De novo DNA methylation suppresses aberrant fate determination during pluripotency transition

Masaki Kinoshita[1,4], Meng Amy Li[1,4], Michael Barber[1], William Mansfield[1], Sabine Dietmann[1] and Austin Smith[1,2,3,5]

[1] Wellcome-MRC Cambridge Stem Cell Institute
Jeffrey Cheah Biomedical Centre
University of Cambridge
Cambridge CB2 0AW
United Kingdom

[2] Department of Biochemistry
University of Cambridge
Cambridge CB2 1GA
United Kingdom

[3] Living Systems Institute
University of Exeter
Exeter EX4 4QD
United Kingdom

[4] Equal contribution

[5] Author for correspondence: austin.smith@exeter.ac.uk

ABSTRACT

Genome remethylation is essential for mammalian development. Here we examined cell fate in the absence of *de novo* DNA methyltransferases. We found that embryonic stem (ES) cells deficient for *Dnmt3a* and *Dnmt3b* are rapidly eliminated from chimaeras. Pluripotency progression is derailed towards extra-embryonic trophoblast. This aberrant trajectory is propelled by failure to methylate and suppress expression of *Ascl2* during formative transition. *Ascl2* deletion rescues transition and improves contribution to chimaeric epiblast but mutant cells are progressively lost in later development. These findings indicate that methylation constrains transcriptome trajectories during developmental transitions by silencing potentially disruptive genes.

Keywords: DNA methylation; pluripotency; epiblast; cell competition; lineage commitment; embryonic stem cells; de novo methyltransferase, trophoblast, chimera

INTRODUCTION

The mammalian genome is characterised by widespread methylation of cytosine residues. After fertilisation, however, both maternal and paternal genomes undergo extensive demethylation, reaching a low point in the blastocyst (Monk et al. 1987; Borgel et al. 2010; Smallwood et al. 2011; Smith et al. 2012). The embryo genome is then remethylated by the activity of de novo DNA methylation enzymes (Auclair et al. 2014). Mouse embryonic stem (ES) cells exhibit global hypomethylation, similar to the in vivo blastocyst profile (Ficz et al. 2013; Habibi et al. 2013; Leitch et al. 2013). Methylation is gained during the formative pluripotency transition to lineage competence, recapitulating early post-implantation development in vivo (Shirane et al. 2016; Kalkan et al. 2017).

Mammals have three DNA methyltransferases. Dnmt1 maintains methylation during DNA replication, while Dnmt3a and Dnmt3b are responsible for de novo methylation. DNA methylation is not required for general cell viability and, with the exception of imprint control regions, largely occurs at seemingly non-functional regions of the genome (Edwards et al. 2017). Nonetheless, knockout of *Dnmt1* in mice results in embryonic lethality around E10.5 (Li et al. 1992). *Dnmt3a* mutants die during puberty, but *Dnmt3b* mutant embryos fail from E9.5 onwards, exhibiting multiple abnormalities. When both de novo methyltransferases are inactivated, development is disrupted by E8.5 with defective somitogenesis and abnormal morphogenesis (Okano et al. 1999). Apparently normal progress to late gastrulation suggests that remethylation in the early post-implantation embryo is not critical for epiblast transition or germ layer formation. However, the reason(s) for the subsequent catastrophic failure is unclear.

We previously showed that Dnmt3a and Dnmt3b contribute to timely exit from naïve pluripotency in vitro (Li et al. 2017). Here we investigate the functional consequences of lack of de novo methylation for pluripotency progression and lineage potential at the cellular level.

RESULTS and DISCUSSION

Chimaera colonisation and lineage potential of Dnmt3a/b deficient ES cells

To investigate chimaera colonisation we introduced a constitutive mKO2 fluorescent reporter into previously generated *Dnmt3a* and *Dnmt3b* double knockout (Dnmt3dKO) ES cells (Li et al. 2017). Compound *Dnmt3a/b* mutant embryos are reportedly normal until somitogenesis (Okano et al. 1999). However, after blastocyst injection we found very sparse contributions of mutant cells in pre-somite stage embryos at E7.5 (Fig. 1A). Even at E6.5 contributions were reduced compared with typical ES cell chimaeras (Fig. 1B). Furthermore, some mutant donor cells were located in the extraembryonic ectoderm, a rare occurrence with wild type ES cells (Posfai et al. 2020).

To improve survival of Dnmt3dKO ES cells in chimaeras we introduced a constitutive *BCL2* transgene (Yamane et al. 2005). In E6.5 embryos we observed higher contributions to epiblast, comparable to wildtype ES cells (Fig. 1C). However, we also saw donor cells in extraembryonic regions. We confirmed localization in extraembryonic ectoderm and ectoplacental cone in 5 out of 6 chimaeras examined by confocal microscopy (Fig S1A). We inspected blastocyst stage chimaeras to test whether *Dnmt3a/b* deficiency or *BCL2* expression enabled primary trophectoderm colonization. However, donor cells were correctly localized to the inner cell mass and did not contribute to trophectoderm (Fig. S1B). Thus, the presence of Dnmt3dKO cells in post-implantation trophoblast likely arises by displacement from the epiblast rather than differentiation via trophectoderm.

The poor and aberrant colonisation behaviour of Dnmt3dKO cells prompted us to investigate in vitro differentiation competence. In defined conditions permissive for neural induction (Mulas et al. 2019), Dnmt3dKO cells showed only weak up-regulation of *Sox1* and *Pax6* (Fig. S1C). In response mesendoderm induction, mutant cells up-regulated *T* and *Foxa2* but to a lower level than parental cells (Fig. S1D). In neural conditions we detected substantial and sustained up-regulation of *Ascl2*, *Hand1* and *Tpbpa*, transcription factor genes associated with the trophoblast lineage (Fig. S1C, S1D). To examine further whether Dnmt3dKO cells adopt trophoblast identity, we applied two culture conditions for trophoblast cells (Tanaka et al. 1998; Ohinata and Tsukiyama 2014). Unlike parental cells, Dnmt3dKO cells gained expression of trophoblast progenitor and differentiation markers and showed no or low up-regulation of formative pluripotency factors Otx2 and Oct6 (Fig. 1D, S1E).

We subjected single KO ES cells (Li et al. 2017) to lineage induction (Fig. S1F). Expression of somatic lineage markers was reduced in both mutants and in neural conditions trophoblast genes were up-regulated (Fig. S1F). The phenotype was more marked in *Dnmt3b* KO cells, with higher trophoblast gene induction and lower neural and mesendodermal gene activation. We introduced expression constructs for *Dnmt3a* and *Dnmt3b* into Dnmt3dKO cells (Fig. S1G). Rescued cells displayed normal differentiation with suppression of trophoblast genes (Fig. S1H).

Culutre in activin, FGF2 and KSR (AFK) induces ES cell conversion into post-implantation formative epiblast-like cells (EpiLCs) (Hayashi et al. 2011). Dnmt3dKO cells showed delayed morphological change on day 1 but by day 2 appeared similar to parental cells with a comparable increase in cell number (Fig S1I, S1J). The Rex1::GFPd2 (RGd2) reporter allows reliable monitoring of ES cell exit from naïve pluripotency (Kalkan et al. 2017). Dnmt3dKO cells showed delayed down-regulation in AFK (Fig. S1K), consistent with findings using siRNA (Li et al. 2017). We sorted the GFP low (GFP^{lo}) fraction at day 2 and found that trophoblast genes *Ascl2* and *Tfap2c* were mis-expressed in mutant cells (Fig. 1E). Levels increased further upon continued culture in N2B27 and additional trophoblast markers became prominent. However, the order of appearance differed from the in vivo developmental programme. *Ascl2* and *Hand1*, evident after 48h in AFK, are characteristic of post-implantation trophoblast, whereas *Cdx2* and *Elf5*, associated with primary trophectoderm, only became expressed after further culture (Fig. 1E). Mutant cells gained no or low expression of neural markers, *Sox1* and *Pax6* (Fig. S1L). GFP^{lo} cells transferred into activin and CH gained only modest upregulation of *T* and FoxA2 but did not express most trophoblast markers. We transferred Dnmt3dKO cells from AFK into trophoblast medium (Tanaka et al. 1998). In contrast to parental cells, mutant cells flattened and some became very large with prominent nuclei (Fig. 1F). Propidium iodide staining showed a fraction with greater than 4N DNA content, consistent with polyploid trophoblast giant cell formation, (Fig. 1G).

These findings indicate that Dnmt3a/b function to suppress trophoblast gene activation and facilitate formative transition to somatic lineage competence.

