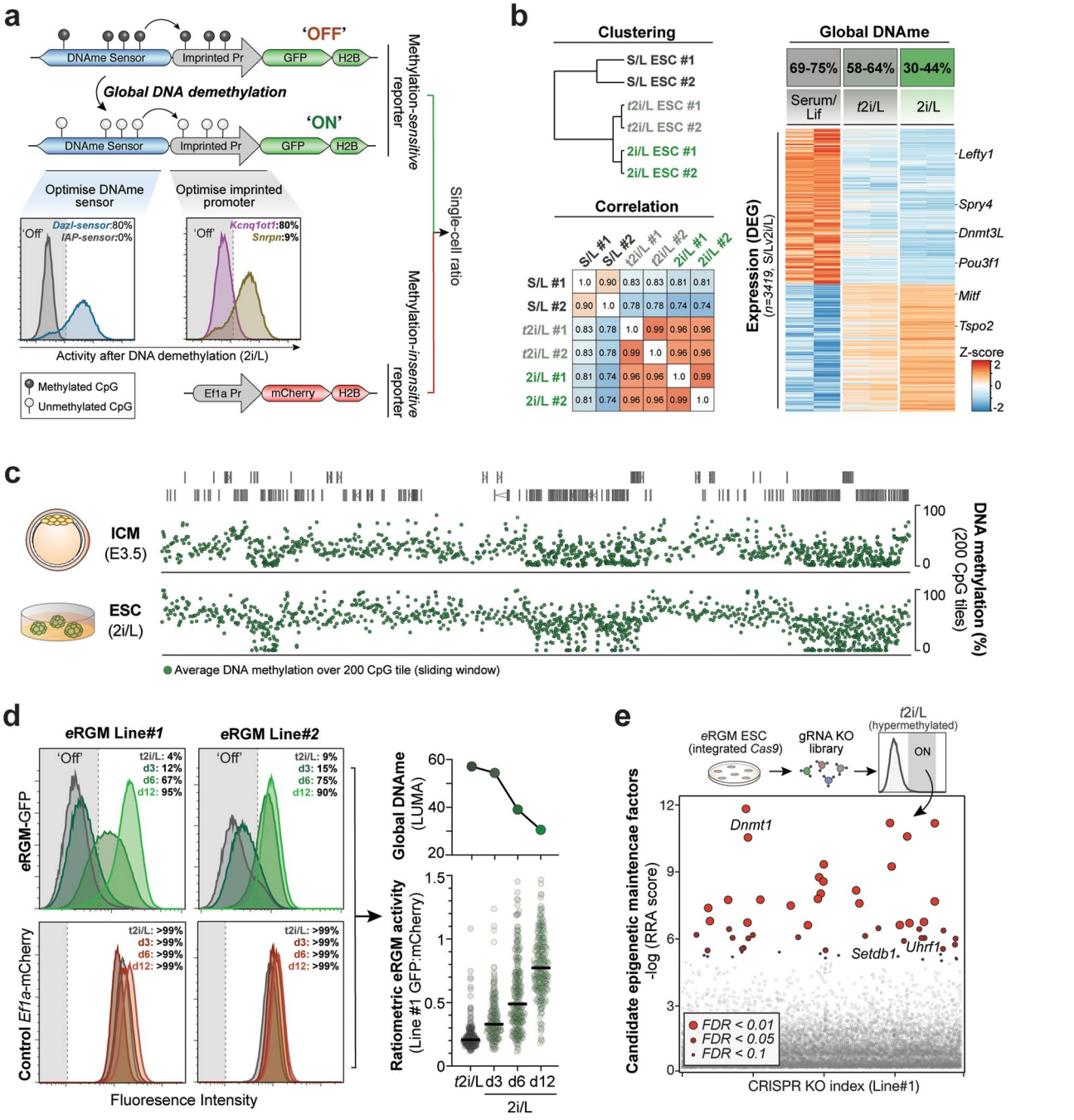
Extended data is available for this paper at https://doi.org/10.1038/s41594-020-0445-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-020-0445-1.

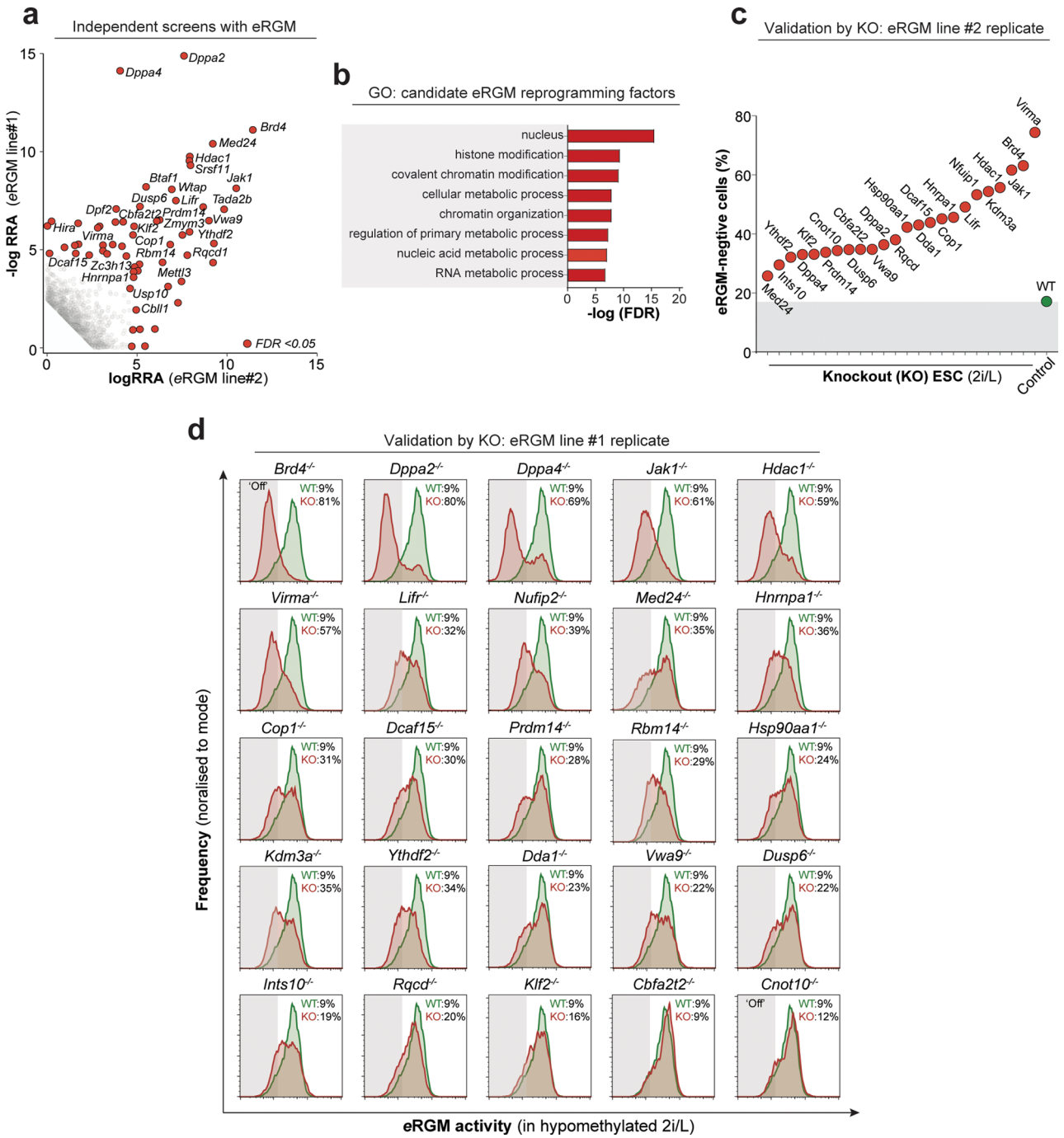
Correspondence and requests for materials should be addressed to J.A.H.

Peer review information Peer reviewer reports are available. Anke Sparmann was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

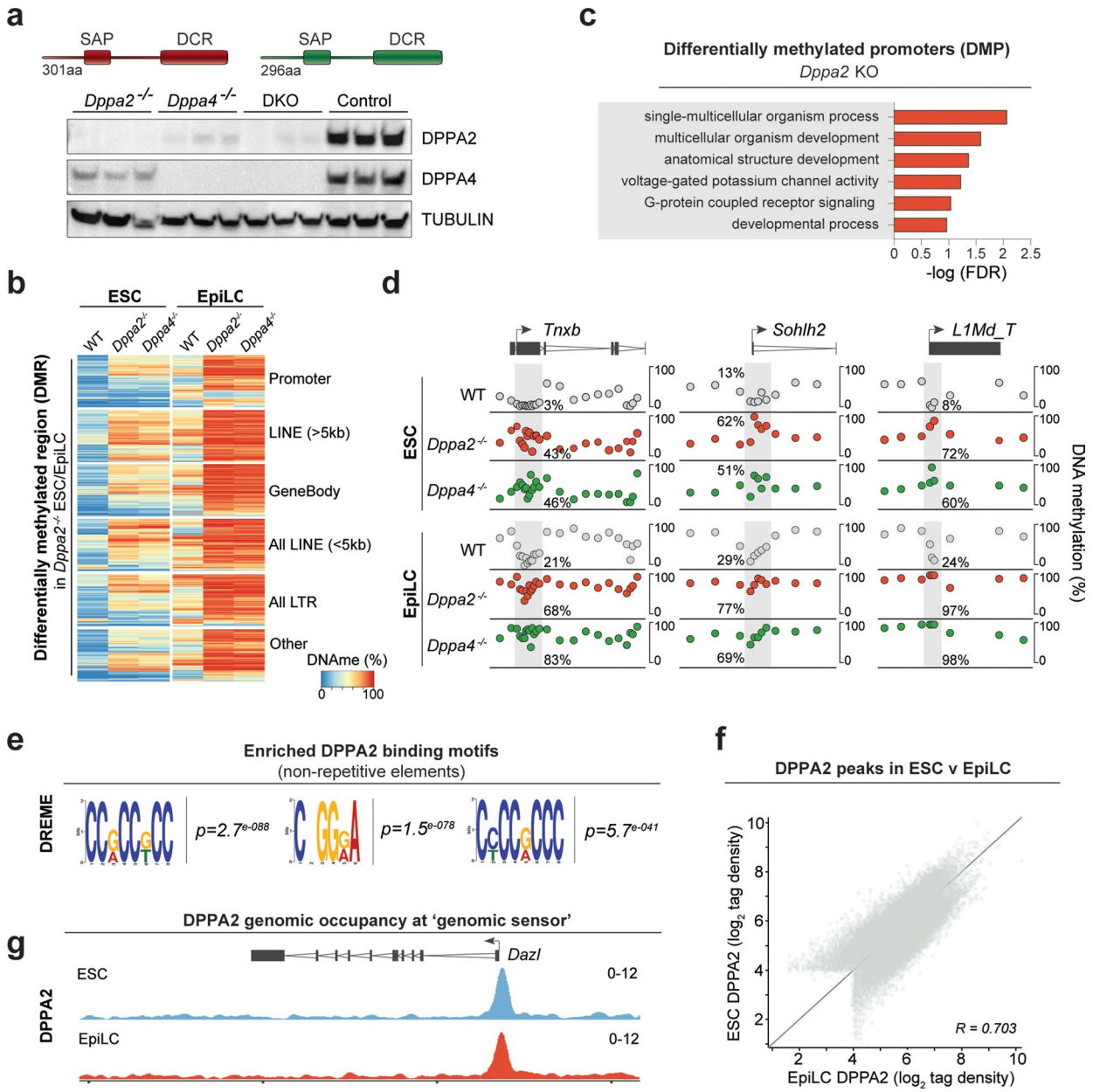
Reprints and permissions information is available at www.nature.com/reprints.



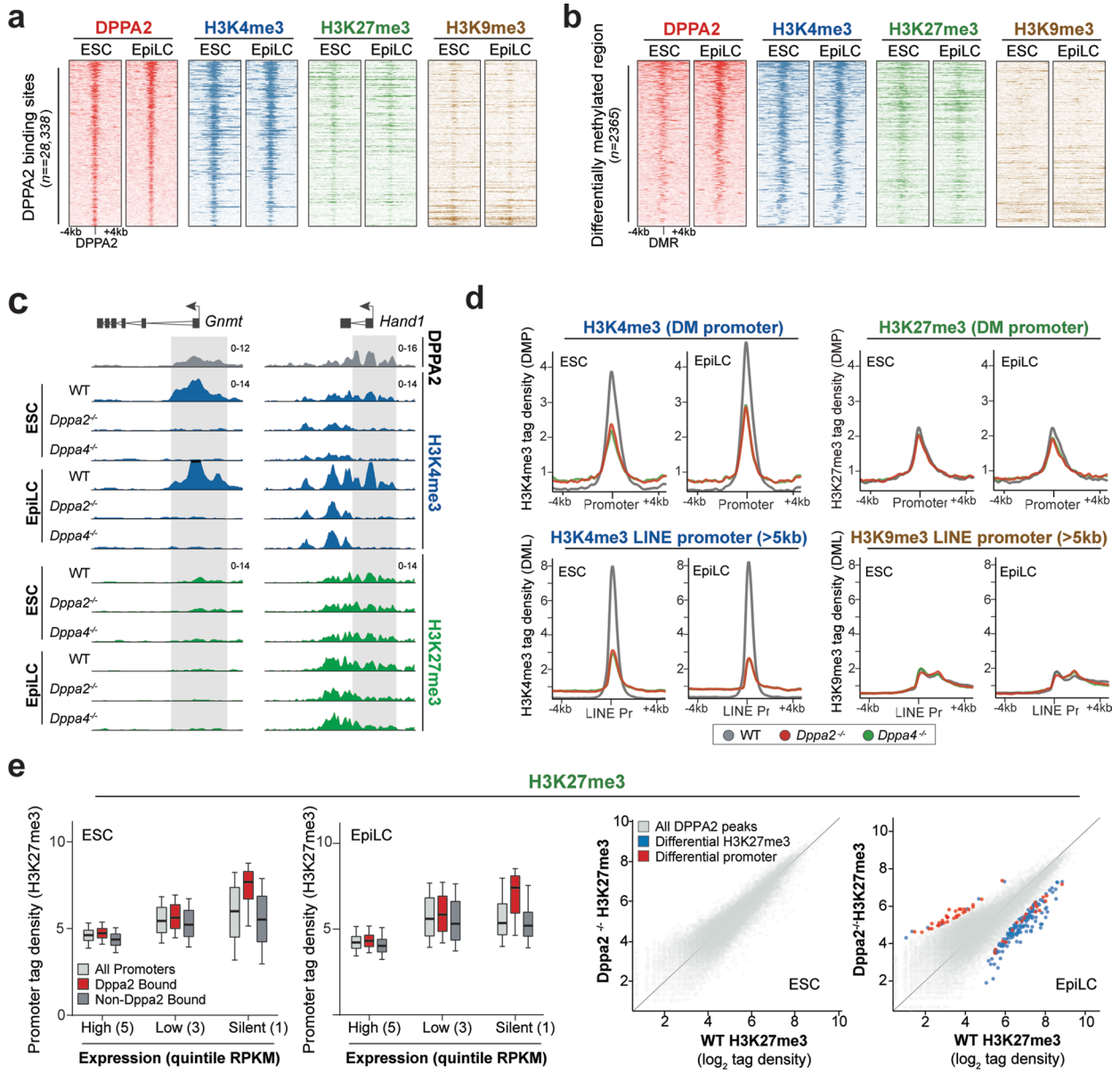
Extended Data Fig. 1 | Model and enhanced ratiometric reporter for developmental DNA demethylation. a, Schematic for design and optimisation of the ratiometric enhanced eRGM cellular DNAme reporter. The system consists of a methylation-*sensitive* imprinted promoter, which controls expression of GFP according to its level of DNA methylation. The DNA methylation level is set by an *antisense* upstream genomic region (DNAme sensor), which acquires a level of DNAme that tracks the global DNA methylation state in the cell. The sensor subsequently adjusts DNAme at the imprinted promoter to equivalence via proximity spreading. This two-stage system generates a robust read-out of global DNAme status (upper panels). Changing the DNAme sensor to a region that is resistant to DNA demethylation (*for example IAP*) prevents eRGM activation in hypomethylated conditions (left FACS plot), confirming that the sensor DNAme status controls activity. Moreover switching the imprinted promoter from *Snrpn* to *Kcnq1ot1* enables a greater degree of expression upon DNA hypomethylation, thereby increasing the dynamic range of the reporter (right FACS plot). Shown in grey is activity in the 'off' hypermethylated state. Finally, by coupling eRGM with a second methylation-*insensitive* reporter (*Ef1a*-mCherry), a single-cell ratiometric score can be generated that normalises for confounding factors. **b**, Transcriptomics from ESC maintained in Serum/Lif (hypermethylated), titrated *t*2i/L (hypermethylated) or 2i/L (hypomethylated). The *t*2i/L and 2i/L transcriptomes are highly comparable despite distinct global methylation states (hyper—and hypo-methylation, respectively), implying transition between these conditions isolates epigenetic resetting without confounding changes in cell identity. **c**, Screen shot of genomic methylation pattern from naïve E3.5 epiblast and naïve ESC demonstrating *in vitro* resetting from *t*2i/L to 2i/L establishes a highly comparable methylome as *in vivo* resetting. **d**, Representative FACS plots of progressive eRGM (GFP) activation by DNA demethylation during 12 day transition from *t*2i/L to 2i/L in independent eRGM cell lines (Line #1 and #2). Normalising to mCherry (lower panel) enables a ratiometric single-cell readout (each datapoint is a single cell) (lower right panel), which closely tracks global methylation levels (upper right). **e**, CRISPR screen for gene KO that enable eRGM activation even under hypermethylated conditions (*t*2i/L) identifies key known DNA methylation and chromatin regulators, confirming eRGM specificity and sensitivity to modulation of epigenetic systems.