Transcriptome analyses of ES cell differentiation trajectory in the absence of Dnmt3a/b

We examined the initial mis-regulation of gene expression by single cell RT-qPCR. We used *Nanog* (naive), *Otx2* (formative), and *Ascl2* (trophoblast) as representative markers (Fig. 2A, S2A). The majority of parental cells down-regulated *Nanog* and gained *Otx2* after 48h in AFK. Dnmt3dKO cells similarly gained Otx2 but generally retained higher Nanog levels. Most strikingly, *Ascl2* was up-regulated in more than half of the Dnmt3dKO cells and many of these were also positive for both Nanog and Otx2. Unexpectedly, we also detected a fraction of triple positive cells among parental cells (Fig. 2A, S2A).

We extended this analysis using the 10x Genomics platform for single cell RNA-sequencing (scRNA-seq). UMAP analysis clustered cells by culture condition in the first dimension and genotype in the second dimension (Fig 2B). We inspected markers for pluripotency states, germ layers and trophoblast (Fig. S2B). In ES cells expression was not significantly altered between parental and mutant. However, in mutant 48h AFK cells we observed persistent expression of multiple naïve genes, lower up-regulation of formative genes, and ectopic expression of several trophoblast genes. We examined *Nanog*, *Otx2* and *Ascl2* using a UMI count above zero to classify expression. *Nanog*$^+$*Ascl2*$^+$, *Otx2*$^+$*Ascl2*$^+$ and *Nanog*$^+$*Otx2*$^+$*Ascl2*$^+$ cells were present in both parental and Dnmt3dKO cells at 48hr (Fig. 2C). However, the combined proportion of dual and triple positive cells involving Ascl2 was three times higher in the mutants, consistent with single cell RT-qPCR.

Pseudo-time analysis using Monocle 2 (Qiu et al. 2017b) indicated a branchpoint in differentiation trajectory (Fig. 2D). We arbitrarily partitioned cells at and after the branchpoint into 5 groups (a-e). Parental cells were predominantly distributed along path 1 whereas mutant cells were almost exclusively located on path 2. Notably, however, 6.4 % of parental cells initiated path 2, although very few reached the endpoint. Differentially expressed genes in parental cells featured formative markers on path 1 and trophoblast genes on path 2 (Fig. S2C). Differential expression analysis without considering genotypes substantiated these alternative fates (Fig. 2E).

We investigated relatedness between path 2 and in vivo trophoblast. From published data (Smith et al. 2017) we identified genes up-regulated in E3.5 trophectoderm or E6.5 trophoblast relative to inner cell mass and post-implantation epiblast respectively. Correlation was low for E3.5 trophectoderm-enriched genes with either pathway. In contrast, many E6.5 trophoblast-enriched genes were upregulated on path 2 (Fig. S2D).

These analyses indicate that the majority of Dnmt3dKO cells exiting naïve pluripotency adopt a deviant trajectory and develop features of post-implantation trophoblast. Moreover, a small fraction of parental cells initiate this trajectory.

Chromatin accessibility and Ascl2 misexpression in the absence of de novo DNA methylation

Chromatin is remodelled during formative transition (Buecker et al. 2014). We used ATAC-seq (Buenrostro et al. 2013) to survey chromatin in the absence of de novo DNA methylation. We analysed ES cells, 48h GFP^{Hi} transitional cells, and GFP^{Lo} post-transition cells and identified loci that are more open at 48h in Dnmt3dKO cells than parental cells (Log2 fold change >0.7, P-value<0.05). We classified 4 groups according to opening in transitional (GFP^{Hi}, Group I and III) or post-transition populations (GFP^{Lo}, Group II and IV), and presence or absence of a CpG island (GGI), annotated by UCSC genome browser (Fig. 3A). The strongest peaks were associated with CGIs, which were lowly methylated at all stages (Fig. 3B and S3A, S3B). Non-CGI peaks were weaker and within regions that are methylated in parental cells, although a short stretch of reduced methylation was apparent at many Group IV peaks (Fig. 3B and S3A, S3B). As expected, CGI peaks are mainly associated with annotated promoters and non-CGI peaks with enhancers (Fig. S3C).

The analysis revealed chromatin regions that opened transiently in parental cells but remained accessible in Dnmt3dKO cells (group II and IV, Fig 3A, S3B). Example genome browser tracks are shown in Fig. 3C. Genes associated with these peaks showed high correlation with differentially expressed genes in Dnmt3dKO GFP^{Lo} cells (Fig.S3D). We also found a positive correlation with the set of E6.5 trophoblast-enriched genes (Fig. S3E).

Thus, during exit from naïve pluripotency Dnmt3dKO ES cells fail to close down loci that are normally accessible only during transition. These regions encompass promoters and proximal enhancers for a subset of E6.5 trophoblast genes that become mis-expressed.

We conducted transcription factor motif enrichment analysis across ATAC peaks at CGI loci in Dnmt3dKO GFP^{Lo} cells and identified Ascl1 (Fig. 3D). Ascl1 was not expressed in any of the samples studied, but the motif is shared with Ascl2, which, as noted above, is upregulated in transitioning mutant cells (Fig. 3E). The Ascl1/2 motif was present in 531 of 609 promoter regions (-2000 to +500bp around TSS) of E6.5 trophoblast enriched genes. The *Ascl2* locus opened during formative transition in parental ES cells (Fig S3F) but in Dnmt3dKO cells, *Ascl2* promoter accessibility increased further post-transition (Fig. S3F).

Inspection of bisulfite sequencing data (Shirane et al. 2016) revealed that *Ascl2* has a CpG-rich promoter that is lowly methylated in ESCs and gains methylation during naïve to EpiLC transition (Fig. 3F). *In vivo*, *Ascl2* remains hypomethylated in trophoblast lineages (Tanaka et al. 1999). *Ascl2* is located in the *Kcnq1* genomic imprinting cluster. lncRNA, *Kcnq1ot1* suppresses paternal expression of *Ascl2 in cis* in placenta (Bogutz et al. 2018). We observed no change in expression of *Kcnq1ot1* in Dnmt3dKO cells, however (Fig S3G).

We investigated whether expression of *Ascl2* can impose trophoblast-like differentiation by introducing an Ascl2-ER fusion construct into ES cells. Upon tamoxifen (Tx) treatment in serum and LIF, cells changed morphology within two days, becoming larger and flattened (Fig. 4A). *Oct4* was down-regulated and trophectoderm genes *Elf5*, *Cdx2* and *Tpbpa* were up-regulated (Fig. 4B). Thus, misexpression of *Ascl2* can initiate trophoblast-like differentiation.

We then tested whether activation of *Ascl2* was necessary for trophoblast-like differentiation of Dnmt3dKO cells. We deleted *Ascl2* to create Dnmt3dKO∆A2 cells (Fig. S4A) saw that the mis-regulation of trophoblast genes was abolished (Fig. 4C). Furthermore, at 24h Dnmt3dKO∆A2 cells expressed of *Otx2*, *Oct6* and *Fgf5* formative markers (Fig S4B). However, neural genes were not subsequently upregulated and mesendodermal differentiation remained inefficient (Fig. 4D).

These findings indicate that absence of de novo DNA methylation causes failure to silence *Ascl2,* which derails the formative pluripotency transition. It should be noted that misexpression of *Ascl2* in Dnmt3dKO cells does not direct normal trophoblast lineage development but triggers a transdifferentiation process.

Dnmt3a/b deficient cells are outcompeted by wildtype cells

We investigated whether deletion of *Ascl2* and elimination of deviant pluripotency transition may restore chimaera contributions. We saw substantial contributions of Dnmt3dKO∆A2 cells in 5 out of 10 epiblasts at E6.5 (Fig. 5A, S5A). Moreover, we observed no donor cells in extraembryonic regions. At E7.5 contributions were reduced and appeared biased to the posterior region (Fig. 5B, S5B). Immunostaining for Stella and Tfap2c showed no overlap with primordial germ cells (Fig S5C, D). By E9.5, we observed only sparse contributions in 4 of 11 embryos (Fig. 5C, S5E).

Progressive dilution of mutant cells in chimaeras suggested a competitive disadvantage with wild type host epiblast cells (Amoyel and Bach 2014). To examine this possibility we used formative pluripotent stem (FS) cells related to E6.0 epiblast (Kinoshita et al. 2020). We were unable to establish FS cells from Dnmt3dKO cells, likely due to disruptive effects of *Ascl2* misexpression. However, Dnmt3dKO∆A2 ES cells could be converted into stable FS cell lines that expressed the core pluripotency factor Oct4 together with formative markers (Fig. 5D, E). We co-cultured equal numbers of GFP expressing wild type and mKO2 expressing

Dnmt3dKO∆A2 FS cells. Proportions of each genotype were quantified by flow cytometry. We observed reduction in the fraction of mutant cells by passage 0 that was further increased by passage 1 (Fig. 5F). In co-cultures Annexin V staining was higher for Dnmt3dKO∆A2 FS cells during self-renewal and mesendoderm induction ($p<0.05$) though not neural induction ($p>0.05$) (Fig. 5G, S5G).

Relative level of cMyc is a key determinant in cell competition (de la Cova et al. 2004; Claveria et al. 2013; Sancho et al. 2013). By immunostaining we saw that Dnmt3dKO∆A2 FS cells have lower levels of cMyc protein than wild type cells (Fig. 5H). Immunoblotting confirmed that cMyc expression was reduced in mutant cells (Fig. S5F). These observations suggest that cMyc-based cell competition may cause elimination of Dnmt3dKO cells mixed with wild type cells (Sancho et al. 2013) and that inactivation of *Ascl2* avoids only the first round of competition. The mechanism that reduces cMyc is unclear but we surmise is secondary to other perturbations in the absence of de novo methylation.

We introduced *BCL2* into Dnmt3dKO∆A2 ES cells. Under in vitro differentiation conditions *BCL2* transfectants showed up-regulation of *Sox1, Pax6, T and Foxa2*, although still below the levels observed for parental transfectants (Fig. 5I). After blastocyst injection we recovered 5 chimaeras among 7 embryos at E9.5. Contributions were variable, but substantial in 3 of the 5 (Fig. 5J). We also collected embryos at E12.5 and detected mKO2 fluorescence in 2 out of 4 specimens, with donor-derived cells in neural tissues and mesenchyme (Fig. 5K, S5H). These contributions were noticeably lower than at E9.5, however, suggesting ongoing loss of *Dnmt3a/b*-deficient cells even in the presence of a strong survival factor.

Overall, our findings demonstrate that in the absence of de novo DNA methyltransferases the ability to execute a cell state transition is compromised due to misexpression of a gene with fate switching potency. The example of *Ascl2* illustrates how genome-wide methylation has been co-opted to constrain expression of a pivotal gene during network rewiring. This scenario may replay at other critical loci that become accessible incidentally during cell transitions. Multiple abnormalities in *Dnmt3a/b* mutant embryos and inability of mutant ES cells to persist in any lineage (Yagi et al. 2017) are consistent a recurrent contribution of DNA methylation to safeguarding transcriptome trajectories. Mutations in *DNMT3a* and *DNMT3b* are associated with tumourigenesis in various tissues (Zhang and Xu 2017; Yagi et al. 2020), possibly due to corruption of cellular transitions.

MATERIALS AND METHODS

Cell Culture

ES cells were maintained in 2iL medium on gelatin coated plates as described (Mulas et al. 2019). 2iL medium consists of 10ng/ml mouse LIF, 1μM of Mek inhibitor PD0325901 and 3μM of GSK3 inhibitor CHIR99021 in N2B27 basal medium. EpiLCs were induced by plating $2x10^5$ ES cells in 20ng/ml activin A, 12.5ng/ml bFgf and 1% knockout serum replacement (KSR), N2B27 medium on fibronectin coated 6-well plate. Neural differentiation was performed by plating $1x10^5$ cells in N2B27 basal medium on laminin coated 6-well plate. Mesendoderm induction was performed by plating $1x10^5$ cells in 10ng/ml Activin A and 3μM CHIR99021 in N2B27 medium on fibronectin coated 6-well plate. Trophoblast culture media were formulated as described previously (Tanaka et al. 1998; Ohinata and Tsukiyama 2014). For chimaera studies, cells were stably transfected with pPBCAG-mKO2-IP plasmid by TransIT LT1 (Mirus) with pCAG-PBase plasmid and selected with 1 μg/ml puromycin. Human *BCL2* was introduced by pT2PyCAG-hBCL2-IH by TranIT LT1 with pCAG-T2ase plasmid and selection with 100μg/ml of hygromycin B. To establish tamoxifen inducible *Ascl2* expression lines, pPBCAG-

Ascl2ER-IN plasmid was co-transfected with pCAG-PBase plasmid into E14Tg2a ES cells and clones picked after selection with 250µg/ml G418. DnmtdKO∆A2 FS cells were established by conversion of ES cells and maintained in 3ng/ml of activin A, 2µM XAV939 and 1µM BMS493 in N2B27 medium on fibronectin coated plates as described (Kinoshita et al. 2020). Alexa Fluor 647 conjugated Annexin V (Biolegend, Cat#640911) was used for apoptotic cell analysis.

Dnmt3a/3b, _Ascl2_ knockout (KO) ES cells

Dnmt3a or *Dnmt3b* single KO and double KO cells were established previously using CRISPR/Cas9 (Li et al. 2017). Two gRNAs were designed to excise the catalytic domain. *Ascl2* KO was also performed by CRISPR/Cas9. gRNAs were designed to excise Exon2 which encodes full length of protein coding sequences. Clones were screened by genomic PCR. gRNAs and primers are listed in Table S1.

Chimaeric embryo experiments

Reporter expressing ES cells were dissociated and 10-15 cells injected into each blastocyst. Injected embryos were transferred to the uteri of pseudo-pregnant females or cultured in vitro for 24h in M2 medium (Sigma Aldrich) in 7% CO_2 at 37˚C.

Immunostaining

Embryos and culture cells were fixed with 4% PFA at RT (for 15 minutes for cells and upto E7.5 stage embryos, 60 minutes for E9.5 and E12.5). Whole E6.5 embryos were stained with rabbit anti-Eomes antibody (Abcam, Cat#ab23345). Blastocyst were stained with rabbit anti-GFP (Invitrogen, Cat#A-11122) and rat anti-Sox2 (eBioscience, Cat#14-9811-82). After the fixation, E9.5 and E12.5 embryos were incubated with 20 % sucrose/PBS overnight at 4˚C, then embedded in OCT compound. Sections were stained with rat anti-Sox2 (eBiosciencce) and FITC-conjugated mouse anti-cTnT antibodies (Abcam, Cat#ab105439) (E9.5 only). Nuclei were stained with DAPI. Embryos and sections were imaged by Leica SP5 or Zeiss LSM880 confocal microscope. E7.5 PGC were stained with anti-Tfap2c (Santa Cruz, Cat# sc-8977), anti-Stella (Abcam, Cat#ab19878) and rat anti-Sox2 (eBioscience, Cat#14-9811-82). FS cells were stained with rabbit anti-cMyc antibody (Abcam, Cat#ab30272).

Western Blot

Cells were lysed with RIPA buffer in the presence of Protease/Phosphatase inhibitor cocktail (Invitrogen). Lysed cells were rotated for 20 minutes and sonicated in Bioruptor (Diagenode). Cell lysates were cleared by centrifugation, and the supernatant was recovered. Protein concentrations were measured by the BCA method (Pierce). 15 µg of protein was loaded in each well. Blots were blocked with 5% BSA/TBS 0.1 % Triton-X for 1 hour at RT and incubated overnight with primary antibodies at 4˚C. Secondary antibodies were incubated for 1 hour at RT and signals were detected with ECL Select (GE Healthcare) and Odyssey Fc (Li-Cor). NaOH (0.2N) was used for stripping. Anti-cMyc (Abcam, Cat#ab30272) and anti-αTubulin (Thermo Fisher, A-11126) were used.

qRT-PCR

Total RNA was isolated with Relila RNA miniprep systems (Promega) and cDNA was synthesized using GoScript Reverse Transcriptase system (Promega). qRT-PCR was performed with Taqman Gene Expression (Thermo Fisher Scientific) or Fast SYBR Green Master Mix (Thermo Fisher Scientific). Primers are listed in Table S1.

Single cell qPCR

Custom Taqman probes were ordered for Nanog, Otx2, Ascl2 and Actin-beta (Thermo Fisher Scientific). Single cells were collected directly into 96-well plate by flow cytometry (MoFlo, Beckman Coulter) and reverse transcription was performed as described previously (Dunn et

al. 2019). Cells which had Ct-value of Actin-beta >14 and no amplification were excluded from the analysis.

Single cell RNA-seq

ES cells and 48 hours AFK cells were dissociated with accutase (Biolegend) and live cells were collected by flow cytometry (MoFlo, Beckman Coulter). Single cell cDNA libraries were constructed using 10x Genomics technology by Genomics Core Facility, Cancer Research UK Cambridge Institute.

ATAC-seq

50,000 ES cells and AFK cells at 48h were collected by flow cytometry (MoFlo, Beckman Coulter). Cells washed with ice-cold PBS once then lysed in the buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ 0.1% IGEPAL). The nuclear pellets were collected and Tn5 tagmentation and library construction were performed with Illumina Nextera kit (FC-121-1030). DNA was purified with AMPure XP beads (Beckman Coulter).