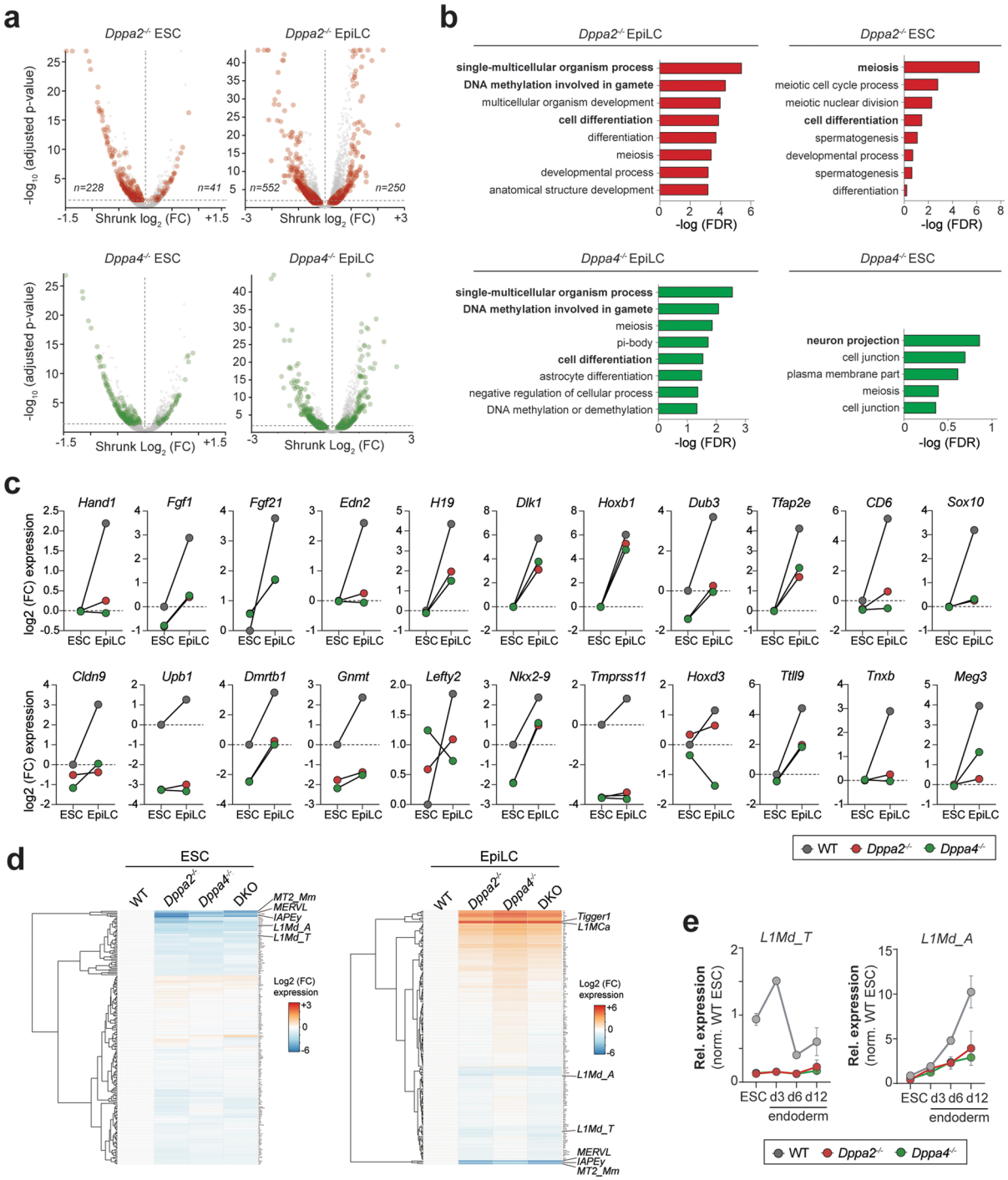
Extended Data Fig. 2 | Validation of CRISPR screen candidates for epigenetic reprogramming. a, Scatter plot showing significance (RRA) values for candidate reprogramming factors from screens of independent eRGM cell lines are highly correlated. A FDR < 0.05 was used as a threshold to identify final candidates. **b**, Gene ontology (GO) analysis of final candidate factors from the screens shows enrichment for nuclear activity, and involvement in chromatin and/or nucleic acid processes consistent with being epigenetic regulators. **c**, Percentage of cells that remain eRGM-negative in hypomethylation-inducing 2i/L upon knockout of the indicated candidate reprogramming factor. Shown is data from KO generated in eRGM line#2, analogous to independent KO in eRGM line#1 (Fig. 2a) **d**, Flow cytometry histograms (GFP on x-axis) demonstrating knockout of most candidates leads to a significant block of eRGM activation amongst single cells, implying altered epigenetic resetting. Shown is percentage single-cells with eRGM-negative 'off' (marked in grey) after 12 days in DNA hypomethylation-inducing 2i/L culture.



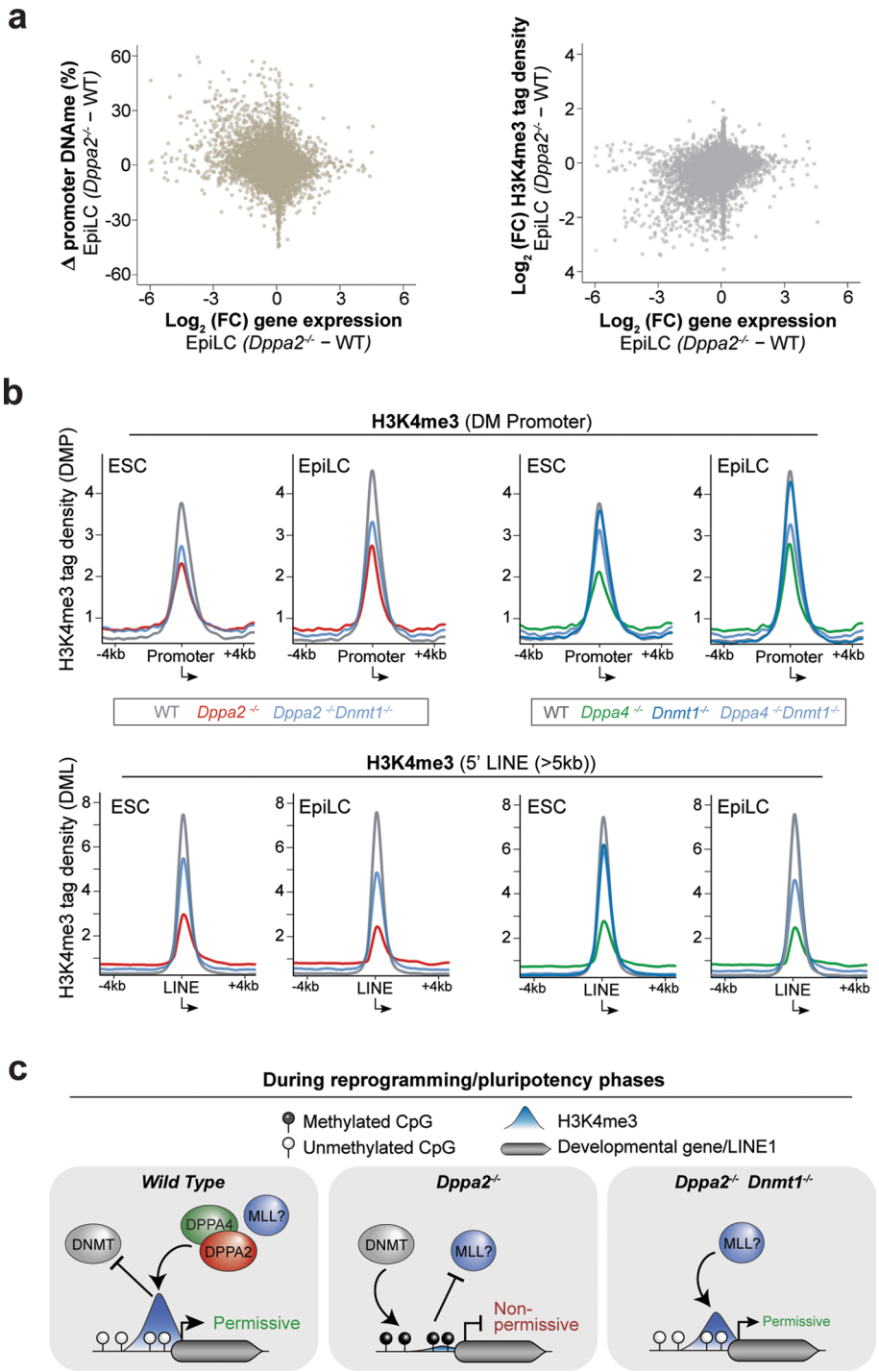
Extended Data Fig. 3 | The DNA methylation landscape in *Dppa2/4* knockout. a, Western blot confirming loss of DPPA2, DPPA4, or both (DKO) protein(s) upon generation of clonal knockout ESC lines. Note loss of DPPA2 protein leads to depletion of DPPA4, and reciprocally, potentially due to disruption of the heterodimeric complex that stabilises each protein. Uncropped image of the blot available as source data. Shown above are schamtics of DPPA2 (red) and DPPA4 (green) **b**, Heatmap showing methylation status of significant differentially methylated regions (from sliding 50 CpG windows) (DMR) in *Dppa2* and *Dppa4* knockout ESC or EpiLC. Note *Dppa4* DMRs are highly correlated with *Dppa2*. **c**, Gene ontology (GO) of genes associated with differentially methylated promoters (DMP) in *Dppa2* KO ESC or EpiLC, determined by direct analysis of +1 kb to -1kb of TSS, reveals enrichment for developmental-associated gene classes. **d**, Genome view showing DNA methylation patterns in WT and *Dppa2/4* KO ESC and EpiLC. Each datapoint represents the windowed average methylation of 15-20 CpG sites. **e**, Top three enriched DPPA2 binding motifs from DREME analysis of DPPA2 binding peaks at non-repetitive elements, reveal preference for GC. **f**, Scatter plot of DPPA2 enrichment at all DPPA2 binding peaks in ESC and EpiLC demonstrating a similar binding pattern of DPPA2 in ESC and EpiLC. **g**, Genome view of DPPA2 CUT&RUN-seq tracks showing that DPPA2 binds at the genomic sensor region used in eRGM (upstream of *Dazl)* in ESC and EpiLC and protects it from *de novo* methylation.



Extended Data Fig. 4 | Chromatin state at DPPA2 binding sites and upon knockout. a, Aligned probe plot showing enrichment of DPPA2, H3K4me3, H3K27me3, and H3K9me3 centered on DPPA2 binding sites +/-4kb in WT ESC and EpiLC. Plots are ordered equivalently, by DPPA2 binding enrichment, which is shown in red. H3K4me3 shows strong enrichment over nearly all DPPA2 sites, whilst H3K27me3 shows modest enrichment. **b**, As in (a), but chromatin enrichment is cantered on differentially methylated regions (DMR). H3K4me3 is enriched in WT ESC at sites prone to hypermethylation upon *Dppa2/4* knockout **c**, Genome view showing H3K4me3 and H3K27me3 in in WT and *Dppa2/4* KO ESC and EpiLC, over representative developmental genes. **d**, Trend plot showing the enrichment of chromatin marks (H3K4me3 and H3K27me3 or H3K9me3) over gene promoters (upper panels) or over full-length LINE1 promoters (lower panels) that acquire DNAme in *Dppa2 KO*. H3K4me3 exhibits a dramatic depletion in *Dppa2* or *Dppa4 KO*, whilst H3K27me3 and H3K9me3 exhibit no significant changes at these loci. **e**, Left: Boxplot showing H3K27me3 enrichment over all, DPPA2-bound and non-DPPA2 bound promoters binned for expression quintile. Box indicates the 25th, median and 75th percentiles, whiskers the 10th to the 90th percentiles. Right: Scatter plot of H3K27me3 enrichment at DPPA2-binding sites in WT and *Dppa2*$^{-/-}$ ESC and EpiLC. Significant differentially H3K27me3 enriched sites, all (blue) and overlapping promoters (red) are highlighted. Significance by LIMMA < 0.01. Showing that unlike H3K4me3, very few DPPA2 binding sites exhibit a change in H3K27me3 upon *Dppa2* knockout.



Extended Data Fig. 5 | Transcriptional and developmental competence upon *Dppa2/4* deletion. a, Volcano plot showing all differentially expressed genes (DEG) in *Dppa2*$^{-/-}$ *and Dppa4*$^{-/-}$ ESC and EpiLC compared to WT. Note many more genes are repressed than activated. **b**, Gene ontology (GO) of DEGs in ESC and EpiLC reveals a strong enrichment for developmental-associated terms, which is driven by silenced developmental genes in *Dppa2/4* KO **c**, Representative examples of developmental genes that fail to activate in *Dppa2/4* KO EpiLC. **d**, Heatmap showing log fold-change of expression normalised to WT of all transposable elements (TE) in WT, *Dppa2*$^{-/-}$, *Dppa4*$^{-/-}$ and DKO ESC and EpiLC. **e**, qRT-PCR quantification of expression of L1Md_T and L1Md_A during ESC to endoderm differentiation in WT, *Dppa2*$^{-/-}$ *and Dppa4*$^{-/-}$ normalised to WT ESC. Data represent mean ± s.d. (n = 2 biologically independent experiments). Confirming that evolutionary young *L1Md_T* and *L1Md_A* exhibit impaired expression in *Dppa2/4* KO, implying *Dppa2/4* are required to maintain competence for TE activity.



Extended Data Fig. 6 | Functional interaction between DNAme, H3K4me3, and gene silencing. a, Scatter plot showing inter-relationships between changes in H3K4me3 and DNA methylation versus gene expression upon *Dppa2 KO* in EpiLC **b**, Trend plots showing the enrichment of H3K4me3 over gene promoters (upper panels) and over full-length LINE1 promoters (lower panels) that acquire DNAme in *Dppa2 KO*. Shown for WT, *Dppa2 KO* (or *Dppa4 KO*), *Dnmt1 KO* and *Dnmt1/Dppa2 KO* (*or Dppa4 KO*) in ESC and EpiLC**. c**, Proposed model of the interplay between H3K4me3 and DNA methylation at *Dppa2/4* targets in regulating gene expression competence.

Appendix

Genome-scale CRISPR Screening for Regulators of Cell Fate Transitions

Valentina Carlini[1,2], Kristjan H. Gretarsson[1,2], Jamie A. Hackett[1,3]

[1] European Molecular Biology Laboratory (EMBL), Epigenetics & Neurobiology Unit, via Ramarini 32, 00015, Rome, Italy.

[2] Equal contribution.

[3] Author for correspondence.

jamie.hackett@embl.it

Running title: CRISPR Screening for Developmental Regulators

Keywords: pluripotency, stem cell, PGC, germ cell, lentivirus, CRISPR, protocol, lineage regulator

Abstract

Knockout CRISPR screening enables the unbiased discovery of genes with a role in any phenotype of interest. The approach couples a genome-scale library of guide RNA (gRNA), the *Cas9* endonuclease, and a faithful phenotypic read-out to systematically identify candidate genes via their loss-of-function effect. Here we provide a detailed description of the CRISPR screen protocol and outline how to apply it to decipher the gene networks that underlie developmental cell fate decisions. As a paradigm we use the *in vitro* model of cell state transition(s) from naive pluripotency to primordial germ cell (PGC) fate, exploiting the *Stella*-GFP:*Esg1*-tdTomato (SGET) mouse ESC line. The principles in this protocol can be readily adapted to characterise lineage regulators for other cell fate models and/or for other species.

1. Introduction

The emergence of CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 has heralded an era of precision editing of genetic sequence to modify or investigate its function. This powerful approach enables the functional role(s) of genes to be rapidly assessed through loss-of-function (LOF) genetics or other strategies, and has proved a landmark technological advancement more generally for biological research [1]. The LOF strategy typically exploits a guide RNA (gRNA) to direct the Cas9 nuclease to a complementary locus, where it induces a precise double-stranded break (DSB). When targeted to coding exons this leads to frameshifting-indels, resulting in homozygous gene 'knockout' due to loss of functional protein. The approach has now been scaled to a genome-wide level by using libraries of gRNA to systematically perturb the function of every gene in the genome in a pooled-cell context (one gRNA per cell) [2-6]. Such knockout 'CRISPR screening' determines whether gene LOF alters cellular response to any given phenotype assay or selection (*e.g.* survival or reporter activity). Any change implies a functional interaction between the gene and the phenotype of interest. The specific gene(s) involved are identified by measuring the enrichment or depletion of gRNAs (indicative of knockouts) within the population after selection, relative to control, via next-generation sequencing (NGS). This enables unbiased forward genetics at unparalleled resolution and consequently, CRISPR screening is a key tool to disentangle the complex interaction between gene function and myriad biological processes, for example, disease, drug response, molecular mechanisms, and cancer [7-11]. Indeed, CRISPR screening has now been adapted to unravel the gene networks that underpin developmental cell fate transitions, including regulators of early embryogenesis and primordial germ cell (PGC) specification [12,13].

PGC are the founding population of the germ cell lineage, which transmits heritable genetic and epigenetic information to the next generation [14,15]. In mammals PGCs arise after a series of cell state transitions that includes transit through naïve pluripotency, and subsequently through a formative state of pluripotency that is primed to give rise to both the somatic lineages and to PGC [16,17]. The decision to form PGC represents a critical developmental event, and is driven by exposure of formative cells to WNT and BMP signalling. This activates key germline genes including *Blimp1, Prdm14, Zfp296* and *Nr5a2* and promotes epigenetic reprogramming [18,19,12,20,21]. Recent progress has recapitulated specification of both human and mouse PGC in an *in vitro* model, by utilising pluripotent embryonic stem cells (ESC) [22-25]. These are induced into a formative pluripotent state (called epiblast-like cells (EpiLC) in mice), which in the presence of appropriate signalling cues (WNT and BMP) can give rise to PGC-like cells (PGCLC). These PGCLC exhibit the key molecular and functional properties of authentic PGC *in vivo* [26,27], and therefore represent an ideal model to investigate the (epi)genetic and molecular mechanisms that underpin the germ cell lineage.

This protocol describes a strategy to exploit unbiased CRISPR screening to identify regulators of cell fate decisions, using mouse PGCLC induction as a paradigm. To denote each cell state transition we use the *Stella*-GFP:*Esg1*-tdTomato (SGET) compound-reporter ESC line [12] (Fig. 1). This enables identification of factors important for naïve pluripotency *per se*, in addition to key regulators for the transition to the formative state (exit from naïve pluripotency), and subsequently for PGC specification. The approach iteratively purifies cells that have successfully acquired the appropriate precursor state for the subsequent developmental transition. Each 'competent' precursor population is then utilised for two purposes: to give rise to the next cell fate, and also as the reference population to identify enriched or depleted gRNAs for that specific cell state transition (and therefore genes with a functional role) (Fig. 2). This iterative strategy distinguishes the approach from the orthogonal comparisons in classical CRISPR screening. Whilst this protocol is focussed on identifying regulators of developmental events towards mouse PGC, it can be adapted for use in human PGC biology or indeed in any epigenetic or developmental transition model for which each successive cell state can be identified (*e.g.* by reporter or cell surface markers). Moreover, here we discuss CRISPR knockout screening but the principles can be readily applied to activation (CRISPRa) or inhibition (CRISPRi) screens to modulate, rather than delete, gene activity.