DATA ANALYSIS

Single Cell RNA-seq analysis

Preliminary sequencing results (bcl files) were converted to fastq files with CellRanger (version 3.0) following the standard 10x Genomics protocol. Barcodes and unique molecular identifier (UMI) ends were trimmed to 26 bp, and mRNA ends to 98 bp. Reads were then aligned to the mouse reference genome (mm10) and gene counts were obtained using the GRCm38.92 annotation in CellRanger. We used Seurat (version 3.1.5) (Butler et al. 2018) to further process the single-cell RNA-seq data. Cells that have between 200 and 5500 unique gene counts, and with less than 5% mitochondrial gene content were filtered. We analysed 2,724 parental and 2,459 Dnmt3dKO ESCs with 4,776 parental and 4,563 Dnmt3dKO 48hr cells. Log-normalization using a scale factor of 10,000 was performed, and the data was scaled to produce standardized expression values for each gene across all cells (z-score transformation), while also regressing out unwanted variation in the percent of mitochondrial gene content. We determined the top 2,000 most highly variable genes, and performed principal component analysis. Visualisation was performed using the UMAP projection method on the first 18 principal components. Differential expression was performed using the Wilcoxon rank sum test, using a threshold of log2 fold change greater than 0.1 and p-value less than 0.05. The enrichment of TE (E3.5) and ExE (E6.5) genes was visualised using the AddModuleScore function of Seurat to show the average expression of sets of genes. The Co-expression of Nanog, Ascl2, and Otx2 was calculated for each cell having at least one unique read for each of the genes.

Pseudotemporal analysis

The pseudotime trajectory was calculated using Monocle2 (version 2.14) (Bell et al. 2012; Trapnell et al. 2014; Qiu et al. 2017a) using the top 1000 differentially expressed genes between each cluster. To find differences between cells at the end and middle of each branch of the trajectory it was split into sections. This was done using an arbitrary pseudotime cut-off of 21.4 for path 1 and 22.25 for path 2 based on visualizing the trajectory. Analysis of trophoblast and trophectoderm gene correlation across pseudotime was performed as follows. The expression of genes enriched in the trophoblast and in trophectoderm plotted across pseudotime, split into the two branches coloured red and blue with bold lines showing the average expression. Pseudotime was scaled up to be in 100 bins with an equal number of cells in each, and then only the second 50 plotted to show the latter half of the time course. Loess regression was used to smooth each of the lines.

Bulk RNA-seq analysis

Genes enriched in the trophoblast over epiblast at E6.5 and enriched in the trophectoderm over ICM at E3.5 were determined by differential expression analysis using publicly available data sets (GSE84236). This was done using tximport (Soneson et al. 2015) to load the dataset into DESeq2 (Soneson et al. 2015) for differential expression analysis using a threshold of log2 fold change > 2 and an adjusted p-value of less than 0.05.

ATAC-seq analysis

Raw reads were pre-processed and quality-filtered using Trim Galore! (https://github.com/FelixKrueger/TrimGalore), and reads shorter than 15 nt were discarded. Processed reads were then aligned to the mouse reference genome (mm10) using bowtie with parameters "-m1 -v1 --best --strata -X 2000 --trim3 1". Duplicates were removed using Picard tools. Read pairs larger than one nucleosome length (146 bp) were discarded, and an offset of 4 nts was introduced. Peaks were called with MACS2 and parameters "--nomodel --shift -55 --extsize 110 --broad -g mm --broad-cutoff 0.1". Then the R package Diffbind[8](Stark 2011) was used to calculate reads across the merged peaks and calculate differential peaks for each cell type utilising the edgeR method (Robinson et al. 2010; McCarthy et al. 2012). Peaks were also identified as being within a CpG island or not and assigned genes based on a window of 2kb around each peak. Scatterplots were plotted using the differential ATAC-seq fold changes from edgeR against the RNA-seq fold changes from Seurat. In each of the quadrants of the scatterplot Fisher's exact test was used to calculate whether the differential expressed genes were overrepresented. Motif analysis was performed by using the findMotifsGenome tool in homer (version v4.10) (Heinz et al. 2010) on peaks that overlapped with CpG islands in the KO using the WT peaks as background sites.

Data Availability

10xGenomics and ATAC-seq data are deposited in GEO: GSE158347.

Acknowledgments

We thank Brian Hendrich and Jennifer Nichols for comments on the manuscript and Giuliano Stirparo for advice on informatics. We are grateful to Tim Lohoff and Jennifer Nichols for discussion of unpublished data. We thank Maike Paramor, Vicki Murray, Peter Humphreys, Darran Clements, Andrew Riddell and biofacility staff for technical support and the CSCI core bioinformatics team for data processing. Sequencing was performed by the CRUK Cambridge Institute Genomics Core Facility. Rosalind Drummond and James Clarke provided laboratory assistance. This research was funded by the Biotechnology and Biological Sciences Research Council (BB/P009867/1, BB/P021573/1). The Cambridge Stem Cell Institute receives core funding from Wellcome (203151/Z/16/Z) and the Medical Research Council (MC_PC_12009). AS is a Medical Research Council Professor (G1100526/1).

Author Contributions

Conceptualisation, M.K., M.A.L., A.S.; Methods, M.K., M.A.L.; Formal Analysis, M.B., S.D.; Investigation M.K., M.A.L., W.M.; Writing M.K., M.A.L., A.S.; Supervision, A.S.

Kinoshita_Fig.1

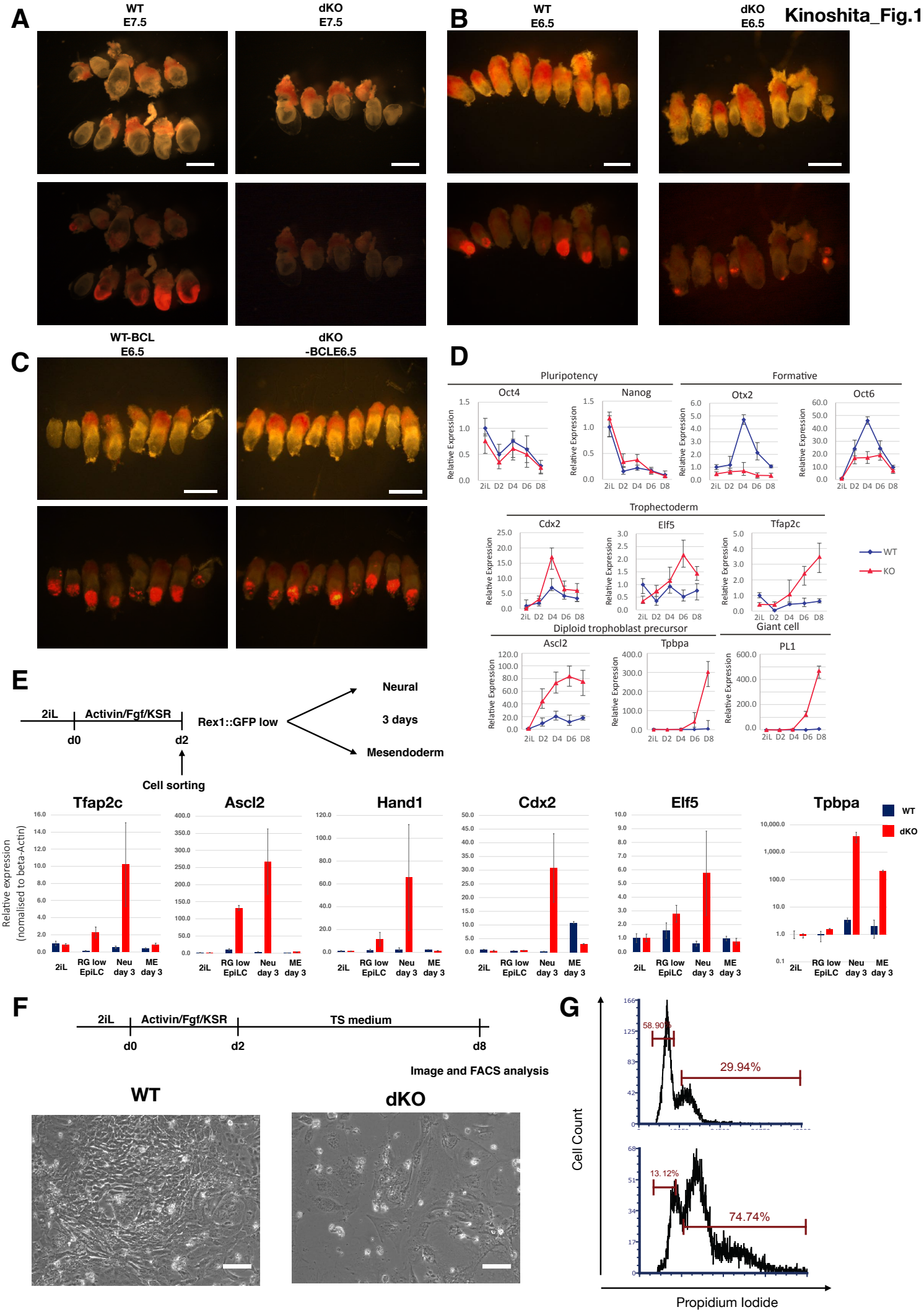


Fig.1 (A-C) Chimaeric embryos obtained with parental, Dnmt3dKO (A-B), and hBCL2 expressing (C) mKO2 reporter ES cells. 30% opacity bright field images overlaid onto fluorescent images. Scale bars, 1 mm (A), 500 μm (B-C). (D) qRT-PCR analysis of marker during culture in trophoblast medium (Tanaka et al. 1998). (E) qRT-PCR analysis of undifferentiated ES cells, Rex1::GFP low sorted AFK cells, and further differentiated cells. (F) Cell morphology of WT and Dnmt3dKO cells in TS cell medium for 8 days. Scale bars, 50 μm. (G) DNA content quantified by propidium iodide (PI) staining at day 8 in TS cell medium. qRT-PCR data were normalized to beta-Actin. Error bars represent S.D. from technical duplicates.

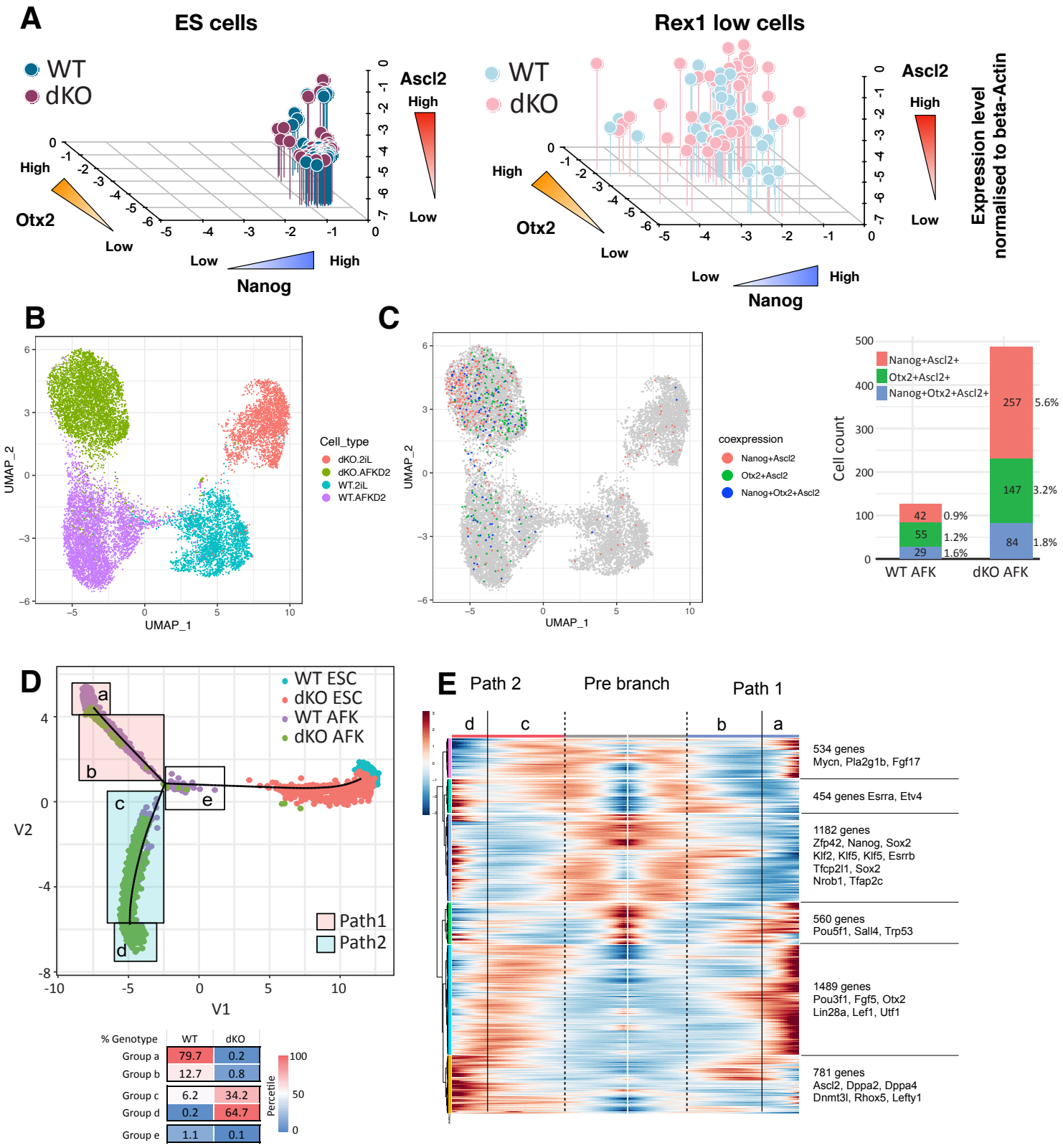
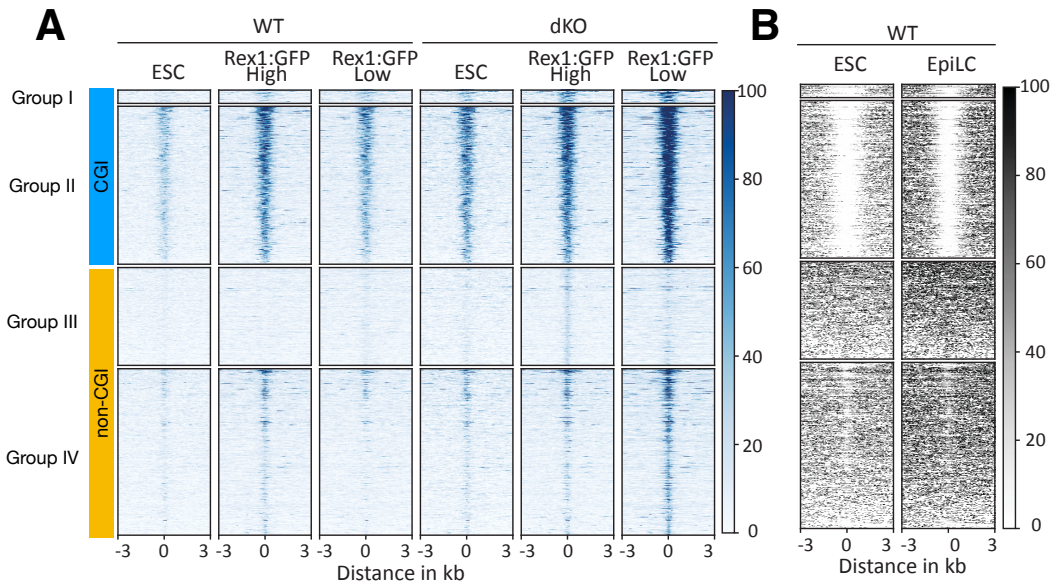
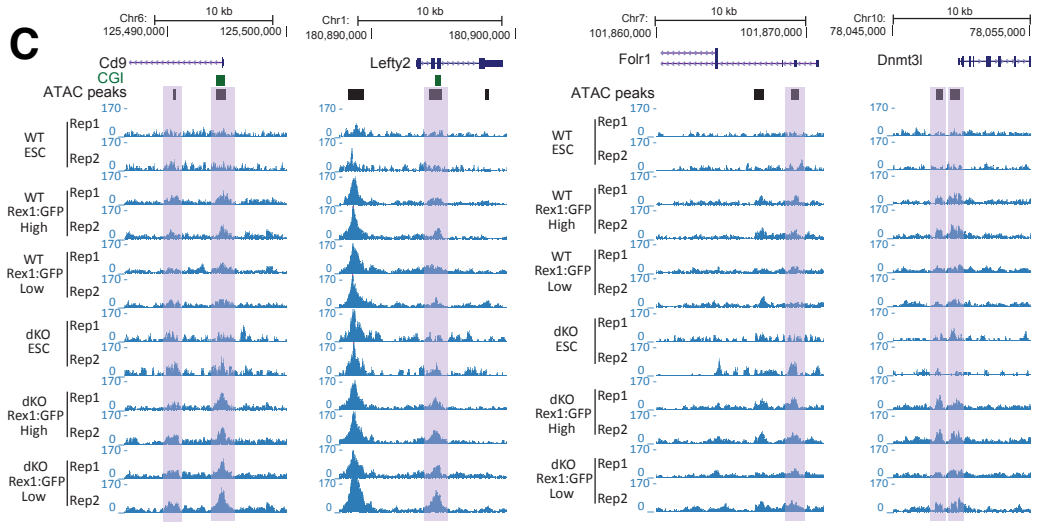
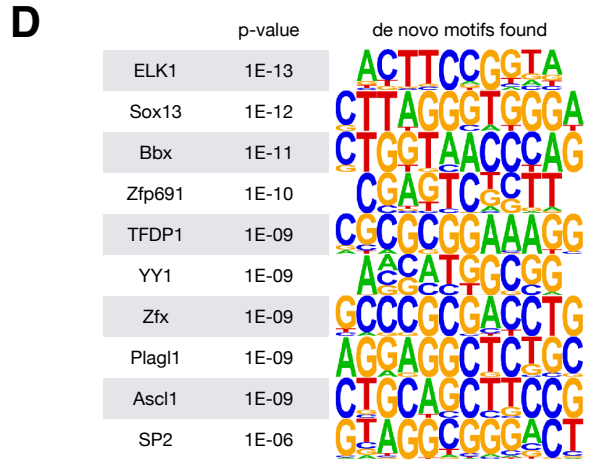
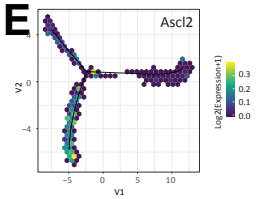