2. Materials

2.1 Equipment

1. Biosafety level 2 (BSL2) room and tissue-grade laminar flow hood
2. Microcentrifuge
3. High capacity centrifuge
4. Fluorescence activated cell sorter (FACS) facility
5. CO_2 controlled humidified incubator
6. Automated cell counter *e.g.* Countess II
7. Thermal cycler
8. Qubit fluorometer
9. Tapestation system
10. Electrophoresis equipment
11. PCR hood
12. Magnetic stand for PCR tubes
13. Mr. Frosty freezing container

2.2 Reagents

1. SGET embryonic stem cells *(Stella*-GFP:*Esg1*-tdTomato)
2. Lenti-X HEK 293T cells (Takara)
3. Genome-wide sgRNA lentiviral library plasmid (*e.g.* Brie, Addgene #73633)
4. Lentiviral packaging plasmid (*e.g*. psPAX2, Addgene #12260)
5. Lentiviral envelope plasmid (*e.g*. pMD2.G Addgene #12259)
6. Lenti-X Concentrator (Takara)
7. *Stable* competent *E. coli* cells (High efficiency) (New England Biolab)
8. Plasmid maxiprep kit (*e.g* Zymopure II plasmid maxiprep)
9. Luria Bertani (LB) broth
10. Ultra-low attachment microwell plate (*see* **Note 1**)
11. 6- and12-well cell culture plates
12. T225 filter-top culture flasks
13. Stericup filter unit, 0.22µM
14. Tissue culture grade phosphate-buffered saline (PBS) solution
15. NDIFF 227 (N2B27)
16. GMEM
17. OPTI-MEM
18. Knockout serum replacement (KSR)
19. Foetal bovine serum (FBS)

20. Dimethyl sulfoxide (DMSO)
21. Penicillin/Streptomycin
22. Puromycin
23. Ampicillin
24. Polybrene
25. Fibronectin
26. Gelatin
27. TrypLE
28. PD0325901 (MEK inhibitor)
29. CHIR99021 (GSK3 inhibitor)
30. Non-essential amino acids (NEAA)
31. Sodium pyruvate
32. β-mercaptoethanol
33. L-glutamine
34. Activin-A
35. Basic fibroblast growth factor (bFGF)
36. Leukaemia inhibitory factor (LIF)
37. Bone morphogenetic protein 4 (BMP4) (R&D Systems)
38. Bone morphogenetic protein 8 (BMP8) (R&D Systems)
39. Epidermal growth factor (EGF), (R&D Systems)
40. Stem cell factor (SCF) (R&D Systems)
41. RNaseA, (DNase- and Protease- free)
42. Genomic DNA extraction kit genomic DNA, for > 1mio cells (*e.g.* DNeasy Blood and Tissue kit, Qiagen)
43. Genomic DNA extraction kit, for < 1mio cells (*e.g* Quick-DNA micro-prep kit, small elution volume >10 ul, Zymo)
44. Qubit dsDNA BR Assay Kit
45. Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolab)
46. P5 stagger F primers (*see Section 2.4*)
47. P7 indexed R primers (*see Section 2.4*)
48. Ultrapure nuclease Free water
49. D1000 Reagents (Agilent) & D1000 ScreenTape (Agilent)
50. SPRI beads (Beckman Coulter)
51. 80 % EtOH (freshly prepared)
52. MAGeCK analysis tool (https://sourceforge.net/p/mageck/wiki/Home/)
53. Cutadapt analysis tool (https://cutadapt.readthedocs.io/en/stable/)

2.3 Cell culture media

1. **ESC culture media**: NDIFF 227 supplemented with 1μM PD0325901, 3μM CHIR99021, 1000 U/ml LIF, 1% Penicillin/Streptomycin (*see* **Note 2**). Pass through 0.22μM filter unit.

2. **EpiLC culture media**: NDIFF 227 supplemented with 1% Knockout serum replacement, 20 ng/ml Activin-A, 12.5 ng/ml bFGF, 1% Penicillin/Streptomycin. Pass through 0.22μM filter unit.

3. **PGCLC culture media**: GMEM supplemented with 15% Knockout serum replacement, 0.1 mM Non-essential amino acids, 1 mM Sodium pyruvate, 1% Penicillin/Streptomycin, 0.1mM B-mercaptoethanol, 1mM L-glutamine, 500 ng/ml BMP4. Add the following cytokines just before use: 1000 U/ml LIF, 100 ng/ml SCF, 500 ng/ml BMP8a (*see* **Note 3**), 50 ng/ml EGF. Pass through 0.22μM filter unit.

4. **293T media** (viral production): DMEM supplemented with 10% FBS, 1% Penicillin/Streptomycin, 6mM L-Glutamine, 25mM HEPES, 1mM Sodium pyruvate.

2.4 NGS Oligos for library preparation

see **Note 4**

Oligo Name	Sequence (5'-3')
P7 Index A1	CAAGCAGAAGACGGCATACGAGATCGGTTCAAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P7 Index A2	CAAGCAGAAGACGGCATACGAGATGCTGGATTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P7 Index A3	CAAGCAGAAGACGGCATACGAGATTAACTCGGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P7 Index A4	CAAGCAGAAGACGGCATACGAGATTAACAGTTGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P7 Index A5	CAAGCAGAAGACGGCATACGAGATATACTCAAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P7 Index A6	CAAGCAGAAGACGGCATACGAGATGCTGAGAAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P7 Index A7	CAAGCAGAAGACGGCATACGAGATATTGGAGGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT

P7 Index A8	CAAGCAGAAGACGGCATACGAGATTAGTCTAAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT
P5 0bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTTTGTGGAAAGGACGAAACACCG
P5 1bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTCTTGTGGAAAGGACGAAACACCG
P5 2bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTGCTTGTGGAAAGGACGAAACACCG
P5 3bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTAGCTTGTGGAAAGGACGAAACACCG
P5 4bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTCAACTTGTGGAAAGGACGAAACACCG
P5 6bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTTGCACCTTGTGGAAAGGACGAAACACCG
P5 7bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTACGCAACTTGTGGAAAGGACGAAACACCG
P5 8bp stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTGAAGACCCTTGTGGAAAGGACGAAACACCG

P5/P7 flow cell sequences

Illumina sequencing primer

Vector binding sequence

Stagger sequence

Barcode sequence

3. Methods

The method described here is optimised for the *Brie pooled gRNA library* [5], which contains 78,637 gRNA that target 19,674 genes in conjunction with streptococcus pyogenes (sp) *Cas9*, and also carries puromycin resistance. The antibiotic selection, number of cells to maintain in culture, and *Cas9* variant described here are specific to the gRNA library and should be optimised if using an alternative.

3.1 Preparation of the lentiviral gRNA library.

A lentiviral preparation can be obtained directly from suppliers (*e.g. Addgene*) or can be generated in a BSL2 facility, as below. Ensure all institutional biosafety guidelines relating to lentiviral usage and CRISPR are adhered to, and the appropriate personal protective equipment (PPE) is in use.

This section will take approximately five days.

1. If necessary, amplify the gRNA library plasmid by transforming 100ng into each of four aliquots of stable competent *E.coli* (0.05 ml each), following the manufacturer's instructions. Expand each outgrowth directly into 100 ml LB with 50µg/ml ampicillin for 14-16 h at 37°C. Maxiprep the library plasmid using endotoxin-free guidelines (*see* **Note 5**). Confirm that the amplified library has maintained sgRNA complexity and representation by NGS sequencing *before* lentiviral preparation (*see* **Note 6**).

2. Day 1. To generate lentiviral particles, seed $1.5x10^7$ Lenti-X HEK 293T cells into each of 5x T225 filter cap flasks, with 40 ml 293T medium (total of $7.5x10^7$ cells). Incubate in a humidified 5% CO_2 chamber at 37°C for approximately 24 hours.

3. Day 2. Transfect cells (when they reach 70-80% confluence) by generating a plasmid mix consisting of 26 µg *pPax2* plasmid, 12 µg *pMD2.G* plasmid, and 32 µg *Brie gRNA library* plasmid in 2 ml Opti-MEM per T225 flask. Additionally prepare 150 µl of lipofectamine in 2 ml Opti-MEM per T225 flask. Combine the plasmid and lipofectamine preparations together and mix well to create the transfection mix (4 ml total per T225 flask). Incubate at room temperature for 20 minutes, then pipette the 4ml dropwise into each flask. Return flasks to incubator (*see* **Note 7**).

4. Six hours post-transfection aspirate the media and replace with 40 ml of fresh 293T media.

5. Day 4. Harvest the first batch of lentivirus-containing supernatant (media) 48 hours after transfection (caution: supernatants contain live lentivirus). Remove medium from each T225 flask, pool and filter with a 0.22µm low protein binding filter unit. Store at 4°C (*see* **Note 8**).

6. Add 40ml of fresh 293T media and incubate for an additional 24 hours at 37℃ with 5% CO_2.

7. Day 5. Harvest the second batch of viral containing supernatant from each flask 72 hours after transfection, pass through a 0.22µm filter unit, and pool with previous harvest (*see* step 5 in section 3.1).

8. In order to concentrate the viral particles, transfer all clarified supernatants from step 7 (media) to a sterile container and combine 1 volume of Lenti-X Concentrator with 3 volumes of supernatant. Mix by gentle inversion and incubate the mixture at 4°C for 30 minutes to overnight.

9. Centrifuge the sample(s) at 1,500 g for 45 minutes at 4°C. After centrifugation, an off-white pellet will be visible.

10. Carefully remove supernatant, taking care not to disturb the pellet. Residual supernatant can be removed with a pipette after a brief centrifugation at 1,500 g. Gently resuspend the pellet in ~10 ml of NDIFF227 and aliquot the concentrated lentivirus into labelled 1.5 ml screw-cap tubes. Store viral aliquots at -80°C. If stored for more than 6 months recalculate the infection efficiency of the lentiviral preparation (*see* section 3.2) as prolonged storage might result in infectivity loss.

3.2 Optimization of Transduction Efficiency

It is critical to assess the infection efficiency of the lentiviral preparation. The objective is to identify the optimised titration ratio such that cells receive a single viral particle on average (and therefore a single integrating gRNA): suggested infection efficiency is 30-50%.

This section will take approximately six days:

1. Day 1. Pre-coat a 12-well plate (12WP) with 0.1% gelatin for 1 hour, aspirate and allow to air-dry. Seed $1.5x10^5$ SGET mESC into nine wells of the 12WP (n=9) with 800µl ESC culture media per well.

2. Day 2. Replace with fresh ESC culture media. Rapidly thaw a lentiviral aliquot at 37°C (*see* step 10 in section 3.1) then keep on ice. Transduce the cells across a titration curve by adding the viral

supernatant to each well in varying amounts (for example: 300 µl, 100 µl, 30 µl, 10 µl, 3 µl, 1 µl), leaving two wells virus-free (0 µl) (non-transduced controls) (n=8) (*see* **Note 9**). The cells in the final well should be collected and counted on this day 2, giving the number of cells that were transduced.

3. Day 3. Aspirate the virus and wash five times with PBS. Add 800µl of fresh ESC media containing puromycin (1.2 µg/ml final concentration) to the transduced wells (n=6) and one non-transduced well (*selection control,* n=1) to start the antibiotic selection. Additionally, add 800µl of ESC media without puromycin to the final non-transduced well (*background control,* n=1).

4. Day 5-6. Once no viable cells remain in the *selection control* (non-transduced) well, accurately count *live* cells in all wells using a cell counter such as countess II. Divide the number of cells in each transduced well by the total number of cells in the *background control* (without puromycin) to calculate the infection efficiency of the viral supernatant (*see* **Note 10**).

5. Identify the ratio of viral supernatant to number of cells that gives an infection efficiency (puromycin resistance) of 30-50%.

3.3 Lentiviral transduction into SGET ESC

This section describes how to generate an ESC line carrying an integrated genome-wide gRNA library to enable subsequent phenotype screening. It is essential that the cells express *Cas9.* The SGET ESC described here already carry an integrated single-copy of constitutively expressed *Cas9*.

The health of the target ESC cell line is critical for obtaining accurate results. Cells should be checked regularly for mycoplasma, have less than 25 passages, proliferate well with no overt signs of differentiation, and be routinely passaged before sub-confluence with regular (daily) media changes ('feeding').

This section will take approximately two weeks:

1. Day 1. Pre-coat 3 T225 flasks and one well of a 12-well plate with 0.1% gelatin for 1 hour, aspirate and allow to air-dry. The day before transduction seed $1x10^7$ SGET mESC into each of the 3 gelatin pre-coated T225 flasks with 40ml ESC culture media. In parallel plate $1.5x10^5$ SGET mESC into one well of a 12-well plate to estimate the proliferation rate of your cells needed for step 2 in section 3.3 (*growth control*).

2. Day 2. Count the cells from the *growth control* and use the proliferation rate to estimate the total number of cells in each T225 flask to be transduced. Rapidly thaw the appropriate number of lentiviral aliquots at 37°C, then keep on ice. Add the optimized volume of lentiviral supernatant to cells calculated at steps 4-5 in section 3.2 to obtain an infection efficiency of 50% (*see* **Notes 10** and **11**).

3. Day 3. Aspirate the media and wash the mESC 5x times with PBS. Add fresh ESC culture media supplemented with puromycin (1.2 µg/ml final).

4. Day 4-9. Perform daily media changes and maintain the antibiotic selection for 7 days. Cells should be passaged when/if they reach 70% confluence. It is critical to passage sufficient cells to maintain coverage of the gRNA library (*see* **Notes 10**, **12** and **13**).

5. Day 10-12. Following selection, the knockout SGET ESC population should be maintained in T225 flasks in ESC media without puromycin for a further 3 days to allow recovery/induce knockout, ensuring >$3.2x10^7$ cells are passaged. Cryofreeze aliquots of $5x10^7$ cells (*see* **Note 14**) and proceed to induction of EpiLC-PGCLC for CRISPR screening when convenient (directly or later from frozen cells) (*see* section 3.4).

6. Day 12. Before continuing with section 3.4 confirm that naïve SGET ESC are >90% double-positive (SG^+ET^+) using fluorescence cytometry (Fig. 1).

[Insert Figure 1 here]

3.4 Induction of EpiLC and PGCLC with CRISPR library

This section describes how to induce EpiLC and PGCLC from naïve ESC, and when to collect samples for screen analysis and validation.

This section will take approximately 8 days:

1. Day 1. After step 5 in section 3.3, split the transduced SGET mESC by washing 2 times with PBS and dissociating into single-cells with 4 ml of TrypLE per T225, incubate at 37°C for 5 minutes and then resuspend in ESC media to dilute out the TrypLE. Count the cells and set aside the 'ESC' sample of >$3.2x10^7$ cells for later genomic DNA extraction (Fig. 2) (*see* **Note 15** and section 3.5).

2. Induce EpiLC from the remaining SGET mESC by plating $1.2x10^7$ cells into each of 3 T225, pre-coated with fibronectin (*see* **Note 16**), and with 40 ml of fresh EpiLC culture media. Change the media daily.

3. Day 3. After 42 hours of EpiLC differentiation, wash the cells with PBS and dissociate using TrypLE (see step 1 section 3.4). Resuspend EpiLC in PBS with 1% KSR. Use fluorescent activated flow cytometry (FACS) to isolate the EpiLC population that has successfully transited to the formative pluripotent state (SG^-ET^+) and that which has failed to exit naïve pluripotency (SG^+ET^+) (Fig. 2).

4. Set aside a fraction of both the purified 'SG^-ET^+ EpiLC' sample and the 'SG^+ET^+ EpiLC' sample for genomic DNA extraction (*see* **Note 17)**. Pellet cells by centrifugation, remove the supernatant, and flash freeze. Store at -80C for later DNA extraction (*see* **Note 18** and section 3.5).

5. For induction of PGCLC: Plate $1.5x10^6$ of the purified SG^-ET^+ EpiLC per well of an ultra-low attachment microwell 6-well plate, using 3ml of PGCLC media per well. A total of at least 23 wells should be used to ensure >400x coverage (>$3.2x10^7$ cells).
   This will induce the formation of 2,400 individual 'embryoids' per well (approximately 55,000 embryoids in total).
   As a control, setup one well of SG^-ET^+ EpiLC in PGCLC media without cytokines (without BMP4, BMP8, SCF, EGF and LIF), which should not be permissive for PGCLC induction (SG^-ET^-).

6. Day 4-9. Perform a half-media change every day, taking care not to disturb the embryoids in each well.

7. Day 9. Six days after PGCLC induction, collect all media and spin down embryoids at 200 g for 4 minutes. Aspirate media and dissociate embryoids resuspending them with TrypLE (use 100 µl of TrypLE for each well of a 6-well) and incubating at 37°C for 5 minutes. Before proceeding, ensure they are resolved into single-cells by trituration. Resuspend cells in PBS with 1% KSR.

8. Isolate PGCLC by fluorescent activated flow sorting (FACS), gating for SG^+ET^{low} (Fig. 2). These correspond to authentic PGC-like cells. Additionally, for control analysis, flow sort cells that have acquired somatic fate (SG^-ET^-).

9. Pellet purified populations at 2000 g for 5 minutes, remove supernatant, and either flash freeze or proceed straight to DNA extraction (*see* section 3.5).

[Insert Figure 2 here]

3.5 Library Preparation

This section describes extraction of genomic DNA from each isolated population and amplification of all gRNA sequences within for NGS analysis. Specific indexed oligo sequences are used to denote each sample being amplified.

This section will take 1-2 days:

1. Perform genomic DNA (gDNA) extraction of each isolated population from the PGCLC cell fate transition protocol (*see* section 3.4). Minimally this should include:
    a. naïve ESC
    b. $SG^{-}ET^{+}$ EpiLC (competent)
    c. $SG^{+}ET^{+}$ EpiLC (failed to exit naïve pluripotency)
    d. $SG^{+}ET^{low}$ PGCLC (authentic PGC)
    e. $SG^{-}ET^{-}$ soma

Isolate genomic DNA with the DNeasy Blood and Tissue kit (for > 1mio cells) or the Quick-DNA micro-prep kit (for <1mio cells), following the manufacturers' instructions. It is important to perform the RNAse step. Accurately quantify DNA concentration with a Qubit dsDNA BR Assay Kit.

2. Amplify gRNA from each sample using *all* the purified gDNA (up to 60 µg) by PCR. Additionally, if not already performed, amplify the original library plasmid. For each sample use a unique 'indexed' P7 primer. For every sample use an equal ratio mastermix of all eight 'stagger' P5 primers (*see* **Note 19**). We suggest to prepare the PCR reactions inside a clean PCR hood.

    Set up each PCR reaction in 50 µl final volume (*see* **Notes 20** and **21**).
    a. 1 µg gDNA (or 50 ng plasmid)
    b. 25 µl Q5 High-Fidelity Master Mix (2X)
    c. 2.5 µl staggered P5 primer mastermix (10 µM)
    d. 2.5 µl unique P7 primer (10 µM) (indexed for each sample)
    e. nuclease free water up to 50 µl.

    Perform the PCR in a thermal cycler as follows:
    a. 98°C for 2 minutes
    b. 98°C for 10 seconds
    c. 60°C for 20 seconds
    d. 72°C for 30 seconds

e. 72°C for 2 minutes

Repeat from steps b. to d. for 22 cycles (*see* **Note 22**)

3. Use SPRI beads to purify and size-select the gRNA library amplicon by removing the remaining genomic DNA and any primer-dimers with a magnetic stand following the manufacturer's protocol. Follow the instructions for double size selection: first use a 0.5x SPRI bead volume to sample volume ratio, then a 1.2x final ratio. Elute in 20 µl of nuclease free water.
4. Check the purified gRNA library amplicons after size-selection using a tape station automated electrophoresis system or a 2% agarose gel. The expected amplification product is 350bp-358bp (depending on the stagger).
5. Quantify and pool together an equal amount of the purified samples into a multiplexed library. Perform next generation single-end 50 (SE50) sequencing, using standard Illumina primers. PhiX should be spiked in to >10% to increase library diversity, whilst optional dark cycles can also be used.

3.6 CRISPR Screen Data Analysis

For counting and analysing gRNA frequency (indicative of knockout frequency) we use the Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout (MAGeCK) tool [28]. The program takes in .fastq files containing sequencing reads of the gRNA, and generates a count table. Normalised count tables of control and treated groups are analysed using the MAGeCK negative binomial distribution significance test to generate a list of enriched or depleted sgRNAs. The relative ranking algorithm (RRA) is then used to identify genes exhibiting a significant difference across multiple targeting gRNAs between control (*e.g.* "$SG^{-}ET^{+}$ EpiLC") precursors, and the next fate transition (*e.g.* "$SG^{+}ET^{low}$ PGCLC"). Below are some suggested steps for an initial overview analysis.

Because the P5 primers contain stagger sequences, the variable sequence of nucleotides before the gRNA sequence needs to be removed using the CutAdapt tool [29]. Here the following example .fastq files are used:

a. *"SGET_EpiLC.fastq" (sequenced $SG^{-}ET^{+}$ EpiLC)*
b. *"SGET_PGCLC.fastq" (sequenced $SG^{+}ET^{low}$ PGCLC)*

1. Remove the vector and stagger sequence with CutAdapt by running the following command in the terminal:
   a. *cutadapt -g TTGTGGAAAGGACGAAACACCG -SGET_EpiLC_trimmed.fastq SGET_EpiLC.fastq*
   b. *cutadapt -g TTGTGGAAAGGACGAAACACCG -SGET_PGCLC_trimmed.fastq SGET_PGCLC.fastq*

2. Download the list of gRNA used in the library [5] and make a new text file ("BrieLibraryControl.txt") with 3 columns: "gRNA_Name", "gRNA_Target_Sequence" and "Target_Gene_Symbol".

3. Generate normalized (to total read counts) count tables ("CountTable_EpiLC_PGCLC") with MAGeCK with the following command:
    a. *mageck count -l BrieLibraryControl.txt –n CountTable_EpiLC_PGCLC --sample-label SGET_EpiLC, SGET_PGCLC --fastq SGET_EpiLC_trimmed.fastq SGET_PGCLC_trimmed.fastq --norm-method total*

4. With the normalized read count table generated from step 3, section 3.6 ("CountTable_ EpiLC_PGCLC.count_normalized.txt") it is now possible to compare the conditions using the test command:
    a. *mageck test -k CountTable_EpiLC_PGCLC.count_normalized.txt -t SGET_PGCLC -c SGET_EpiLC -n PGCLC_LOF_results --pdf_report*

The PDF report generated will reveal the top hits from the screen and changes in single gRNA frequency between the two populations (*see* **Note 23**). Alternatively, visualization of the screening data can be generated using log transformed RRA values or p-values from the readout file (PGCLC_LOF_result.gene_summary.txt) against all genes (in a custom order).
Ideally the top hits should have false discovery rate (FDR) less than 0.05, at least three enriched gRNAs ('goodsgRNA') and log fold change higher than 2. Significant hits should be *technically* validated by introducing the corresponding gene knockout in the SGET cells and registering the effect on PGCLC development.

4. Notes

1. For this protocol, we are using Ultra-low attachment microwell plate from Iwaki, Cat. No. 4810-900. Each well contains 2400 ultra-low attachment micro-wells that facilitate efficient and high-throughput embryoid body formation for PGCLC specification.
2. Optional: addition of 1% Knockout serum replacement (KSR) can improve ESC adhesion.
3. BMP8a and BMP8b are interchangeable.
4. We recommend use of high-quality ultramers from IDT. Take special care to avoid cross-contaminating indexed primers via aerosols.

5. A minimum of $5x10^7$ total colony forming units (CFU) is necessary to maintain library complexity. Assess via plating a 10,000-fold dilution on LB agar. Do not overgrow bacteria since this promotes recombination.
6. Before proceeding to generate lentiviral particles (*see* step 2 in section 3.1), we recommend to confirm that the amplified library plasmid has maintained sgRNA complexity and representation by NGS sequencing. To do this, follow from step 2 in section 3.5 using the library plasmid as input, with the expectation that >90% of gRNAs should be within one logarithmic range. Notably, the vector library gRNA frequency counts also act as a control for some final analyses.
7. Transfection can alternatively be performed with other strategies such as with linear PEI (25kDA) to reduce costs.
8. When filtering, use only cellulose acetate or polyethersulfone (PES) (low protein binding) filters. Do not use nitrocellulose filters, which bind surface proteins on the lentiviral envelope.
9. Polybrene (Hexadimethrine bromide) can be added to the culture at a final concentration of 8ug/ml to enhance the efficiency of transduction [30], albeit we observe only negligible effect.
10. A 30-50% infection efficiency ensures that most cells receive only one gRNA. The transduction needs to be at a scale to achieve a gRNA library coverage of 400x in the final transduced cell population. For example, to achieve coverage of 400x with infection efficiency of 50% using the *Brie pooled library* (which has 78,637 gRNAs), a total of $6.3x10^7$ cells need to be infected.
11. To maintain coverage of the screening library it is suggested to have at least $2.1x10^7$ SGET in each T225 ($6.3x10^7$ cells in total are needed assuming an infection efficiency of 50%).
12. Generally, at least 4-7 days are required for the *Cas9*-induced genetic perturbation to manifest efficiently throughout the population.
13. To maintain a sufficient sgRNA library coverage of 400x, ensure a minimum number of $3.2x10^7$ SGET cells are passaged/maintained at any time point during culture.
14. To cryofreeze cells wash them with PBS and dissociate with 4ml of TrypLE per T225 flask. Incubate at 37°C for 5 minutes and resuspend in ESC media to dilute out the dissociating reagent. After counting, spin down $5x10^7$ cells per aliquot and resuspend them in 1,5 ml freezing media (FBS with 10% DMSO). Transfer cells to cryotubes and allow them to cool slowly (1°C/min) at -80°C in a Mr. Frosty cryofreezing container. After 24 hours transfer them to long-term storage in liquid nitrogen.
15. The ESC sample cells collected for DNA extraction can be pelleted by centrifugation, the supernatant removed, and flash frozen. Store cell pellet at -80°C until all samples are obtained.

16. Flasks should be coated with fibronectin (1:60 in PBS) between 6-24 hours prior to cell seeding. Aspirate just before use.
17. The $SG^{-}ET^{+}$ EpiLC sample needs to be divided for both DNA extraction and PGCLC induction (*see* steps 4 and 5 in section 3.4).
18. Ensure an appropriate number of cells (>$3.2x10^{7}$) to maintain library coverage are retained for DNA extraction.
19. You should generate a mastermix of the 8x P5 F primers (10 µM final concentration). Each individual P5 stagger contains a region of different length, which is important to maintain sequence diversity across the flow cell during NGS, by offsetting identical reads coming from the amplicon. The P7 indexed primers contain a unique barcode used to multiplex samples. Use a different P7 primer for each given sample. Both the P5 and P7 primers contain a region to be bound to the flow cell and vector binding sequence.
20. PCR reaction volume can be scaled up or down according to the needs, however we found that performing the PCR in >100 µl volume was inefficient.
21. If less than 4 µg of DNA is obtained from a sample ensure at least 4 parallel PCR reactions are still setup (each with reduced DNA), to minimise stochastic amplification bias. Additional amplification cycles may be necessary.
22. We suggest you test PCR conditions in advance to optimise number of PCR cycles.
23. A good practice when working with new screening data is to see if the genome-wide knockout has been successful. Comparing the distribution of the gRNAs from the plasmid library (control) to the distribution of the gRNAs from the initial transduced ESC cell population (sample) should reveal depletion of essential genes (*e.g.* ribosomal genes such as *Rps5*) and often, enrichment of the tumor suppressor *p53* is also observed.

REFERENCES

1. Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346 (6213):1258096. doi:10.1126/science.1258096
2. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157 (6):1262-1278. doi:10.1016/j.cell.2014.05.010
3. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science 343 (6166):84-87. doi:10.1126/science.1247005
4. Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, Sanjana NE, Zhang F (2017) Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat Protoc 12 (4):828-863. doi:10.1038/nprot.2017.016
5. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol 34 (2):184-191. doi:10.1038/nbt.3437

6. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 32 (3):267-273. doi:10.1038/nbt.2800
7. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol 35 (5):463-474. doi:10.1038/nbt.3834
8. Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, Mupo A, Grinkevich V, Li M, Mazan M, Gozdecka M, Ohnishi S, Cooper J, Patel M, McKerrell T, Chen B, Domingues AF, Gallipoli P, Teichmann S, Ponstingl H, McDermott U, Saez-Rodriguez J, Huntly BJ, Iorio F, Pina C, Vassiliou GS, Yusa K (2016) A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep 17 (4):1193-1205. doi:10.1016/j.celrep.2016.09.079
9. Ruiz S, Mayor-Ruiz C, Lafarga V, Murga M, Vega-Sendino M, Ortega S, Fernandez-Capetillo O (2016) A Genome-wide CRISPR Screen Identifies CDC25A as a Determinant of Sensitivity to ATR Inhibitors. Molecular cell 62 (2):307-313. doi:10.1016/j.molcel.2016.03.006
10. Fukuda K, Okuda A, Yusa K, Shinkai Y (2018) A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. Genome Res 28 (6):846-858. doi:10.1101/gr.227280.117
11. Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosh I, Sanjana NE, Shalem O, Satija R, Raychowdhury R, Mertins P, Carr SA, Zhang F, Hacohen N, Regev A (2015) A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. Cell. doi:10.1016/j.cell.2015.06.059
12. Hackett JA, Huang Y, Günesdogan U, Gretarsson KA, Kobayashi T, Surani MA (2018) Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. Nature Communications 9 (1):4292. doi:10.1038/s41467-018-06230-0
13. Li M, Yu JSL, Tilgner K, Ong SH, Koike-Yusa H, Yusa K (2018) Genome-wide CRISPR-KO Screen Uncovers mTORC1-Mediated Gsk3 Regulation in Naive Pluripotency Maintenance and Dissolution. Cell Rep 24 (2):489-502. doi:10.1016/j.celrep.2018.06.027
14. Tang WW, Kobayashi T, Irie N, Dietmann S, Surani MA (2016) Specification and epigenetic programming of the human germ line. Nat Rev Genet 17 (10):585-600. doi:10.1038/nrg.2016.88
15. Hackett JA, Zylicz JJ, Surani MA (2012) Parallel mechanisms of epigenetic reprogramming in the germline. Trends Genet 28 (4):164-174. doi:10.1016/j.tig.2012.01.005
16. Smith A (2017) Formative pluripotency: the executive phase in a developmental continuum. Development 144 (3):365-373. doi:10.1242/dev.142679
17. Hackett JA, Surani MA (2014) Regulatory Principles of Pluripotency: From the Ground State Up. Cell Stem Cell 15 (4):416-430. doi:10.1016/j.stem.2014.09.015
18. Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, Saitou M (2008) Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet 40 (8):1016-1022. doi:ng.186 [pii]
10.1038/ng.186
19. Tu S, Narendra V, Yamaji M, Vidal SE, Rojas LA, Wang X, Kim SY, Garcia BA, Tuschl T, Stadtfeld M, Reinberg D (2016) Co-repressor CBFA2T2 regulates pluripotency and germline development. Nature 534 (7607):387-390. doi:10.1038/nature18004
20. Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA (2005) Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436 (7048):207-213. doi:10.1038/nature03813

21. Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M (2008) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes & Development 22 (12):1617-1635
22. 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 (4):519-532. doi:S0092-8674(11)00771-9 [pii]
10.1016/j.cell.2011.06.052
23. Saitou M, Miyauchi H (2016) Gametogenesis from Pluripotent Stem Cells. Cell Stem Cell 18 (6):721-735. doi:10.1016/j.stem.2016.05.001
24. Irie N, Weinberger L, Tang WW, Kobayashi T, Viukov S, Manor YS, Dietmann S, Hanna JH, Surani MA (2015) SOX17 is a critical specifier of human primordial germ cell fate. Cell 160 (1-2):253-268. doi:10.1016/j.cell.2014.12.013
25. Sasaki K, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, Ohta H, Moritoki Y, Iwatani C, Tsuchiya H, Nakamura S, Sekiguchi K, Sakuma T, Yamamoto T, Mori T, Woltjen K, Nakagawa M, Yamamoto T, Takahashi K, Yamanaka S, Saitou M (2015) Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. Cell Stem Cell 17 (2):178-194. doi:10.1016/j.stem.2015.06.014
26. Hikabe O, Hamazaki N, Nagamatsu G, Obata Y, Hirao Y, Hamada N, Shimamoto S, Imamura T, Nakashima K, Saitou M, Hayashi K (2016) Reconstitution in vitro of the entire cycle of the mouse female germ line. Nature 539 (7628):299-303. doi:10.1038/nature20104
27. Kurimoto K, Yabuta Y, Hayashi K, Ohta H, Kiyonari H, Mitani T, Moritoki Y, Kohri K, Kimura H, Yamamoto T, Katou Y, Shirahige K, Saitou M (2015) Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. Cell Stem Cell 16 (5):517-532. doi:10.1016/j.stem.2015.03.002
28. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 15 (12):554. doi:10.1186/s13059-014-0554-4
29. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 17 (1):10-12
30. Denning W, Das S, Guo S, Xu J, Kappes JC, Hel Z (2013) Optimization of the transductional efficiency of lentiviral vectors: effect of sera and polycations. Mol Biotechnol 53 (3):308-314. doi:10.1007/s12033-012-9528-5

nature COMMUNICATIONS

Corrected: Publisher correction

ARTICLE

DOI: 10.1038/s41467-018-06230-0 OPEN

Tracing the transitions from pluripotency to germ cell fate with CRISPR screening

Jamie A. Hackett[1,2,3], Yun Huang[1,3], Ufuk Günesdogan[1,3,4], Kristjan A. Gretarsson[2], Toshihiro Kobayashi[1,3,5] & M. Azim Surani[1,3]

Early mammalian development entails transit through naive pluripotency towards post-implantation epiblast, which subsequently gives rise to primordial germ cells (PGC), the founding germline population. To investigate these cell fate transitions, we developed a compound-reporter to track cellular identity in a model of PGC specification (PGC-like cells; PGCLC), and coupled it with genome-wide CRISPR screening. We identify key genes both for exit from pluripotency and for acquisition of PGC fate, and characterise a central role for the transcription regulators *Nr5a2* and *Zfp296* in germline ontogeny. Abrogation of these genes results in widespread activation (*Nr5a2*$^{-/-}$) or inhibition (*Zfp296*$^{-/-}$) of WNT pathway factors in PGCLC. This leads to aberrant upregulation of the somatic programme or failure to activate germline genes, respectively, and consequently loss of germ cell identity. Our study places *Zfp296* and *Nr5a2* as key components of an expanded PGC gene regulatory network, and outlines a transferable strategy for identifying critical regulators of complex cell fate decisions.

[1] Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. [2] Epigenetics and Neurobiology Unit, European Molecular Biology Laboratory (EMBL), via Ramarini 32, 00015 Rome, Italy. [3] Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. [4] Department of Developmental Biology, University of Göttingen, Göttingen Center for Molecular Biosciences, Justus-von-Liebig Weg 11, 37077 Göttingen, Germany. [5] Present address: Center for Genetic Analysis of Behaviour, National Institute for Physiological Sciences, 5-1 Higashiyama Myodaiji, Okazaki, Aichi 444-8787, Japan. These authors contributed equally: Jamie A. Hackett, Yun Huang and Ufuk Günesdogan. Correspondence and requests for materials should be addressed to J.A.H. (email: jamie.hackett@embl.it) or to M.A.S. (email: a.surani@gurdon.cam.ac.uk)

The germ cell lineage generates the totipotent state and transmits heritable genetic and epigenetic information to the next generation. Robust specification of primordial germ cells (PGC), the precursors of sperm and eggs, is therefore a critical developmental event to ensure the propagation of a species. Approximately 30 specified PGCs are detected in mouse embryos at embryonic day (E) 7.25, which arise from 'competent' epiblast precursor cells in response to BMP and WNT signalling[1,2]. The specification of PGCs is accompanied by induction of key germ-cell genes, repression of the nascent somatic-mesodermal programme, and widespread epigenetic remodelling, including global DNA demethylation[3,4].

PGC specification follows WNT-dependent induction of the primitive streak/mesodermal gene *Brachyury (T)*, which in mice promotes activation of key PGC specifiers; *Blimp1*, *Prdm14*, *Cbfa2t2* and *Ap2γ*[5–8]. These transcription factors form a self-reinforcing network that feeds back to suppress other WNT/BMP-induced mesodermal genes (including *T*), thereby repressing the ongoing somatic programme in early PGCs[9]. *Blimp1* and *Prdm14* additionally activate germline-specific genes and initiate epigenome resetting, with mutation of either gene resulting in loss of PGCs by E12.5 [10,11]. The broader gene regulatory network that controls PGC ontogeny has however been relatively uncharted, due to the absence of unbiased functional approaches and the challenges of analysing the limited number of nascent PGCs in embryos.

The recent development of an in vitro model of PGCs, termed PGC-like cells (PGCLC), now facilitates molecular studies of the specification and developmental events of the germ cell lineage[12,13]. PGCLC are derived by inducing naive embryonic stem cells (ESC) that are equivalent to the inner cell mass (ICM) of blastocysts (E3.5–E4.5), towards competent epiblast-like cells (EpiLC), which closely resemble pre-gastrulation mouse epiblast (E5.5–E6.5)[14,15]. EpiLC can in turn be induced to undergo specification as PGCLC in response to BMP and WNT. Specified PGCLC are equivalent to migratory PGCs in vivo (E8.5–E10.5)[12], and have the potential to develop to mature functional gametes[13,16]. This model is therefore appropriate to investigate the inherent regulatory mechanisms of nascent germ cells, and the preceding developmental transitions from ICM, through post-implantation development and PGC specification.

The advent of genome-wide CRISPR screening has enabled unbiased interrogation of recessive gene function in a wide spectrum of biological contexts[17–19]. We reasoned that by designing appropriate reporters, CRISPR screening could be adapted to identify genes involved in controlling sequential cell fate decisions during lineage-specific differentiation. Specifically, by employing the PGCLC model we embarked on identification of genes that are important for (i) maintenance of naive pluripotency, (ii) transition to the germline competent state (EpiLC) and, (iii) specification of the PGC lineage. We further investigate novel candidates to reveal their key role and mechanistic function during nascent germ cell development. From a broader perspective, we demonstrate that unbiased CRISPR screening can be adapted to probe the genetic basis of complex multi-step developmental processes, using the germline lineage as a paradigm.

Results

Compound reporter for developmental transitions to PGC fate. We set out to design a reporter system that can distinguish between successive cell identities during the developmental transitions from naive pluripotency to specified PGC fate. Single-cell RNA-seq data revealed that *Stella* (also known as *Dppa3*) is expressed in the naive pluripotent ICM but is rapidly downregulated in post-implanation epiblast, with re-activation occurring specifically in nascent PGCs (from E7.5)[20]. In contrast *Esg1* (also known as *Dppa5*) is also expressed in the ICM, but maintains high expression during post-implanation development, and subsequently becomes strongly downregulated in early PGCs; with low germline expression reacquired later (>E9.5) (Fig. 1a). To exploit the mutually exclusive expression of these genes, we generated an ESC line with compound *Stella*-GFP and *Esg1*-tdTomato (SGET) reporters. In this 'traffic-light' system naive pluripotent cells are double-positive (yellow), post-implantation epiblast cells are *Esg1*-positive (red) and nascent PGCs are *Stella*-positive (green). We monitored SGET expression during mouse development by tetraploid complementation and chimera formation, and observed strong double-activation in naive pluripotent epiblast at E4.5 (Fig. 1b). This resolved to single *Esg1-tdTomato* activity in E6.0 epiblast, with both reporters subsequently silenced in somatic tissues by E7.0. Importantly *Stella*-GFP is specifically reactivated in PGCs by E8.5, until E12.5, with *Esg1*-tdTomato additionally showing weak expression in later PGC stages (Fig. 1b). Thus, SGET faithfully recapitulates expression of the endogenous genes during the transitions towards PGC fate.