A
ES cells
Rex1 low cells
WT
dKO
Ascl2
High
Low
Otx2
Nanog
Expression level normalised to beta-Actin
B
UMAP_2
UMAP_1
Cell_type
dKO.2iL
dKO.AFKD2
WT.2iL
WT.AFKD2
C
coexpression
Nanog+Ascl2
Otx2+Ascl2
Nanog+Otx2+Ascl2
Nanog+Ascl2+
Otx2+Ascl2+
Nanog+Otx2+Ascl2+
Cell count
WT AFK
dKO AFK
42 0.9%
55 1.2%
29 1.6%
257 5.6%
147 3.2%
84 1.8%
D
WT ESC
dKO ESC
WT AFK
dKO AFK
Path1
Path2
V1
V2
% Genotype WT dKO
Group a 79.7 0.2
Group b 12.7 0.8
Group c 6.2 34.2
Group d 0.2 64.7
Group e 1.1 0.1
Percetile
E
Path 2
Pre branch
Path 1
534 genes
Mycn, Pla2g1b, Fgf17
454 genes Esrra, Etv4
1182 genes
Zfp42, Nanog, Sox2
Klf2, Klf5, Klf5, Esrrb
Tfcp2l1, Sox2
Nrob1, Tfap2c
560 genes
Pou5f1, Sall4, Trp53
1489 genes
Pou3f1, Fgf5, Otx2
Lin28a, Lef1, Utf1
781 genes
Ascl2, Dppa2, Dppa4
Dnmt3l, Rhox5, Lefty1

Fig. 2 (A) Single cell qRT-PCR analysis of ES cells and sorted Rex1::GFP low AFK cells. Expression levels of Nanog (X axis), Otx2 (Y axis) and Ascl2 (Z axis) were normalized to beta-Actin. (B) UMAP of parental and dKO ES cells and AFK 48h cells. (C) UMAP colored to show cells co-expressing $Nanog^{+}Ascl2^{+}$ (red), $Otx2^{+}Ascl2^{+}$ (green) or $Nanog^{+}Otx2^{+}Ascl2^{+}$ (blue) (left) with enumeration (right). (D) Pseudo-time ordering. Proportions of each genotype of AFK 48h cells in areas a-e are shown below. (E) Gene expression heatmap of 5,000 differentially expressed genes in pseudotime.


A
WT
dKO
ESC
Rex1:GFP High
Rex1:GFP Low
ESC
Rex1:GFP High
Rex1:GFP Low
Group I
Group II
Group III
Group IV
CGI
non-CGI
100
80
60
40
20
0
-3 0 3
Distance in kb
B
WT
ESC
EpiLC
Distance in kb


C
Chr6: 125,490,000 – 125,500,000
Chr1: 180,890,000 – 180,900,000
Chr7: 101,860,000 – 101,870,000
Chr10: 78,045,000 – 78,055,000
10 kb
Cd9
Lefty2
Folr1
Dnmt3l
CGI
ATAC peaks
WT ESC
WT Rex1:GFP High
WT Rex1:GFP Low
dKO ESC
dKO Rex1:GFP High
dKO Rex1:GFP Low
Rep1
Rep2
170
0


D
p-value
de novo motifs found
ELK1 1E-13
Sox13 1E-12
Bbx 1E-11
Zfp691 1E-10
TFDP1 1E-09
YY1 1E-09
Zfx 1E-09
Plagl1 1E-09
Ascl1 1E-09
SP2 1E-06


E
Ascl2
V1
V2
Log2(Expression+1)

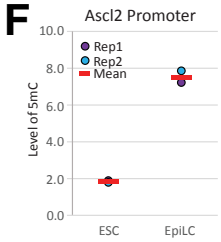

F
Ascl2 Promoter
Rep1
Rep2
Mean
Level of 5mC
10.0
8.0
6.0
4.0
2.0
0.0
ESC
EpiLC

Fig. 3 (A) Heatmaps of ATAC-seq intensity distributions grouped according to increased accessibility in Dnmt3dKO GGP^{Hi} (I and III) or GFP^{Lo} (II and IV) presence or absence of CGI. (B) Heat map of CpG methylation across ATAC-seq peaks in ES cells and EpiLCs (Shirane et al. 2016). (C) Genome browser examples of genes that are more open in Dnmt3dKO cells. (D) De novo motif finding analysis of differential ATAC peaks. (E) Expression of *Ascl2* in pseudo-time plot. (F) *Ascl2* promoter methylation during ES to EpiLC transition (Shirane et al. 2016).