Next, we examined SGET activity during in vitro specification of PGCLC, which initiates from naive ESC and transitions through 'competent' EpiLC[12]. SGET ESC were largely double-positive (SG^+ET^+), but resolved to *Esg1*-only expression (SG^-ET^+) in >97% of cells upon induction into EpiLC. In contrast, nascent PGCLC (day 2) reactivated *Stella* while concomitantly repressing *Esg1* (SG^+ET^-), with later stage PGCLC (day 6) exhibiting low *Esg1* expression (SG^+ET^{low}) (Fig. 1c). PGCLC carrying SGET therefore recapitulate the in vivo dynamics. To functionally validate the SGET reporter system, we generated ESC with a mutation in the key PGC-specifier *Blimp1* (Supplementary Fig 1A), which resulted in a significant reduction (up to 2.8-fold) in the efficiency of PGCLC generation (Fig. 1d; Supplementary Fig 1B). The remaining $Blimp1^{-/-}$ PGCLC exhibited aberrant gene expression (Supplementary Fig 1C), consistent with in vivo studies where a fraction of nascent 'PGCs' remain but with an altered transcriptome[10]. The SGET reporter system is thus ideal for monitoring successive changes in cell identity, from naive ESC to specified PGCLC.

CRISPR screen identifies key genes for naive ESC. We introduced a single-copy of *Cas9* into SGET ESC (Supplementary Fig 1D), and subsequently infected this line with an integrating lentiviral library of exon-targeting guide RNAs (gRNA)[21], in independent biological replicates. Statistical analysis of gRNA frequency in the ESC population, relative to the initial frequency, reveals essential genes because their cognate gRNAs become depleted concomitant with cell loss. We reasoned that by flow sorting SGET cells that successfully acquire each sequential fate during PGCLC specification, we could identify essential genes for each transition by comparing gRNA frequency to the preceding cell population (Supplementary Fig 1E). Accordingly, this approach filters out genes with a role in prior fate transitions and reveals the critical regulators for each stage of multi-step differentiation events, based on functional requirement.

Following introduction of the CRISPR library, we identified 627 genes that lead to a loss of naive ESC maintained in 2i/LIF when knocked-out ($p < 0.01$ in all independent replicates using the relative ranking algorithm (RRA) in the MAGeCK tool[22]) (Fig. 2a). Gene ontology (GO) terms for these genes were highly enriched for fundamental biological processes such as ribosome biogenesis (Bonferroni adjusted $p = 2.88 \times 10^{-27}$; e.g. *Rps5*) and protein translation (Bonferroni adjusted $p = 4.25 \times 10^{-30}$; e.g. *Eif6*), supporting the efficacy of the library (Supplementary Fig 2A). Notably ESC carrying mutations for the core

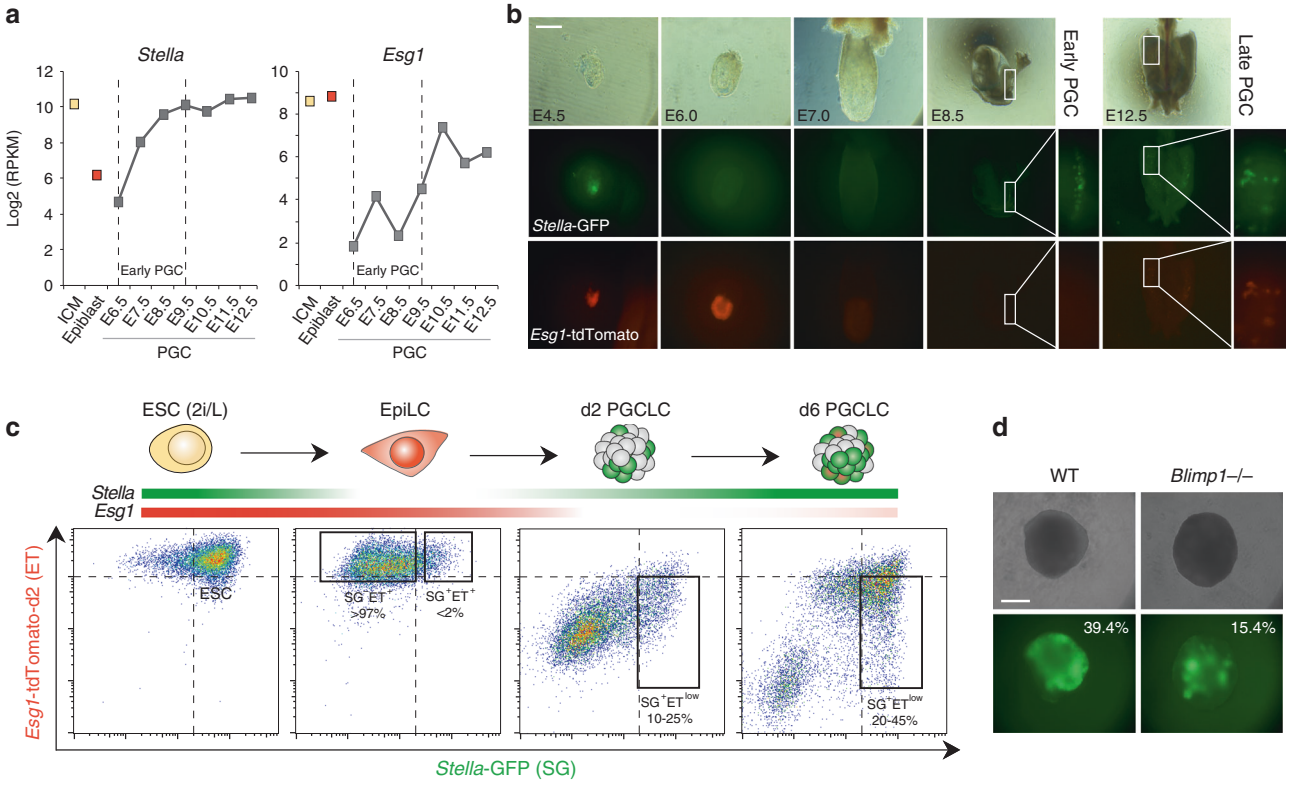


Fig. 1 The SGET reporter for tracking cell fate transitions toward PGCs. **a** Single-cell RNA-seq analysis showing reciprocal gene expression patterns of endogenous *Stella* and *Esg1* during early development (ICM; E3.5 or Epiblast; E6.5) and in primordial germ cells (PGC). **b** Chimera and tetraploid complementation assays confirm faithful expression of the *Stella*-GFP (green) and *Esg1*-tdTomato (red) SGET reporter during early development and in PGCs. **c** Representative FACS analysis of SGET activation during in vitro cell fate transitions of ESC into EpiLC and subsequently PGCLC. **d** Representative images of impaired PGCLC induction from *Blimp1*$^{-/-}$ SGET ESC. Numbers indicate percentage of SG^{+}ET^{low} PGCLC at day 5 as determined by FACS. Scale bar(s): 200 μm

pluripotency factors *Oct4* and *Sox2* were also highly depleted, consistent with their essential role in propagating pluripotent ESC[23] (Fig. 2a, b; Supplementary 2B). The majority of naive pluripotency genes were not depleted however, which is in line with the capacity of 2i/LIF media to buffer against a single genetic perturbation to the naive pluripotency network (Fig. 2b)[24]. Importantly, this enables the role of such naive pluripotency genes to be examined in PGC specification, where they are also typically expressed, since knockout cells remain in the population. Exceptions however are *Nanog* and *Tfcp2l1*, which are likely depleted owing to a proliferation disadvantage upon knockout in ESC. In contrast to the many depleted genes, we observed that only 21 genes become enriched in the population upon knockout, primarily corresponding to tumour suppressors (Supplementary Fig 2B & C). Among these *Trp53* (*p53*) is the top hit, suggesting that *p53*-mutant mouse ESC acquire a major selective advantage, similar to recent reports in human ESC[25].

We next compared the genes essential for naive ESC (in 2i/LIF) with those essential for ESC in conventional serum/LIF culture, which represents an alternative pluripotent state[23]. Using datasets from the same CRISPR library[21], we observed good overlap (62%) between ESC conditions, but noted 38% of genes appear to be critical under only specific pluripotent culture parameters (Fig. 2c). Among the 2i/LIF specific genes were *Ogdh* and *Dlst;* two enzymes that mediate α-ketoglutarate metabolism which feeds into the TCA cycle. We used siRNAs to test acute loss of *Ogdh* and observed a significant reduction of proliferation and/or increased cell death in 2i/LIF ESC but not serum/LIF ESC (Fig. 2d). We also tested *Txn1*, which was scored as essential in both pluripotent ESC states, and found an equivalent cell reduction in 2i/LIF and serum/LIF (Fig. 2d). These data suggest ESC in distinct pluripotent states rely, in part, on distinct genetic networks, with ESC in 2i/L uniquely reliant on *Ogdh* metabolism of α-ketoglutarate for example. More generally, establishing this mutant SGET ESC population enables screening for functional regulators of EpiLC, and subsequently PGCLC, without confounding general survival genes.

Acquisition of post-implantation epiblast fate. We next used fluorescence-activated cell sorting (FACS) to isolate EpiLC that successfully transited to the SG^{-}ET^{+} epiblast state, as well as cells that maintained *Stella* expression (SG^{+}ET^{+}) (<2%), indicative of failure to exit naive pluripotency. Analysis of this SG^{+}ET^{+} EpiLC population revealed enrichment for 21 candidate genes with potential intrinsic roles in dissolution of naive pluripotency ($p <$ 0.01 using the RRA[22]), including the two prototypical regulators *Tcf3* and *Zfp281* (Fig. 2e; Supplementary Fig 3A)[26,27]. The candidates additionally included the epigenetic regulator *Dnmt1* and also *Rest*, which are known to be important during early differentiation[28,29].

To test the role of novel candidates in driving exit from pluripotency we generated ESC knockouts for *Zmym2* (also known as *Zfp198)*, *FoxP1*, *Uchl5* and *Zfp281* as a positive control, using CRISPR targeting. Consistent with their failure to repress *Stella*-GFP during EpiLC formation, mutant EpiLC were also impaired in activation of key epiblast markers *Fgf5*, *Dnmt3b* and *Otx2*, while their proliferation appeared unaffected (Fig. 2f; Supplementary Fig 3B-C). We observed both a delay in activation of epiblast genes and a reduction in their absolute levels in

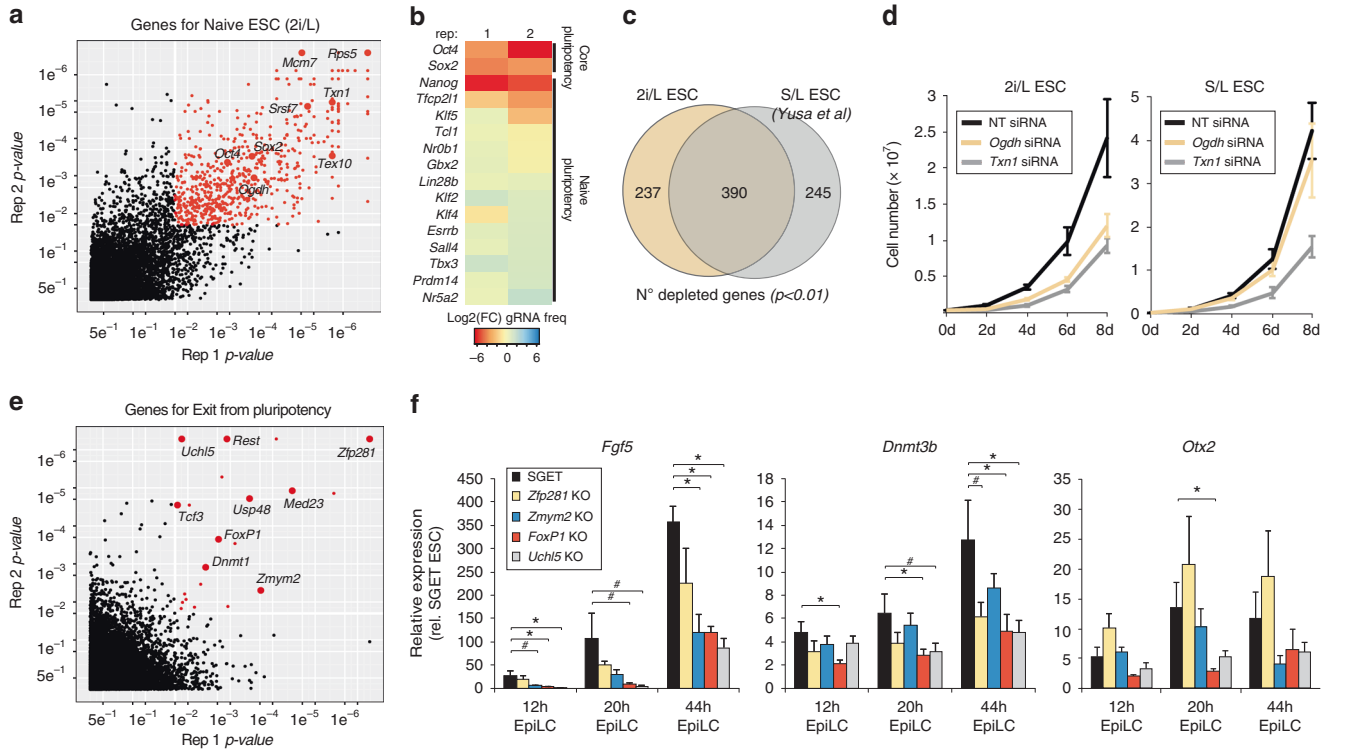


Fig. 2 Identification of important genes for ESC and EpiLC induction. **a** Scatter plot showing essential genes for ESC propagation in 2i/LIF (red datapoints) as determined by a CRISPR screen. **b** Heatmap showing mean fold change (FC) in normalised frequency of gRNAs targeting core- and naive- pluripotency genes in ESC relative to frequency in starting gRNA library. Reduced frequency indicates a functionally important gene for ESC in 2i/LIF. **c** Venn diagram intersecting significant depleted genes in 2i/LIF ESC (2i/L) with serum/LIF (S/L) maintained ESC. **d** Proliferation of ESC following siRNA knockdown of *Ogdh* or *Txn1* in ESC maintained in 2i/L or S/L culture conditions. **e** Scatter plot showing significantly enriched gene knockouts in $SG^{+}ET^{+}$ EpiLC that have failed to exit naive pluripotency. **f** Gene expression showing impaired activation of key epiblast markers in EpiLC carrying knockouts for candidate exit from pluripotency regulators. Significance was determined using a one-tailed *t*-test; $^{*}p < 0.05$; $^{\#}p < 0.1$. Error bars show s.e.m. of triplicate independent experiments

knockout lines, despite high levels of powerful differentiation-inducing FGF and ACTIVIN in the culture medium. Notably, the *Zmym2*-, *FoxP1*-, *Uchl5*- mutant lines exhibited expression defects comparable with the bona fide exit from pluripotency regulator *Zfp281*, supporting an important role for them in the timely acquisition of post-implantation cell identity. Consistently all three candidates are transcriptionally upregulated during EpiLC induction (Supplementary Fig 3D).

Candidate genes for specification of PGC fate. We next focused on genes involved in mouse germline specification by inducing PGCLC from competent $SG^{-}ET^{+}$ EpiLC. In order to obtain sufficient numbers of PGCLC for coverage of the gRNA library, we optimised and scaled-up the induction (Supplementary Fig 4A). Specified $SG^{+}ET^{low}$ PGCLC were FACS purified at day (d) 6 of differentiation to allow time for depletion of cells carrying mutations in critical germline genes. Because PGCLC numbers remained a limiting factor however, we relaxed the candidate threshold to $p < 0.05$, while retaining a $p < 0.01$ in at least one replicate, to account for increased noise (Supplementary Fig 4B). This resulted in identification of 23 candidate genes involved in specification and/or development of nascent PGC. The frequency of gRNAs targeting these genes was highly depleted in PGCLC relative to preceding EpiLC (Fig. 3a), while negative control gene families were not depleted (Supplementary Fig 4C). Moreover, genes with an established critical role in PGC specification, such as *Blimp1*, *Cbfat2t2* and *Prdm14*, also showed marked reduction in gRNA frequency in PGCLC, supporting the efficacy of the screen (Fig. 3a). These genes narrowly failed to meet significance across all replicates however; *Cbfa2t2* scored *p*-values of $p = 0.00006$ and $p = 0.22$ for example. We also noted that several pluripotency genes (*Nr5a2*, *Esrrb* and *Sall4*) were depleted specifically from PGCLC (Supplementary Fig 4C), suggesting that some genes typically linked with pluripotency have a potentially important function in the germline.

To refine our candidate list we examined their expression profile during induction of PGCLC from the transduced SGET ESC by RNA-seq. This confirmed that $SG^{+}ET^{low}$ PGCLC activate high levels of key PGC markers such as *Blimp1*, *Stella*, *Itgb3*, *Nanos3* and *Prdm14*, as well as repressed *Uhrf1* and *Klf4*, supporting their authentic germline identity (Supplementary Fig 4D). The dynamics of candidate gene expression revealed four broad clusters, with cluster 1 and 4 corresponding to genes that are expressed relatively stably throughout ESC to PGCLC transition; the latter is nonetheless only just above the detection threshold implying it may filter out false-positives. In contrast, cluster 2 and 3 reflected dynamically regulated genes that are activated in PGCLC, with cluster 3 corresponding to genes fully silenced in EpiLC before strong germ cell re-activation, and including known PGC specifiers (Fig. 3b). From these, we chose to focus on *Zfp296* and *Nr5a2*, since they rank 3 and 7 overall in the screen, and exhibit comparable PGCLC expression dynamics to *Prdm14* and *Blimp1* (Fig. 3b; Supplementary Fig 4D). *Nr5a2* encodes an orphan nuclear receptor previously associated with reprogramming towards pluripotency, and also has a role in extra embryonic development, which leads to embryonic lethality when

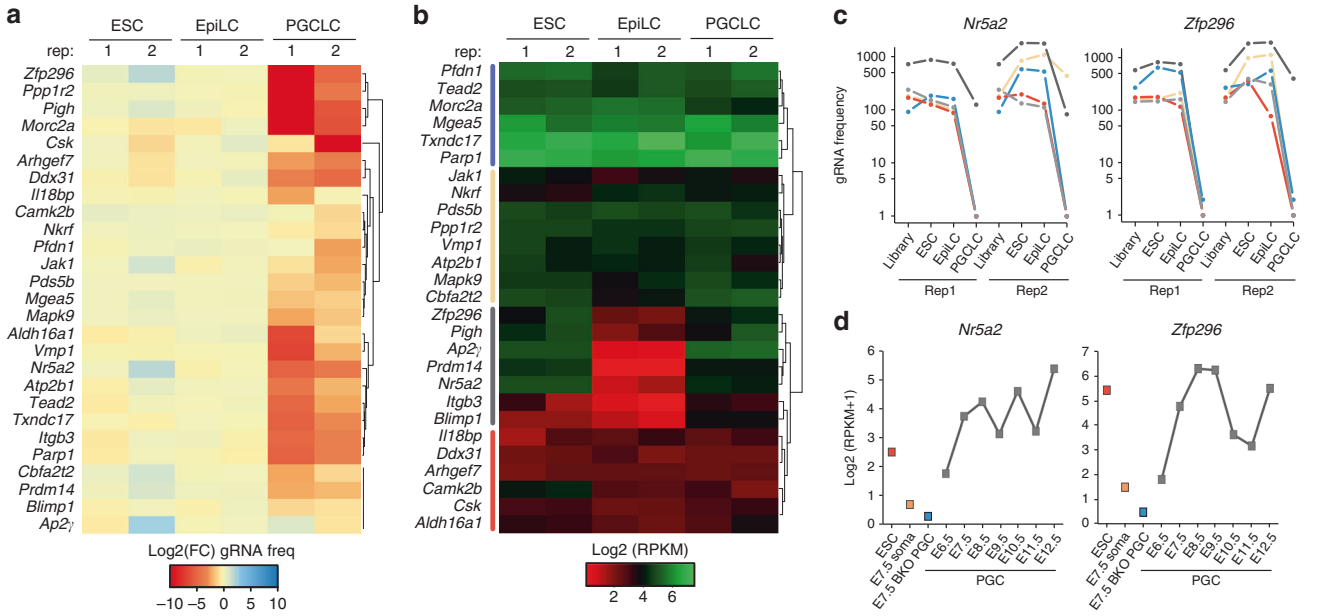


Fig. 3 Candidate genes for primordial germ cell fate. **a** Heatmap showing mean fold-change (FC) in normalised frequency of gRNAs that target candidate PGC regulators, relative to their frequency in the preceding cell state. **b** Log_2 RPKM gene expression dynamics of candidate PGC regulators during induction of SGET ESC into d6 PGCLC in the screen. **c** Normalised frequency of individual gRNAs targeting *Nr5a2* and *Zfp296* during induction of PGCLC. **d** Expression dynamics of *Nr5a2* and *Zfp296* during in vivo formation of PGCs by single-cell RNA-seq. *Blimp1*$^{-/-}$ PGC (BKO) and soma are shown for reference

abrogated[30]. *Zfp296* encodes a zinc-finger protein that was among the original 20 Yamanaka factors, and has recently been shown to influence heterochromatin and induced pluripotent stem cell (iPSC) generation[31–33].

Inspection of the individual gRNAs that target *Nr5a2* and *Zfp296* revealed a strong and reproducible reduction of their frequency specifically in PGCLC, but not in the preceding developmental transitions (Fig. 3c). Moreover, single-cell RNA-seq indicates both genes are highly upregulated in nascent PGC in vivo at the very onset of specification, but not in adjacent somatic cells (Fig. 3d). Both *Nr5a2* and *Zfp296* are also strongly upregulated in early human PGCs (Supplementary Fig 4E), and fail to be activated in *Blimp1*$^{-/-}$ PGC that are destined to diverge from the germline lineage (Fig. 3d). The striking upregulation of these genes during incipient PGC specification, coupled with their high rank in the functional screen, prompted us to investigate their role in detail.

Nr5a2 and Zfp296 regulate germline ontogeny. We used CRISPR editing to generate frame-shifted null-alleles of *Nr5a2* or *Zfp296* in SGET ESC. *Nr5a2*$^{-/-}$ ESC proliferated normally, expressed comparable levels of naive pluripotency genes, and exhibited indistinguishable morphology from parental and matched WT ESC derived from the same targeting process. However upon induction of PGCLC, we observed a significant reduction in specification efficiency of *Nr5a2*$^{-/-}$ cells at d2 relative to WT controls (WT 19.8% ± 8.9; KO 6.4% ± 2.3) (Fig. 4a). The impaired induction of *Nr5a2*-mutant cells was further reflected in d6 PGCLC, and across multiple independent knockout lines (WT 44.7% ± 4.0; KO 17.4% ± 4.6), suggesting an important role for *Nr5a2* in PGC specification (Fig. 4a; Supplementary Fig 5A). We next tested *Zfp296*$^{-/-}$ ESC lines, which also exhibited normal morphology and pluripotent gene expression profiles, albeit with modestly reduced proliferation rate. Upon induction, PGCLCs were specified from *Zfp296*-mutant cells at apparently normal efficiency at d2 (WT 14.9% ± 3.2; KO 16.2% ± 2.2), as judged by the SGET reporter. However, by d6 we observed a highly significant depletion of PGCLC in multiple mutant lines (WT 43.1% ± 3.5; KO 8.9% ± 2.8), suggesting the effects of *Zfp296* abrogation on PGCLC are progressive and manifest after specification (Fig. 4b). More generally, these data indicate that loss-of-function of either *Nr5a2* or *Zfp296* leads to a significant impairment of germ cell development.

Despite the dramatic reduction in numbers, both mutant lines give rise to some putative PGCLC. To examine the identity of these cells, and their developmental trajectory, we performed RNA-seq. Unbiased hierarchical clustering revealed that *Nr5a2*$^{-/-}$ EpiLC were interspersed with WT, but that PGCLC were clustered according to genotype, implying transcriptional differences in the absence of *Nr5a2* primarily arise upon induction of germline fate (Supplementary Fig 5). To investigate this further we used principal component analysis (PCA), which confirmed that mutant EpiLC cluster closely with WT but that *Nr5a2*-null PGCLC progressively diverge from WT controls during their development (Fig. 4c). By d6 a strong distinction is apparent. This is consistent with *Nr5a2*$^{-/-}$ EpiLC being competent for the germline lineage but undergoing an abnormal developmental trajectory upon PGCLC specification, which coincides with the timing of *Nr5a2* upregulation in WT cells.

Global transcriptome clustering of *Zfp296*-null cells indicated that precursor EpiLC were distinct yet highly comparable between WT and knockout lines (Supplementary Fig 5). Indeed, PCA showed *Zfp296*$^{-/-}$ EpiLC cluster closely with WT. In contrast, d2 mutant PGCLC exhibited a significantly different state from WT counterparts, that is closer to precursor EpiLC and thus apparently impaired in its developmental path towards germline fate, which is consistent with the subsequent loss of PGCLC numbers (Fig. 4d). The remaining *Zfp296*-mutant PGCLC at d6 appeared to partially re-establish the expected global profile, albeit with greater transcriptome variance between lines than WT. This implies that despite widespread

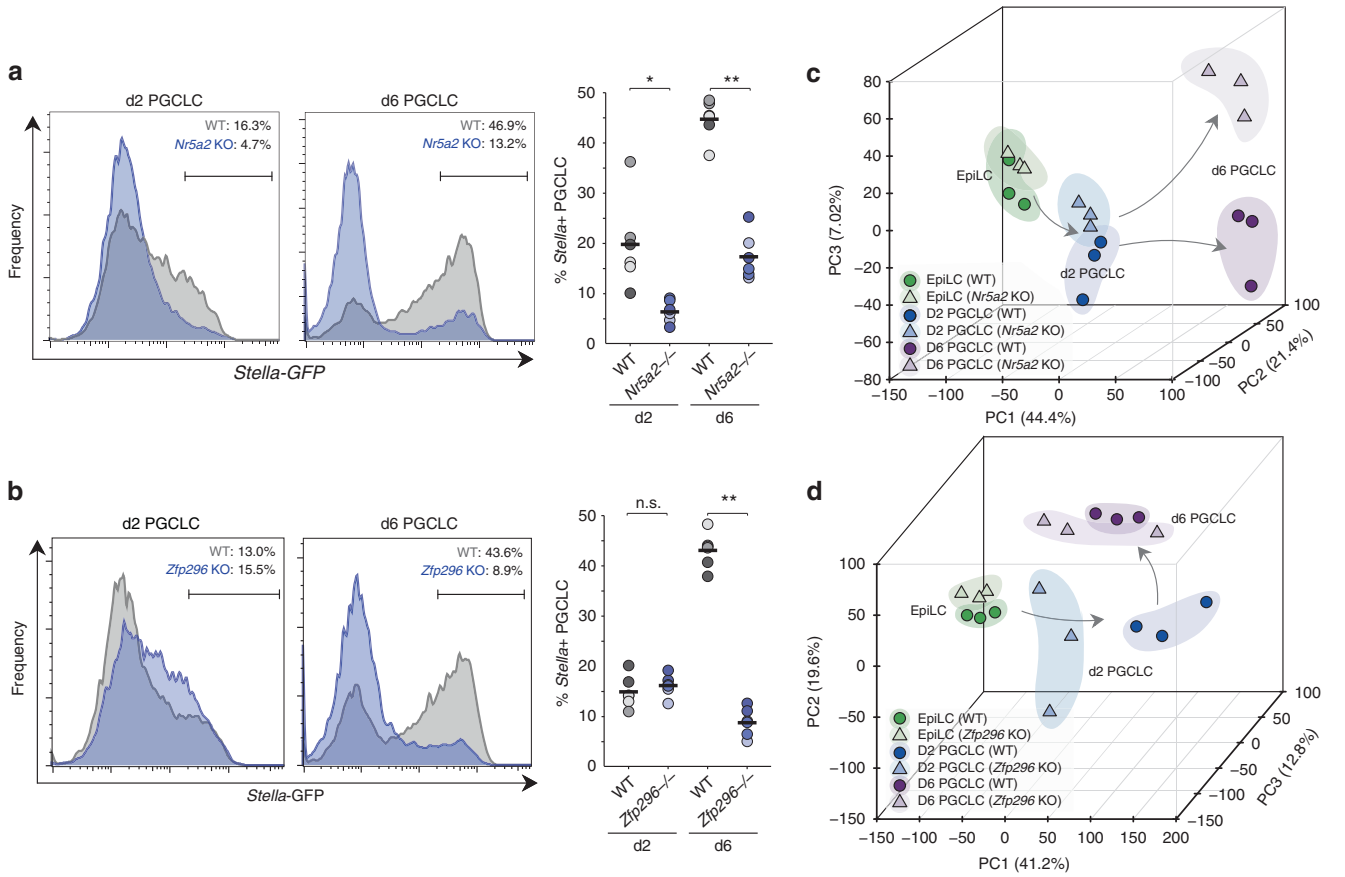


Fig. 4 *Nr5a2* and *Zfp296* are key regulators of germ cell development. **a** Representative FACS plot showing impaired induction of PGCLC in *Nr5a2*-knockout (KO) cells. Shown right is replicate quantifications of independent WT and KO lines. Independent lines are colour-matched. Bar indicates mean value. **b** Representative FACS of PGCLC in *Zfp296*-knockout SGETs, with replicate quantification shown right. **c** Principal component analysis (PCA) showing the developmental trajectory of independent $Nr5a2^{-/-}$ and matched-WT SGET lines during induction of PGCLC, based on global transcriptome. **d** PCA of independent $Zfp296^{-/-}$ and matched-WT SGET transcriptomes. $^{*}p < 0.05$; $^{**}p < 0.01$

transcriptional mis-regulation, a minor proportion of mutant PGCLC can overcome the effects of *Zfp296* abrogation, and give rise to putative germ cells (Fig. 4d). Overall, these data suggest that loss of *Nr5a2* or *Zfp296* has a dramatic influence on PGC development.

Nr5a2 and Zfp296 converge on reciprocal regulation of WNT. To investigate the mechanisms that impair development of mutant PGCLC and underpin their aberrant transcriptomes in more depth, we identified differentially expressed genes (adjusted $p < 0.05$; >2 FC) in EpiLC and PGCLC. This revealed that downregulation of naive pluripotency genes and upregulation of post-implantation markers proceeded similarly between $Nr5a2^{-/-}$ and WT EpiLC counterparts (Fig. 5a). Moreover, there was appropriate upregulation of early PGC markers such as *Nanos3*, *Stella*, *Blimp1* and *Prdm14* in d2 $Nr5a2^{-/-}$ PGCLC, that was at least equivalent to WT, albeit with a modest failure to upregulate *Nanog*. However, we noted d2 *Nr5a2*-null PGCLC exhibited a highly significant overexpression of WNT pathway genes and targets, including *Cdx2*, *T* and *Mixl1* (Fig. 5a). This was further reflected in gene ontology (GO) analysis of differentially expressed genes at d2, which highlighted canonical WNT signalling pathway (Bonferroni adjusted $p = 0.017$) and pattern specification process (Bonferroni adjusted $p = 0.014$) as top hits (Table S1). We confirmed hyperactivation of key WNT pathway genes *T* and *Wnt3* in early $Nr5a2^{-/-}$ PGCLC using qRT-PCR on independent replicates (Fig. 5b). Strong overexpression of the master mesoderm regulator *T* is predicted to override its role in germ cell induction and divert cells toward a somatic mesendoderm programme. Consistently, the surviving $Nr5a2^{-/-}$ PGCLC at d6 exhibit acute activation of mesendoderm regulators such as *Tbx4*, *Hey1* and *Pou2f3* (Fig. 5a). This collectively suggests that the absence of *Nr5a2* leads to aberrant transduction of WNT signalling in nascent PGCLC, which in turn leads to upregulation of mesendoderm master-regulators and consequently de-repression of a somatic programme. It is possible that the aberrant activation of somatic genes in the absence of *Nr5a2* is a contributing factor that drives the majority of prospective PGCLC out of the germline programme, leading to reduced efficiency of specification.

We next analysed gene expression in *Zfp296*-null EpiLC, which revealed that pluripotency genes including *Nanog*, *Esrrb* and *Tfcp2l1* were repressed as expected, while EpiLC markers such as *Fgf5* and *Dnmt3b* were appropriately upregulated. In contrast mutant PGCLC at d2 exhibited a clear signature of mis-expressed genes (Fig. 5a). Remarkably, GO analysis again suggested this primarily corresponded to WNT targets, but in contrast to hyperactivation of WNT in $Nr5a2^{-/-}$ PGCLC, *Zfp296*-null PGCLC exhibited a severe block of WNT-associated gene activation; the top enrichment is negative regulation of canonical WNT signalling (Bonferroni adjusted $p = 0.0064$). Among all genes *T* is the most significantly downregulated in early *Zfp296*-mutant PGCLC, with *Cdx2*, *Notum* and *Dkk1* also being among

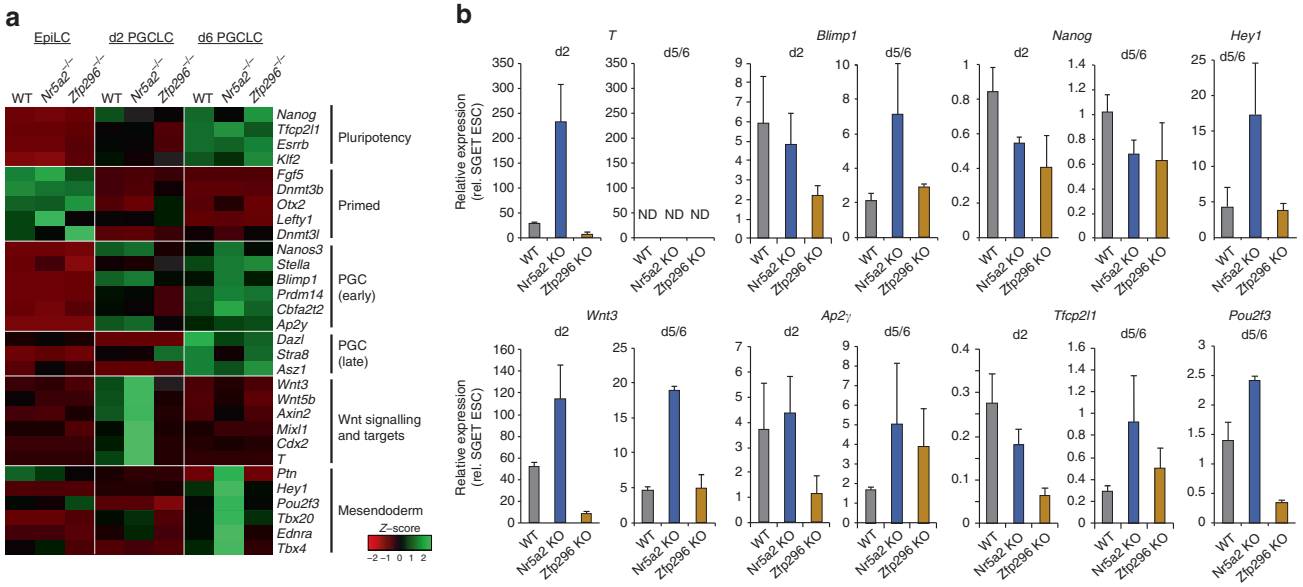


Fig. 5 Compound mis-regulation of WNT and germ cell genes in the absence of *Nr5a2* and *Zfp296*. **a** Heatmap showing change in mean expression profile of selected genes from triplicate independent WT, *Nr5a2*$^{-/-}$ or *Zfp296*$^{-/-}$ lines during PGCLC induction by RNA-seq. **b** Quantitative validation of gene expression changes in PGCLC from WT, *Nr5a2*$^{-/-}$ and *Zfp296*$^{-/-}$ lines by independent qRT-PCR experiments. Error bars show s.e.m. of duplicate biological experiments, each comprising at least twelve independent replicate inductions. ND not detected

the top 20. The absence of appropriate level WNT signalling could manifest as impaired activation of critical germline genes that are WNT and/or *T* targets[5]. Indeed, *Zfp296*-null PGCLC at d2 exhibited a block in *Blimp1*, *Prdm14 Cbfa2t2a* and *Ap2γ* activation (Fig. 5a). qRT-PCR analysis independently confirmed highly impaired *T* and *WNT* genes in *Zfp296*-mutant PGCLC and concomitant significant reduced activation of germline regulators *Blimp1* and *Ap2γ* (Fig. 5b). This suggests that absence of *Zfp296* establishes a reciprocal developmental failure relative to deletion of *Nr5a2*, with both genes converging on regulation of appropriate WNT signalling in nascent PGCLC. In the case of *Zfp296*, this appears to manifest in a delayed but dramatic loss of PGCLC by d6. Notably, the SG^{-}ET^{-} cells from both *Nr5a2* and *Zfp296* embryoids, which have putatively acquired somatic fate, exhibited comparable expression of lineage-specific markers *Sox17*, *Hoxa1* and *Sox7* relative to wild type (Supplementary Fig 6A). This implies that the loss of function for both genes does not impede the exit from the pluripotent state towards somatic fates, but these cells are impaired specifically in their capacity to form PGCs.

Rescue of Nr5a2 and Zfp296 deficiency. We considered that the impaired specification and aberrant transcriptome of mutant PGCs could be due to impaired epigenetic reprogramming, as reported in *Prdm14*$^{-/-}$ PGCs[11]. To investigate this, we quantified global DNA methylation erasure and observed that WT, *Nr5a2*$^{-/-}$ and *Zfp296*$^{-/-}$ PGCLC all underwent extensive and comparable DNA demethylation (Fig. 6a). This was confirmed by bisulfite pyrosequencing analysis of LINE-1 elements (Supplementary Fig 6B). We therefore turned our attention back to understanding whether the observed mis-regulation of WNT could be driving impaired germline fate by exposing incipient PGCLC to WNT inhibitor (WNTin: XAV939) or WNT agonist (WNTag: Chiron). Addition of WNTin led to strong down-regulation of key PGC genes including *T* and *Wnt3*, as well as *Nanog*, consistent with the effects of impaired WNT transduction in *Zfp296*-null PGCLC (Fig. 6b). Reciprocally, WNTag affected an increase in expression of *T* and *Wnt3* at d2, while also down-regulating *Nanog*, which is in line with the effects of WNT over-activity in *Nr5a2*-null PGCLC. Indeed, WNTag also elicited overexpression of mesendoderm gene *Hey1* by d6 PGCLC (Fig. 6c). Notably WNTinh led to a significant depletion of specified PGCLC ($p < 0.05$), while partial WNT activation by WNTag also trended towards impaired PGCLC specification (Fig. 6d). These data suggest that precise levels of WNT activity are necessary for optimal PGC specification.