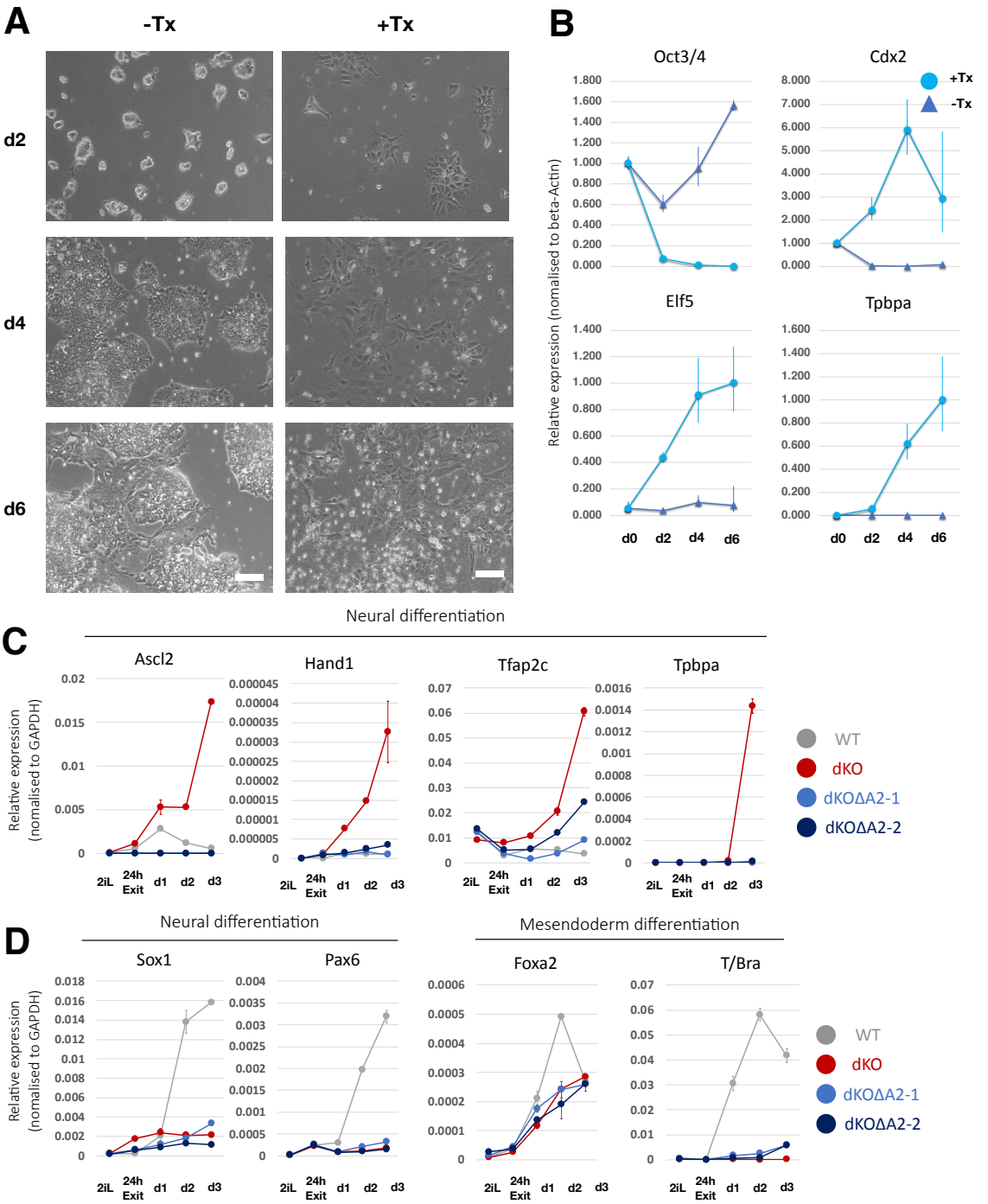
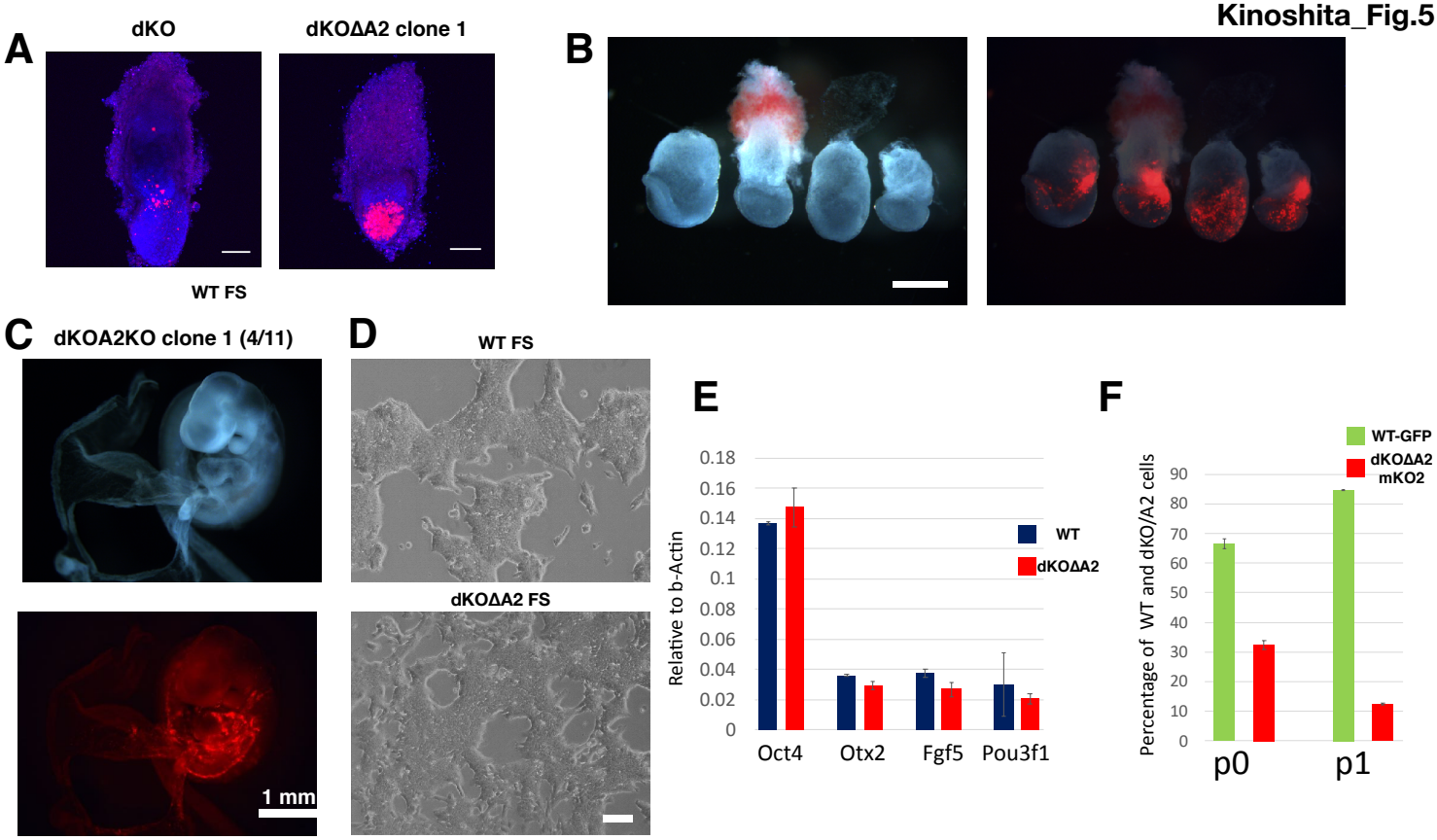
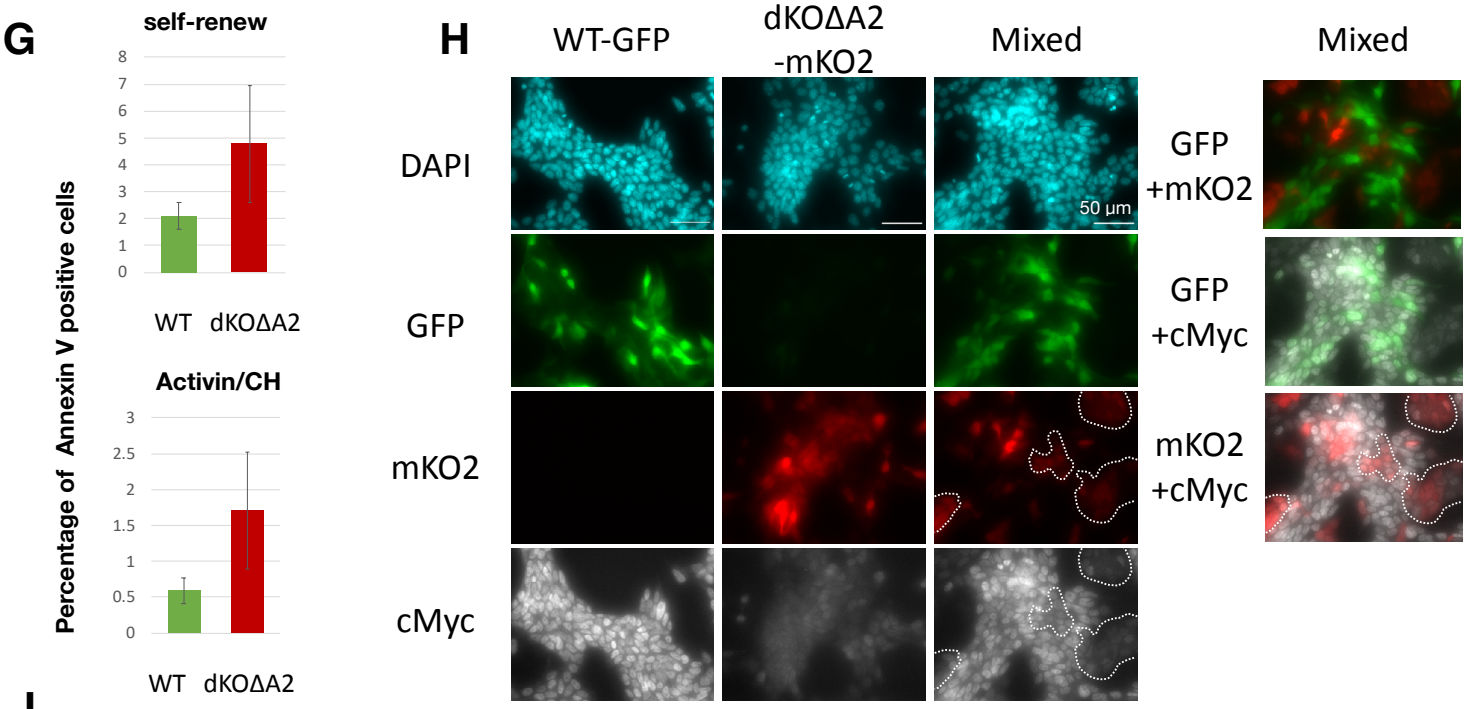
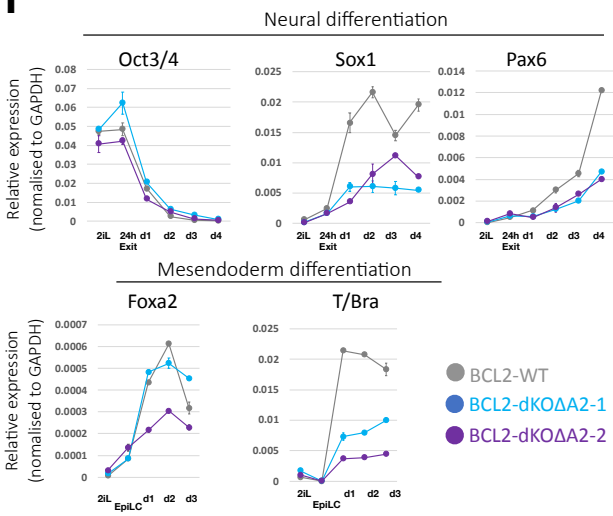

A
-Tx
+Tx
d2
d4
d6
B
Oct3/4
Cdx2
Elf5
Tpbpa
+Tx
-Tx
Relative expression (nomalised to beta-Actin)
d0
d2
d4
d6
C
Neural differentiation
Ascl2
Hand1
Tfap2c
Tpbpa
Relative expression (nomalised to GAPDH)
2iL
24h Exit
d1
d2
d3
WT
dKO
dKOΔA2-1
dKOΔA2-2
D
Neural differentiation
Mesendoderm differentiation
Sox1
Pax6
Foxa2
T/Bra
Relative expression (nomalised to GAPDH)
2iL
24h Exit
d1
d2
d3
WT
dKO
dKOΔA2-1
dKOΔA2-2

Fig. 4 (A) Images of ES cells cultured in serum/LIF with or without tamoxifen (Tx) induced activation of Ascl2-ER. Scale bars, 100 μm. (B) qRT-PCR analysis of marker expression during Ascl2-ER induction. Error bar represents S.D. from technical triplicates. (C) qRT-PCR analysis of two clonal lines of Dnmt3dKO∆A2 cells in neural differentiation culture. Error bars represent S.D. from technical duplicates. (D) Dnmt3dKO∆A2 cells failed to up-regulate lineage markers in neural or mesendoderm differentiation protocol. Error bars represent S.D. from technical duplicates.