Addition of WNTin to *Nr5a2*$^{-/-}$ PGCLC rescued hyperactivated *T* and *Wnt3*, demonstrating that their upregulation in *Nr5a2*-mutant PGCLC is likely a direct consequence of de-regulated WNT activity (Fig. 6b). In contrast addition of excess WNTag to *Zfp296*-mutant PGCLC only resulted in modest rescue of aberrantly repressed WNT targets, suggesting that in the absence of *Zfp296*, germ cells cannot respond to WNT activation, even when forced. We therefore attempted to rescue the specification defects by generating mutant PGCLC that carry Doxycycline (DOX)-inducible expression of wild-type *Nr5a2* and *Zfp296*. We induced *Nr5a2* expression specifically during PGCLC induction in a *Nr5a2*$^{-/-}$ background, which led to a highly significant rescue of PGCLC specification. Moreover, this reimposed suppression of somatic and WNT target genes, implying a direct effect of *Nr5a2* on these pathways (Fig. 6e; Supplementary Fig 6C-D). In contrast, we found expression of *Zfp296* to be highly toxic to both ESC and PGCLC, even at very low background (leaky) levels, suggesting precise control over its expression level is important.

Remarkably DOX-induced *Nr5a2* expression in mutant PGCLC resulted in increased PGCLC induction as compared to wild-type counterparts. This encouraged us to test whether *Nr5a2* could drive germline fate independently of BMP cytokines that are normally requisite. Here, forced expression of *Nr5a2* in wild-type EpiLC led to emergence of SG^{+}ET^{low} PGCLC at >tenfold higher efficiency than background; at levels comparable with forced expression of canonical PGC-specifier *Prdm14*, albeit at low absolute frequency (3.3%) (Fig. 6f). Induction of *Nr5a2* resulted in PGCLC with apparently similar expression profiles as authentic PGCs (Supplementary Fig 6E). This indicates that activation of *Nr5a2* can, at low efficiency, directly induce PGC-like fate from competent EpiLC. More

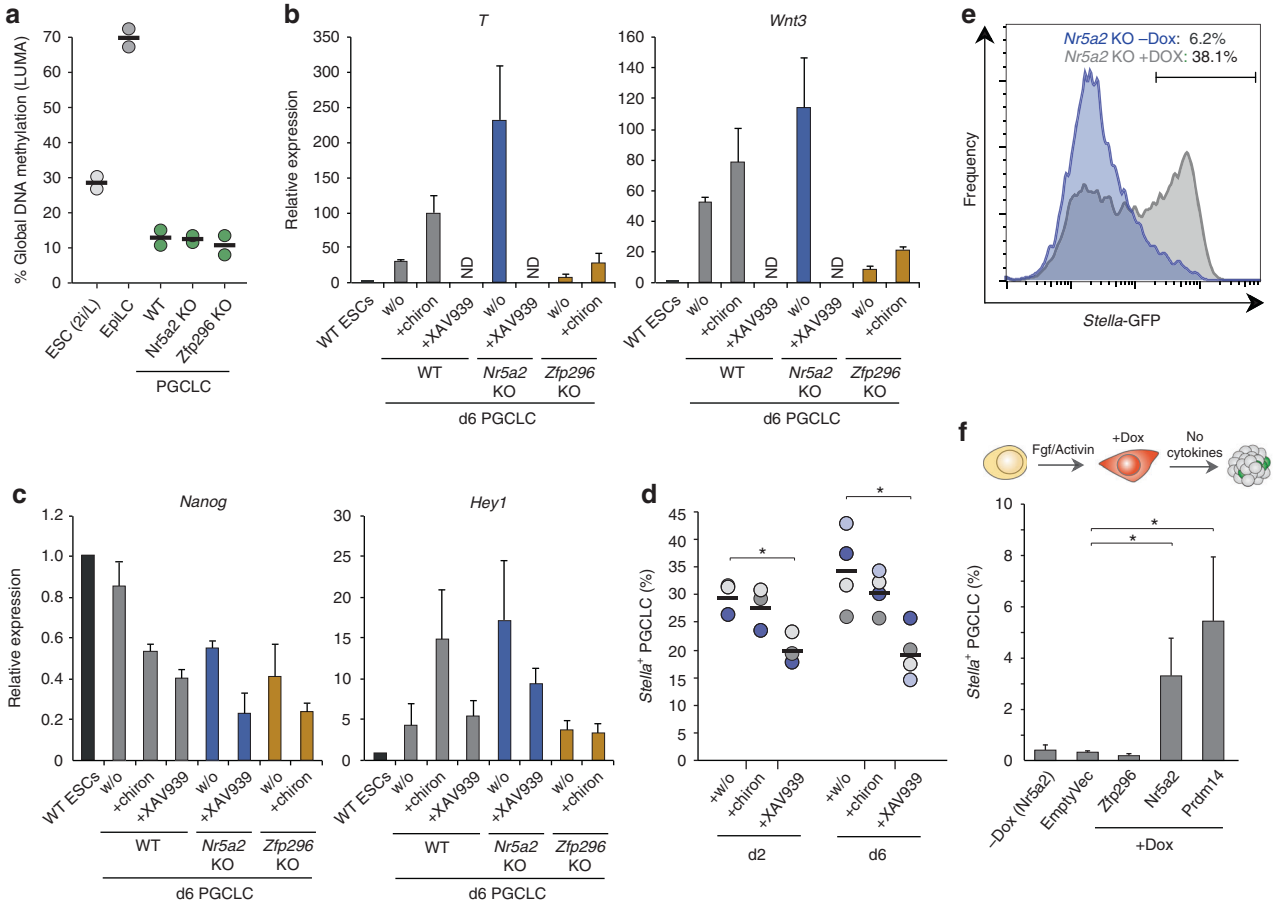


Fig. 6 Role of WNT and rescue of *Nr5a2* defects. **a** Global levels of DNA methylation reprogramming in WT and mutant PGCLC assayed by LUMA. **b** qRT-PCR analysis showing expression of WNT genes/targets are affected by WNT inhibitor (+XAV939) or agonist (+chiron), similarly to knockout PGCLC. w/o: control without WNT modulation. **c** Expression of *Nanog* and mesendoderm gene *Hey1* by qRT-PCR in WT and mutant PGCLC with WNT modulation. **d** Effects of WNT manipulation on efficiency of PGCLC ontogeny in quadruplicate independent experiments. Bar indicates mean value and $^*p < 0.05$. **e** FACS plot showing the *Nr5a2*-mutant PGCLC specification defect at d2 is rescued by Dox-inducible activation of exogenous *Nr5a2* cDNA. Dox is added after EpiLC induction. **f** Percentage *Stella*$^+$ PGCLC induced without cytokines from WT cells upon DOX-inducible expression of the indicated cDNA

generally, our data places *Nr5a2* and *Zfp296* as critical regulators that provide robustness to germline specification by modulating appropriate WNT activity.

Discussion

Our study outlines a principle for using CRISPR screens to deconvolve the genetic basis of successive cell fate decisions. By establishing a 'traffic-light' SGET reporter system we traced the pathway from pluripotency to germ cell fate, and identified important genes for each transition. This strategy can be transposed to interrogate multi-step transitions towards any lineage for which a faithful compound reporter or isolation method can be developed.

Utilising SGET we identified genes involved in exiting naive pluripotency, including the previously described key regulators *Zfp281* and *Tcf3*[34,35], and further validated roles for novel candidates *Zmym2*, *FoxP1* and *Uchl5*. Among these *Zmym2* expression is anti-correlated with naive pluripotency genes at the single-ESC level[36], and is upregulated when cells are induced to egress from naive pluripotency both in vitro and in vivo (Supplementary Fig 3). Indeed, ZMYM2 is known to directly interact with NANOG[37] where its negative regulatory activity could blunt NANOG function, or it could alternatively function in this role as part of the LSD1 complex[38]. Interestingly the transcription regulator *FoxP1* is also strongly activated during EpiLC induction and, while it can be alternatively spliced to generate a pro-pluripotency isoform[39], our data implies canonical *Foxp1* isoforms play a role in antagonising naive pluripotency. The potential mechanism of function for the deubiquitinase *Uchl5* is less clear, but it has previously been associated with regulation of the TGF-β/Smad pathway[40]. Future work will be important to characterise the regulatory role of the full repertoire of candidates in control of pluripotent status.

We further exploited the SGET strategy to identify *Zfp296* and *Nr5a2* as key functional genes for the germline lineage and as important parts of an expanded PGC gene regulatory network. Deletion of either *Nr5a2* or *Zfp296* converges on mis-regulation of the WNT pathway, resulting in its aberrant activation or inhibition, respectively, implying that precisely regulated levels of WNT is fundamental for germ cell ontogeny. Indeed, loss of either gene leads to severely disrupted PGCLC development. Of note *Nr5a2* is a reported WNT target[41] and directly interacts with β-catenin[42], raising the possibility it forms a negative-feedback loop with WNT to modulate appropriate signalling levels in PGCs. This must strike a balance between WNT-mediated activation of germ cell regulators such as *Blimp1* and *Prdm14*[5], and the capacity to subsequently repress the WNT co-activated mesodermal programme. By blunting WNT activity, NR5A2 may ensure the threshold is tipped towards reversibility, since in its

absence PGCLC development is impaired and linked with aberrant activation of mesendodermal genes. Functionally, *Nr5a2* was previously shown to be important for reprogramming somatic cells to iPSC but not for pluripotency per se[43–45]. Similarly, PGC undergo reprogramming but do not acquire pluripotency[4], supporting a role for *Nr5a2* in resetting PGCs away from somatic fates. Notably, NR5A2-binding motifs are among the most significantly enriched within germline-associated promoters[46], consistent with a direct functional role for *Nr5a2* on PGC genes.

In contrast to *Nr5a2*, loss of *Zfp296* results in failure to appropriately upregulate WNT in the germline, as well as affecting some TGF-β targets such as *Lefty1*. Underactive WNT in *Zfp296*-mutant PGCLC manifests as impaired activation of early germline genes, including *Blimp1* and *Prdm14*. This in turn leads to dramatic loss of PGCLC. Indeed, a recent report has shown the absence of ZFP296 leads to sterility, confirming the long-term consequences of its loss-of-function on the germline[32]. The molecular mechanism through which *Zfp296* impacts WNT to regulate PGC ontogeny is an important topic for future research. ZFP296 can, however, influence global H3K9 methylation and has been reported to negatively regulate KLF4 activity[32,47]. The latter counteracts murine *Blimp1* expression[48], implying *Zfp296* could contribute to the PGC programme through multiple regulatory routes. Of note, both *Zfp296* and *Nr5a2* are strongly activated in nascent human PGC (Supplementary Fig 4)[49], comparably to their upregulation in mouse PGC, indicating they may have a broad conserved role in mammalian germline development.

Our approach and experimental design is both objective and based on functional outcomes, thereby delineating an alternative strategy to identify critical components of cell-type specific gene regulatory networks, which in the case of germ cells, has previously been challenging. The approach preferentially identifies intrinsic cell fate regulators however; not weak-acting or extrinsic regulators, for at least two reasons. First, the addition of extrinsic signalling at high levels (FGF and ACTIVIN for EpiLC; BMP among others for PGCLC) overcomes weak- or signalling receptor-specific effects by acting on multiple redundant pathways to drive fate transitions. Second, the pooled screen strategy means that knockout cells for a given extrinsically acting gene (e.g. *Wnt3*) can be rescued by paracrine effects from adjacent cells with wild-type alleles. For these reasons, FGF signalling and WNT components were not identified in the exit from pluripotency and PGCLC screens, respectively. Nevertheless, we identified multiple cell-intrinsic candidates. Indeed, recent optimisations to loss-of-function CRISPR libraries[50,51] will minimise false negatives in future iterations, which can arise due to ineffective gRNAs confounding significance scores, as we observe for *Prdm14* for example. In summary, we have used CRISPR screening to identify key regulators for successive cell fate transitions out of naive pluripotency towards the germ cell lineage. In doing so we characterised *Zfp296* and *Nr5a2* as crucial for driving germ cell development by acting to regulate appropriate WNT levels, thereby controlling the balance between activation of PGC genes and silencing of the somatic programme.

Methods

ESC culture and PGCLC induction. ESCs were maintained in N2B27 medium supplemented with 2i (PD0325901 (1 μM) and CHIR99021 (3 μM; both Stemgent), 1000 U/ml LIF (Cambridge Centre for Stem Cell Research) (2i/L) and 1% Knockout Serum Replacement (KSR) on fibronectin at 37 °C in a 5% CO_2 humidified chamber. For serum/LIF (S/L) culture, cells were maintained on gelatin in GMEM supplemented with 10% FCS, L-Glutamine, non-essential amino acids (NEAA), 0.1 mM β-mercaptoethanol, and 1000 U/ml LIF. Induction of EpiLC and PGCLC was carried out as previously described[12] . Briefly, ESCs cultured in 2i/L were washed thrice with PBS and seeded onto fibronectin (16.7 mg/ml) coated plates with EpiLC medium (N2B27, 1% KSR, bFGF (12 ng/ml), Activin A (20 ng/ml)) for 42 h. Media was changed every day. EpiLC were then gently dissociated and seeded at 3000 cells per embryoid in ultra low-cell binding U-bottom 96-well plates (NUNC) with PGCLC induction medium (GK15: GMEM, 15% KSR, NEAA, 1 mM sodium pyruvate, 0.1 mM β-mercaptoethanol, 100 U/ml penicillin, 0.1 mg/ml streptomycin and 2 mM L-glutamine; supplemented with cytokines: BMP4 (500 ng/ml), LIF (1000 U/ml), SCF (100 ng/ml), BMP8a (500 ng/ml) and EGF (50 ng/ml)). For large scale (CRISPR screen) PGCLC inductions (>4000 embryoids per replicate), 30 μl of PGCLC medium was used for each embryoid, with 20 μl additional PGCLC medium added on d2. For smaller-scale PGCLC inductions, 150 μl of PGCLC medium was used per embryoid and another 50 μl of PGCLC medium was added on d2. To minimise evaporation, only the inner 60 wells were used during the screen and 200 μl of PBS was added to the outer 36 wells. After 2–6 days, EBs were collected for FACS or gene expression analyses. SGET EpiLC were also seeded in GK15 medium without cytokines as a negative control to normalise for background levels of SGET fluorescence in embryoids. Where indicated, WNT inhibitor XAV939 (10 μM) or agonist CHIR99021 (3 μM) were added upon induction of PGCLC in embryoids. Where indicated DOX was added at 200 ng/μl upon induction of PGCLC in embryoids+/− above mentioned cytokines.

Generation and validation of the SGET reporter. Embryonic stem cell lines carrying *Stella*-GFP BAC (SG)[52] were re-derived as previously described, and routinely checked for mycoplasma-negative status using the mycoalert detection kit (Lonza). Low-passage SG lines were used to target an in-frame tdTomato cassette heterozygously into exon 1 of *Esg1* (ET) by homologous recombination, thereby generating SGET ESC. Targeting vector(s) for *Esg1* were constructed by amplifying homology arms from genomic DNA of mouse ESC by PCR using PrimeSTAR MAX (Takara Bio, Otsu, Japan), tdTomato from ptdTomato-N1 (Takara Bio, Otsu, Japan) and a destabilised domain (d2) with T2A BSD from pL1L2_IRESdCherryT2Abla-pAneotk vector (a gift from core facility in Cambridge Stem Cell Institute), and assembled using InFusion HD cloning kit (Takara Bio, Otsu, Japan). For gene targeting, 1×10^7 low-passage (< p10) *Stella*-GFP ESC suspended in PBS were electroporated with 40 μg linearised targeting vector using conditions: 800 V, 10 F 49 in Gene Pulser equipment (Bio-Rad). ESCs were selected with 10 mg/ml blasticidin, and colonies were picked and screened for correct targeting by PCR.

We tested for faithful reporter activity using tetraploid embryo complementation (~E8.5) and chimera formation assays (E12.5). For tetraploid complementation assay, 2-cell stage diploid embryos were collected in M2 medium from the oviduct of mice at E1.5, and washed three times with medium containing 0.01% polyvinyl alcohol (Sigma), 280 mM Mannitol (Sigma), 0.5 mM Hepes (Sigma) and 0.15 mM MgSO4 (Sigma). Electrofusion of blastomeres to produce tetraploid embryos was subsequently carried out using a DC pulse (100 V/mm, 30 ms, 1 time) followed by application of AC pulses (5 V/mm, 10 s) using ECM 200 (BTX, Holliston, MA). Tetraploid embryos were transferred into KSOM medium (Merck Millipore) and cultured for 24 h for 4-cell/morula injection. For chimera formation assay, morula stage diploid embryos were collected in M2 medium from the oviduct of mice at E2.5. For micro-manipulation, a piezo-driven micro-manipulator (Prime Tech, Tokyo, Japan) was used to drill through the zona pellucida and 5–10 SGET ESC were introduced into the perivitelline space of morula stage tetraploid or diploid embryos. They were cultured to the blastocyst stage and then transferred into the uteri of pseudopregnant recipient MF1 female mice (E2.5). Post-implantation embryos were dissected at ~E4.5, E6.0, E7.0, E8.5 and E12.5 to analyse GFP and tdTomato expression. Notably these assays also confirm the authentic pluripotent status of our SGET line(s). All husbandry and experiments involving mice were authorised by a UK Home Office Project License 80/2637 and carried out in a Home Office-designated facility.

Lentiviral CRISPR screen. A genome-wide lentiviral CRISPR gRNA library was utilised that contains 87,897 gRNAs targeting 19,150 mouse protein-coding genes[21], with up to five gRNAs per gene (Addgene: #50947). The gRNA library was amplified with NEB 10-beta electrocompetent *Escherichia coli* (NEB) as per the recommended protocol. Briefly, *E. coli* were transformed via high-efficiency electroporation, to ensure faithful library replication, and incubated at 37 °C for 1 h with SOC recovery medium (ThermoFisher) before growing in 500 ml 2xTY (16 g/l Tryptone, 10 g/l Yeast Extract, 5.0 g/l NaCl) + ampicillin (50 μg/ml) medium at 37 °C overnight with 230 rpm shaking. Plasmid was purified from 500 ml bacterial cultures using an endotoxin-free plasmid maxi kit (Qiagen) as per manufacturer's instructions. Faithful replication/amplification of the library was confirmed by Illumina sequencing. For lentiviral production, 4×10^6 HEK 293T cells cultured in DMEM (Invitrogen), 10% FCS, L-glutamine and Penicillin-Streptomycin were seeded onto 10 cm poly-lysine coated plates. The following reagents were prepared in OptiMEM (ThermoFisher) per plate: 10 μg CRISPR plasmid library, 10 μg VSVG-pcDNA3 (Addgene:#8454), 10 μg pCMV-dR8.2 (Addgene:#8455) and 30 μl of Lipofectamine 2000 (ThermoFisher), and incubated with cells for 5–6 h at 37 °C in a 5% CO_2 incubator. Medium was subsequently replaced to include Forskolin (5 μM/ml), and viral supernatant was collected after 48 h and 64 h. Cell debris were removed by centrifugation at 1200 rpm for 15 min at 4 °C, and the supernatants were filtered with 0.45 μM filter. To concentrate the virus supernatant, it was combined with 4xPEG solution (320 ml of 50% PEG 6000, 40 ml of 5 M NaCl, 20

ml of 1 M HEPES, adjust pH to 7.4 and add ddH20 until 500 ml). The virus/PEG mixture was incubated for 4 h at 4 °C, centrifuged at 2600 × *g* for 30 min at 4 °C, and the supernatant aspirated; the process was repeated twice. The viral titre (multiplicity of infection) was assayed by determining the percentage of cells with viral integration after transduction by flow analysing BFP-positive cells and confirmed by qPCR (for BFP).

For the CRISPR screen we first generated SGET ESC that carried constitutive expression of *Cas9* by introducing a CAG-*Cas9*-pA cassette into SGET cells via PiggyBac-mediated transposition. Clones were subsequently picked and screened to identify a single-copy integration of *Cas9*, and to confirm functional capacity to direct high-efficiency gRNA-directed DNA indels using a GFP reporter. To introduce the CRISPR library, a total of 5×10^7 SGET-*Cas9* ESC were transduced with lentiviral library using a multiplicity of infection (MOI) of 0.3, with biological replicates infected on independent occasions (gRNA fold coverage per replicate 146×; 296× total). Transduced SGET ESC were maintained on fibronectin-coated T175 flasks in N2B27, 2i/LIF and 1% KSR, with 8 ng/µl of polybrene added during lentiviral infection. Medium was changed 24 h later and cells were selected with puromycin (1.2 µg/ml) for 5 days. High numbers (>30mio) of transduced ESC were passaged every 2–3 days at low ratio to ensure the complexity of the gRNA library was maintained. ESC pellets were taken at day 12 for analysis and EpiLC were induced on day 12 post-lentiviral transduction, from which PGCLC were subsequently induced.

To quantify the relative frequencies of integrated gRNAs in each population during cell fate transitions, genomic DNA was isolated from day 12 ESC post-lentiviral transduction, from SG^-ET^+ and SG^+ET^+ EpiLC populations at 42 h, and from SG^+ET^{low} PGCLCs from d6 embryoids, in biological duplicate using the DNeasy Blood & Tissue kit (Qiagen). Purified genomic DNA was used as template for custom primers that specifically amplify the gRNA region, and include overhanging Illumina adaptors and indexes to allow deep sequencing and multiplexing, respectively (synthesised as Ultramers from IDT) (see oligo tables). Prior to PCR, genomic DNA was sonicated with a Biorupter (Diagenode) for 15 s on 'LOW' power to improve efficiency of amplification. gRNA sequences were amplified in multiple PCR reactions to enable all isolated DNA to be utilised, each with NEBNext® Q5 HotStart HiFi PCR Master Mix (NEB), 0.2 µM of universal forward primer (mix of staggers), 0.2 µM of indexed reverse primer and genomic DNA (~650 ng/50 µl rxn). For amplification from the plasmid vector, 15 ng was used as template for replicate amplifications. The following cycling conditions were used: 95 °C for 2 min, then between 21 and 26 cycles of (98 °C for 20 s, 62 °C for 20 s and 72 °C for 20 s) followed by 72 °C for 1 min and hold at 4 °C. The number of cycles was optimised for the point at which PCR products can be first visualised on an agarose gel. The PCR reaction was subsequently purified with AMPure® XP beads (Beckman Coulter) using double size selection to remove primer dimers and genomic DNA. Briefly, 0.55x of AMPure beads were added to PCR reaction (1×) and incubated for 5 min at room temperature. The supernatant was collected; the beads contained the unwanted larger fragments and were discarded. An additional 0.3× of AMPure beads was subsequently added to the supernatant (1×) and incubated for 5 min at room temperature. The supernatant was removed and beads washed twice with 200µl of 80% EtOH, air dried for 5 min and DNA was eluted from AMPure beads with EDTA-low TE buffer. The concentration of adaptor-ligated gRNA amplicons was measured with the Qubit DNA Assay kit (ThermoFisher Scientific) and the fragment distribution determined with an Agilent D1000 ScreenTape System (Agilent Technologies). Libraries were subsequently multiplexed and sequenced with an Illumina HiSeq 2500 using single-end 50 bp reads. As the sequences for all samples were identical up to the gRNA region, these types of low complexity libraries can produce low quality data. To counteract this, forward staggered primers were introduced at equal ratios during PCR amplification, generating offset reads. In addition, libraries were sequenced with four dark cycles and low density (70%) clustering.

Gene editing in ESC cell lines. Targeted knockout ESC lines were generated using CRISPR genome editing by either deleting a critical exon to generate a frame-shifted null-allele or via inducing frame-shifting null indels into early coding exonic sequences. In brief, 20nt gRNAs encoding complementary sequences to the region/gene to be targeted were cloned into px459 (v2.0) (Addgene:#62988) using the *Bbs1* sites and transfected into SGET ESC with lipofectamine 2000 (ThermoFisher), according to the manufacturers guidelines. A table of the gene targeting gRNAs and the editing strategy is shown below. Transfection of ESC was selected for with 1.2 µg/ml puromycin for 48 h, and ESC were subsequently passaged into 6-well plates at low density (1000–5000 ESC per well). One week later colonies were picked and expanded before undergoing genotyping by amplicon sizing or using the tracking of indels by decomposition tool (https://tide.nki.nl). Sanger sequencing on selected clones was subsequently used to confirm the precise mutant sequence of each allele. Multiple mutant lines were catalogued and validated before use in functional assays, where WT and/or heterozygous clones derived from the same targeting process were utilised as control, along with the parental SGET line. For doxycycline-inducible expression, *Zfp296*, *Nr5a2* or *Prdm14* cDNA was cloned into a custom PiggyBac vector downstream of a DOX-responsive promoter, and transfected into SGET ESC in conjunction with the reverse transactivator (TET-3G). Genomic integration was selected for with 250 µg/ml neomycin.

Flow cytometry. For FACS, cultured cells or embryoids were dissociated into single cells with TrypLE, and suspended in phosphate buffered saline (PBS) with 0.5% BSA or 3% FBS. Filtered cell suspensions were subsequently flow sorted using a Sony SH800Z or a MoFlow high-speed cell sorter (Beckman Coulter) or analysed with BD LSRFortessa X-20 (BD Biosciences). For exit from pluripotency analyses, cells were sorted/analysed based on absence or presence of *Stella*-GFP expression, while maintaining *Esg1*-tdTomato activation (SG+/−ET+). For PGCLC, cells were sorted/analysed based on *Stella*-GFP re-activation coupled with low *Esg1*-tdTomato. The threshold(s) levels for sorting was set and normalised based on expression levels in pre-optimised SGET ESC. A negative population of ESC without any fluorescence was used to set absolute thresholds. Forward and side scatters were used to gate for the cell population and doublets. For each sort the maximum number of cells was collected, while for flow analysis at least 10,000 cell datapoints were captured to achieve a representative sample.

RNA-seq and gene expression. For RNA-seq of $Nr5a2^{-/-}$, $Zfp296^{-/-}$ and matched WT control cell lines, EpiLC, d2 or d6 PGCLC were dissociated into single-cell solutions with TrypLE and sorted with a Sony SH800Z or a MoFlow high-speed cell sorter (Beckman Coulter) based on appropriate *Stella*-GFP and *Esg1*-tdTomato expression. Samples were collected in 150 µl of extraction buffer from the PicoPure RNA isolation kit (Life Technologies) and rapidly frozen on dry ice or liquid nitrogen. Total RNA was subsequently extracted using the PicoPure RNA isolation Kit, including a 15 min on-column DNaseI digestion. RNA integrity number was assessed with RNA HS ScreenTape (Agilent), and all samples confirmed to have a RIN > 8.5. 100 ng of total RNA from triplicate biological independent experiments was used as input for the NEBNext Ultra RNA library Prep Kit for Illumina® (NEB), with the library generated as per manufacturer's instructions. During the final PCR stage, 15 cycles of amplification were performed to generate the adaptor ligated, fragmented cDNA for sequencing. Samples were assessed using the High Sensitivity D1000 ScreenTape assay (Agilent) to ensure the library did not contain primer-dimer contamination after last round of AMPure beads cleaning. The Qubit dsDNA HS assay kit (ThermoFisher Scientific) and the NEBNext Library Quant Kit for Illumina® were used to accurately quantify the concentration of each library preparation. Samples were multiplexed with 12 indexes per lane and a total of two lanes of sequencing on an Illumina HiSeq2500 or HiSeq4000, single-end 50 bp, with an average depth of approximately 20 million reads per sample.

For RNA-seq from CRISPR screen samples, 50 ng of total RNA was used as input for the Ovation RNA-seq System v2 (Nugen) as per manufacturer's instructions. Amplified double-stranded (ds) cDNA was diluted into EDTA-low TE (Agilent) and sheared into ~230 bp in length using S220 Focused-Ultrasonicator (Covaris) using settings: duty factor 10%, cycle burst 200, intensity 5, temp at 4 °C and treatment time of 5 min per sample. A high Sensitivity D1000 ScreenTape assay (Agilent) was used to assess the efficiency of library fragmentation. Fragmented ds-cDNA was concentrated with Qiagen Reaction Clean Up kit (MiniElute) and 1 µg of the fragmented ds-cDNA used as input for the library preparation using Encore Rapid DR Multiplex Library System (Nugen). This kit ligated the adaptors to repaired-end ds-cDNA without amplification, to mimimise biases introduced during PCR amplification. The KAPA Library Quantification Kit (Kapa bioscience) was used to quantify the concentration of each adaptor-ligated libraries prior to multiplexing. The ESC, EpiLC and PGCLC RNA sequencing libraries were generated in parallel for each replicate, and the biological replicates were generated on independent occasions. Samples were multiplexed and sequenced with Illumina HiSeq1500, single-end 50 bp read length with a minimum depth of ~15 million reads per sample.

DNA methylation. Global DNA methylation levels were determined using the LUminometric Methylation Analysis (LUMA) method[53] and bisulfite pyrosequencing. Briefly, genomic DNA was isolated from purified PGCLC using the DNeasy Blood & Tissue kit (Qiagen) and treated with RNase. 50–100 ng of DNA was digested with MspI/EcoRI and HpaII/EcoRI (NEB) and the subsequent methylation-sensitive overhangs were quantified by Pyrosequencing (PyroMark Q24 Advance) with the dispensation order: GTGTGTCACACAGTGTGT. Global CpG methylation levels were determined from relative peak heights at position(s) 7, 8, 13 and 14 using the formula: $[(2*(p7*p13))/(p8+p14)]^{HpaII}/[(2*(p7*p13))/(p8+p14)]^{MspI}$. CpG methylation at LINE1 loci was determined by bisulfite pyrosequencing using the EpiTect bisulfite kit (Qiagen). PCR amplification and assay design were performed as previously described[54].

Bioinformatics. For RNA-seq, expression reads were quality-trimmed using *TrimGalore* to remove adaptors, and aligned to the mouse reference genome (GRCm38/mm10) using *TopHat2* guided by ENSEMBL gene models. Mapped reads were imported into *SeqMonk* analysis software, normalised, and quantitated using the RNA-seq quantitation pipeline. Differentially expressed genes (DEG) were identified using the DESeq2 statistical package with *fdr* significance thresholds set as $p < 0.05$, filtering for a fold-change (FC) > 2 and mean RPKM expression > 1 in at least one sample. Analysis of gene ontology enrichment was performed using DEG datasets in the DAVID bioinformatics resource v6.7 (https://david.ncifcrf.gov) and/or the REVIGO visualisation tool (http://revigo.irb.hr). Heatmaps and

plots were generated using custom scripts in the R statistical package. For analysis of the CRISPR screen, demultiplexed gRNA sequences were extracted from Illumina reads using custom scripts to account for staggered start positions. These were quantitated and statistically analysed using model-based analysis of genome-wide CRISPR–Cas9 knockout (MAGeCK) v.0.5.2 software in python[22]. Library sequences were trimmed to exclude the bottom 1% of reads from the initial library, which primarily corresponded to zero-count gRNAs. Fold-changes in gRNA frequency were calculated as the logarithmic of the mean fold-change between each gRNA set(s) that target a common gene between paired fate transitions. Statistical analysis to identify significantly affected genes was performed using a 'total' normalisation method in MAGeCK to compare gRNA counts with the preceding cell population during ESC-EpiLC-PGCLC transit (or with the vector library in the case of ESC). We subsequently applied the relative ranking algorithm, taking genes scoring $p < 0.01$ for negative or positive enrichment in both independent replicate screens as candidates for further validation.

Data availability

All raw sequencing data related to this study has been deposited in the gene expression omnibus (GEO) repository under accession GSE117473. Other relevant data and reagents that support the findings of this study are available from the corresponding author(s) on reasonable request. Information on oligonucleotide sequences can be found in the supplementary material.

Received: 5 February 2018 Accepted: 8 August 2018
Published online: 16 October 2018

References

1. Ohinata, Y. et al. A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571–584 (2009).
2. Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found. Symp.* **182**, 68–84 (1994). discussion 84-91.
3. Saitou, M. & Yamaji, M. Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences. *Reproduction* **139**, 931–942 (2010).
4. Hackett, J. A., Zylicz, J. J. & Surani, M. A. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet.* **28**, 164–174 (2012).
5. Aramaki, S. et al. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell* **27**, 516–529 (2013).
6. Nakaki, F. et al. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222–226 (2013).
7. Tu, S. et al. Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature* **534**, 387–390 (2016).
8. Nady, N. et al. ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *eLife* **4**, e10150 (2015).
9. Tang, W. W., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M. A. Specification and epigenetic programming of the human germ line. *Nat. Rev. Genet.* **17**, 585–600 (2016).
10. Kurimoto, K. et al. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22**, 1617–1635 (2008).
11. Yamaji, M. et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* **40**, 1016–1022 (2008).
12. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
13. Saitou, M. & Miyauchi, H. Gametogenesis from pluripotent stem cells. *Cell Stem Cell* **18**, 721–735 (2016).
14. Buecker, C. et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
15. Zylicz, J. J. et al. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *Elife* **4**, e09571 (2015).
16. Hikabe, O. et al. Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature* **539**, 299–303 (2016).
17. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
18. Han, K. et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35**, 463–474 (2017).
19. Parnas, O. et al. A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* **162**, 675–686 (2015).
20. Hackett, J. A. et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452 (2013).
21. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
22. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
23. Hackett, J. A. & Surani, M. A. Regulatory principles of pluripotency: from the ground state up. *Cell. Stem. Cell.* **15**, 416–430 (2014).
24. Dunn, S. J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. G. Defining an essential transcription factor program for naive pluripotency. *Science* **344**, 1156–1160 (2014).
25. Merkle, F. T. et al. Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229–233 (2017).
26. Martello, G. et al. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491–504 (2012).
27. Fidalgo, M. et al. Zfp281 functions as a transcriptional repressor for pluripotency of mouse embryonic stem cells. *Stem Cells* **29**, 1705–1716 (2011).
28. Jackson, M. et al. Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol. Cell Biol.* **24**, 8862–8871 (2004).
29. Yamada, Y., Aoki, H., Kunisada, T. & Hara, A. Rest promotes the early differentiation of mouse ESCs but is not required for their maintenance. *Cell Stem Cell* **6**, 10–15 (2010).
30. Pare, J. F. et al. The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. *J. Biol. Chem.* **279**, 21206–21216 (2004).
31. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
32. Matsuura, T., Miyazaki, S., Miyazaki, T., Tashiro, F. & Miyazaki, J. I. Zfp296 negatively regulates H3K9 methylation in embryonic development as a component of heterochromatin. *Sci. Rep.* **7**, 12462 (2017).
33. Fischedick, G. et al. Zfp296 is a novel, pluripotent-specific reprogramming factor. *PLoS ONE* **7**, e34645 (2012).
34. Fidalgo, M. et al. Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proc. Natl Acad. Sci. USA* **109**, 16202–16207 (2012).
35. Guo, G., Huang, Y., Humphreys, P., Wang, X. & Smith, A. A PiggyBac-based recessive screening method to identify pluripotency regulators. *PLoS ONE* **6**, e18189 (2011).
36. Kumar, R. M. et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516**, 56–61 (2014).
37. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).
38. Whyte, W. A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
39. Gabut, M. et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146 (2011).
40. Wicks, S. J. et al. The deubiquitinating enzyme UCH37 interacts with Smads and regulates TGF-beta signalling. *Oncogene* **24**, 8080–8084 (2005).
41. Wagner, R. T., Xu, X., Yi, F., Merrill, B. J. & Cooney, A. J. Canonical Wnt/beta-catenin regulation of liver receptor homolog-1 mediates pluripotency gene expression. *Stem Cells* **28**, 1794–1804 (2010).
42. Botrugno, O. A. et al. Synergy between LRH-1 and beta-catenin induces G1 cyclin-mediated cell proliferation. *Mol. Cell* **15**, 499–509 (2004).
43. Aaronson, Y., Livyatan, I., Gokhman, D. & Meshorer, E. Systematic identification of gene family regulators in mouse and human embryonic stem cells. *Nucleic Acids Res.* **44**, 4080–4089 (2016).
44. Guo, G. & Smith, A. A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development* **137**, 3185–3192 (2010).
45. Wang, W. et al. Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc. Natl Acad. Sci. USA* **108**, 18283–18288 (2011).
46. Ng, J. H. et al. In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell.* **24**, 324–333 (2013).
47. Fujii, Y., Kakegawa, M., Koide, H., Akagi, T. & Yokota, T. Zfp296 is a novel Klf4-interacting protein and functions as a negative regulator. *Biochem. Biophys. Res. Commun.* **441**, 411–417 (2013).
48. Hackett, J. A., Kobayashi, T., Dietmann, S. & Surani, M. A. Activation of lineage regulators and transposable elements across a pluripotent spectrum. *Stem Cell Rep.* **8**, 1645–1658 (2017).
49. Tang, W. W. et al. A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**, 1453–1467 (2015).
50. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

51. Tzelepis, K. et al. A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep.* **17**, 1193–1205 (2016).
52. Payer, B. et al. Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genesis* **44**, 75–83 (2006).
53. Karimi, M., Luttropp, K. & Ekstrom, T. J. Global DNA methylation analysis using the luminometric methylation assay. *Methods Mol. Biol.* **791**, 135–144 (2011).
54. Walter, M., Teissandier, A., Pérez-Palacios, R., Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**, e11418 (2016).

Acknowledgements

We thank Andy Riddell, Charles Bradshaw and all members of the M.A.S. and J.A.H labs for discussions and support. Funding for this study came from a Wellcome Trust programme grant (C6946/A14492) and Cancer Research UK/Wellcome Trust (092096) core grants to M.A.S, and from a core European Molecular Biology Laboratory (EMBL) grant to J.A.H. The Newton Trust and a Leverhulme Trust Early Career fellowship supported U.G. The James Baird Fund supported Y.H.

Author contributions

J.A.H. designed the study, performed experiments and bioinformatics, and wrote the manuscript. Y.H. performed experiments and bioinformatics. U.G. performed experiments and wrote the manuscript. K.H.-G. performed experiments. T.K. performed experiments. M.A.S. supervised the study and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06230-0.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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