Kinoshita_Fig.5







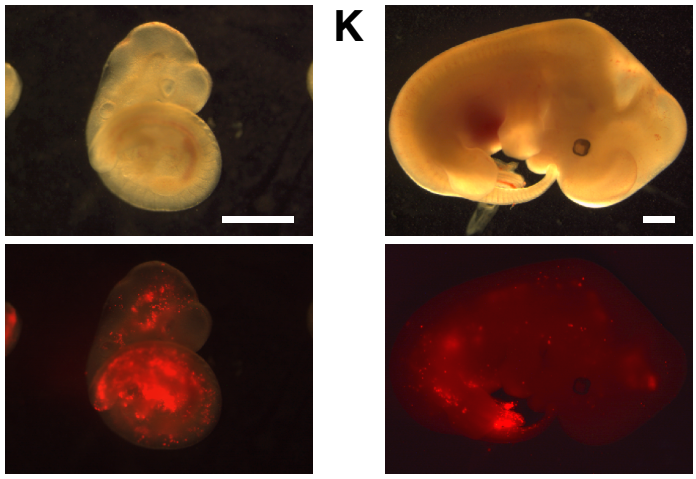


Fig. 5 (A) Maximum projection confocal images of chimeric contributions to E6.5 embryos. Red, mKO2; blue, Eomes immunostaining. Scale bar, 100μm. (B) Dnmt3dKO∆A2 ES cell chimaeras at E7.5. 25% opacity bright field image overlaid with fluorescent image. Scale bar, 250μm. (C) Dnmt3dKO∆A2 ES cell chimaeras at E9.5. Scale bar, 1mm. (D) FS cells established from parental and Dnmt3dKO∆A2 ES cells. Scale bar, 100μm. (E) RT-qPCR analysis of FS cell markers. Error bar represents S.D. from technical duplicates. (F) Percentage of GFP-labelled WT FS cells and mKO2-labelled dKO-A2 FS cells in the mixed co-culture. p0 is two days after plating. Error bar represents S.D from two cultures. (G) Annexin V positive cells quantified by flow cytometry after 24h in indicated co-culture. Error bars represents SD from 6 experiments. $P<0.05$. (H) Immunofluorescence images of cMyc parental FS cells (GFP), Dnmt3dKO∆A2 FS cells (mKO2) and co-culture for 24h. Dashed regions highlight mKO2 positive cells in co-culture. (I) RT-qPCR analysis of marker expression during neural and mesendoderm differentiation of BCL2 transfected ES cells. Error bar represents S.D. from technical duplicates. (J) BCL2-Dnmt3dKO∆A2 ES cell chimaeras at E9.5. 20% opacity bright field image overlaid on fluorescent image. Scale bar, 1mm. (K) BCL2-Dnmt3dKO∆A2 ES cell chimeras at E12.5. Scale bar, 1 mm.

REFERENCES

Amoyel M, Bach EA. 2014. Cell competition: how to eliminate your neighbours. *Development* **141**: 988-1000.

Auclair G, Guibert S, Bender A, Weber M. 2014. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* **15**: 545.

Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A et al. 2012. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* **8**: e1002629.

Bogutz AB, Oh-McGinnis R, Jacob KJ, Ho-Lau R, Gu T, Gertsenstein M, Nagy A, Lefebvre L. 2018. Transcription factor ASCL2 is required for development of the glycogen trophoblast cell lineage. *PLoS Genet* **14**: e1007587.

Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M. 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42**: 1093-1100.

Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, Wysocka J. 2014. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**: 838-853.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**: 411-420.

Claveria C, Giovinazzo G, Sierra R, Torres M. 2013. Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* **500**: 39-44.

de la Cova C, Abril M, Bellosta P, Gallant P, Johnston LA. 2004. Drosophila myc regulates organ size by inducing cell competition. *Cell* **117**: 107-116.

Dunn SJ, Li MA, Carbognin E, Smith A, Martello G. 2019. A common molecular logic determines embryonic stem cell self-renewal and reprogramming. *Embo J* **38**.

Edwards JR, Yarychkivska O, Boulard M, Bestor TH. 2017. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **10**: 23.

Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J et al. 2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**: 351-359.

Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HH, Matarese F, Lepikhov K, Gut M, Brun-Heath I, Hubner NC et al. 2013. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**: 360-369.

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

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Kalkan T, Olova N, Roode M, Mulas C, Lee HJ, Nett I, Marks H, Walker R, Stunnenberg HG, Lilley KS et al. 2017. Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144**: 1221-1234.

Kinoshita M, Barber M, Mansfield W, Cui Y, Spindlow D, Stirparo GG, Dietmann S, Nichols J, Smith A. 2020. Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell*.

Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, Mansfield W, Nashun B, Knezovich JG, Smith A et al. 2013. Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* **20**: 311-316.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915-926.
Li MA, Amaral PP, Cheung P, Bergmann JH, Kinoshita M, Kalkan T, Ralser M, Robson S, von Meyenn F, Paramor M et al. 2017. A lncRNA fine tunes the dynamics of a cell state transition involving Lin28, let-7 and de novo DNA methylation. *Elife* **6**.
McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**: 4288-4297.
Monk M, Boubelik M, Lehnert S. 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**: 371-382.
Mulas C, Kalkan T, von Meyenn F, Leitch HG, Nichols J, Smith A. 2019. Defined conditions for propagation and manipulation of mouse embryonic stem cells. *Development* **146**.
Ohinata Y, Tsukiyama T. 2014. Establishment of trophoblast stem cells under defined culture conditions in mice. *PLoS One* **9**: e107308.
Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.
Posfai E, Schell JP, Janiszewski A, Rovic I, Murray A, Bradshaw B, Pardon T, El Bakkali M, Talon I, De Geest N et al. 2020. Defining totipotency using criteria of increasing stringency. *bioRxiv*: 2020.2003.2002.972893.
Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. 2017a. Single-cell mRNA quantification and differential analysis with Census. *Nature methods* **14**: 309-315.
Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017b. Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* **14**: 979-982.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
Sancho M, Di-Gregorio A, George N, Pozzi S, Sanchez JM, Pernaute B, Rodriguez TA. 2013. Competitive interactions eliminate unfit embryonic stem cells at the onset of differentiation. *Dev Cell* **26**: 19-30.
Shirane K, Kurimoto K, Yabuta Y, Yamaji M, Satoh J, Ito S, Watanabe A, Hayashi K, Saitou M, Sasaki H. 2016. Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Dev Cell* **39**: 87-103.
Smallwood SA, Tomizawa S-i, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**: 811-814.
Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**: 339-344.
Smith ZD, Shi J, Gu H, Donaghey J, Clement K, Cacchiarelli D, Gnirke A, Michor F, Meissner A. 2017. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**: 543-547.
Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**: 1521.
Stark RB, G. 2011. DiffBind : differential binding analysis of ChIP-Seq peak data. *Bioconductor*.
Tanaka M, Puchyr M, Gertsenstein M, Harpal K, Jaenisch R, Rossant J, Nagy A. 1999. Parental origin-specific expression of Mash2 is established at the time of implantation with its imprinting mechanism highly resistant to genome-wide demethylation. *Mech Dev* **87**: 129-142.
Tanaka S, Kunath T, Hadjantonakis A-K, Nagy A, Rossant J. 1998. Promotion of trophoblast stem cell proliferation by FGF-4. *Science* **282**: 2072-2075.
Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**: 381-386.

Yagi M, Kabata M, Tanaka A, Ukai T, Ohta S, Nakabayashi K, Shimizu M, Hata K, Meissner A, Yamamoto T et al. 2020. Identification of distinct loci for de novo DNA methylation by DNMT3A and DNMT3B during mammalian development. *Nature communications* **11**: 3199.

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**: 224-227.

Yamane T, Dylla SJ, Muijtjens M, Weissman IL. 2005. Enforced Bcl-2 expression overrides serum and feeder cell requirements for mouse embryonic stem cell self-renewal. *Proc Natl Acad Sci U S A* **102**: 3312-3317.

Zhang W, Xu J. 2017. DNA methyltransferases and their roles in tumorigenesis. *Biomark Res* **5**: 1.