Tracking the embryonic stem cell transition from ground state pluripotency

Tüzer Kalkan[1], Nelly Olova[2], Mila Roode[1], Carla Mulas[1], Heather J. Lee[2,7], Isabelle Nett[1], Hendrik Marks[3], Rachael Walker[1,2], Hendrik G. Stunnenberg[3], Kathryn S. Lilley[5,6], Jennifer Nichols[1,4], Wolf Reik[2,7,8], Paul Bertone[1] and Austin Smith[1,5]

1. Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, Cambridge CB2 1QR, UK
2. Babraham Institute, Cambridge CB22 3AT, UK
3. Radboud University, Faculty of Science, Department of Molecular Biology, Radboud Institute for Molecular Life Sciences (RIMLS), 6500HB, Nijmegen, The Netherlands
4. Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 4BG
5. Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA
6. The Cambridge Centre for Proteomics, Cambridge System Biology Centre, University of Cambridge, Cambridge, CB2 1QR, UK
7. Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK
8. Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK

Highlights

- The Rex1 destabilized GFP reporter demarcates naive pluripotency.
- Exit from the ES cell ground state is asynchronous across the population.
- Exit is relatively acute in individual cells and precedes lineage priming.
- Transcriptome and DNA methylome changes resemble peri-implantation epiblast.

Author Contributions

Acknowledgements

Development • Advance article

Summary

Mouse embryonic stem (ES) cells are locked into self-renewal by shielding from inductive cues. Release from this ground state in minimal conditions offers a system for delineating developmental progression from naive pluripotency. Here we examined the initial transition process. The ES cell population behaves asynchronously. We therefore exploited a short-half-life *Rex1::GFP* reporter to isolate cells either side of exit from naive status. Extinction of ES cell identity in single cells is acute. It occurs only after near-complete elimination of naïve pluripotency factors, but precedes appearance of lineage specification markers. Cells newly departed from the ES cell state display features of early post-implantation epiblast and are distinct from primed epiblast. They also exhibit a genome-wide increase in DNA methylation, intermediate between early and late epiblast. These findings are consistent with the proposition that naive cells transition to a distinct formative phase of pluripotency preparatory to lineage priming.

Introduction

Epiblast cells, founders of all somatic cells and the germ line, are formed in the inner cell mass (ICM) in the final day of pre-implantation development in mice (Boroviak and Nichols, 2014, Cockburn and Rossant, 2010). This emergent condition of "naive pluripotency" (Nichols and Smith, 2009) is characterized by a unique suite of transcription factors, a hypomethylated genome, and the ability to give rise directly and clonally to embryonic stem (ES) cells (Boroviak et al., 2014, Brook and Gardner, 1997, Smith et al., 2012, Lee et al., 2014, Nichols and Smith, 2012). Upon implantation, ES cell forming capacity is abruptly lost, epithelialisation commences, global gene expression is reconfigured, and DNA methylation increases, indicative of a profound cellular transition (Boroviak et al., 2014, Boroviak et al., 2015, Auclair et al., 2014, Bedzhov and Zernicka-Goetz, 2014). Subsequently egg cylinder epiblast cells are subject to inductive cues leading up to gastrulation and become fated, though not yet lineage-committed (Tam and Zhou, 1996, Osorno et al., 2012, Solter et al., 1970). The late phase of pluripotency during gastrulation is termed primed, reflecting the incipient expression of lineage specification factors (Nichols and Smith, 2009, Hackett and Surani, 2014).

Mouse ES cells cultured in serum-free media supplemented with two chemical inhibitors (2i) of MEK1/2 and GSK3α/β, and optional addition of the cytokine LIF are in a uniform condition of self-renewal termed the "ground state" (Ying et al., 2008). In this in vitro ground state, ES cells show transcriptional and epigenetic similarity to naive pre-implantation epiblast (Ficz et al., 2013, Habibi et al., 2013, Leitch et al., 2013, Nichols and Smith, 2012, Boroviak et al., 2015). Upon withdrawal of 2i, ES cells embark on a path to lineage commitment either in vitro or in vivo when injected into a pre-implantation embryo (Ying et al., 2008, Dunn et al., 2014, Marks et al., 2012). Recent studies have begun to explore the dissolution of naive pluripotency and the route towards multi-lineage differentiation *in vitro* (Buecker et al., 2014, Leeb et al., 2014, Kurimoto et al., 2015, Thomson et al., 2011, Respuela et al., 2016, Yang et al., 2014, Betschinger et al., 2013, Davies et al., 2013, Liu et al., 2015, Acampora et al., 2013). However, differentiating cultures become heterogeneous (Marks et al., 2012, Kalkan and Smith, 2014, Buecker et al., 2014, Hayashi et al., 2011). A means to identify and select cells as they transition from naive pluripotency would facilitate experimental resolution.

We previously generated ES cells carrying a *Rex1::GFPd2* (RGd2) reporter in which the coding sequence of one allele of Rex1 (gene name *Zfp42*) is replaced with a GFPd2-IRES-*bsd* cassette that produces a destabilized version of GFP protein with a 2-hour-half-life (GFPd2) (Wray et al., 2011). Here we exploit this reporter to monitor ES cell exit from naive pluripotency guided by autocrine cues in defined adherent culture. We test the utility of the reporter as a faithful marker of naive pluripotency and survey transcriptomic, metabolic and DNA methylome changes during the initial transition towards differentiation competence.

Results

The RGd2 reporter is a neutral marker of naïve pluripotency in the embryo

The Rex1 coding sequence is entirely deleted in the *RGd2* allele. RGd2 ES cells (Wray et al., 2011) were transmitted through the mouse germline and heterozygous animals were backcrossed twice to strain 129. Following heterozygous intercrosses, homozygous mice were healthy and fertile, although slightly under-represented (Table S1). These results confirm previous reports that Rex1 is dispensable for development (Kim et al., 2011, Masui et al., 2008, Rezende et al., 2011). We could derive wildtype, heterozygous and homozygous ES cells, both male and female, from intercross embryos (Table S2), demonstrating that Rex1 is not significant for ES cell propagation. RGd2 expression should therefore constitute a neutral reporter.

We evaluated reporter expression in the embryo by immunofluorescence staining for GFP. Co-staining for GATA4 revealed that the RGd2 reporter is expressed exclusively and uniformly throughout the naive epiblast (Epi) at E4.5 (Fig 1A), with no GFP in either GATA4-positive primitive endoderm or trophoblast. GFP is downregulated during implantation and becomes undetectable in the epiblast at E5. Expression is up-regulated in the extra-embryonic ectoderm (ExE), however (Fig 1B). These results are consistent with Rex1 mRNA expression in the embryo measured by in situ RNA hybridisation (Pelton et al., 2002), RT-qPCR (Boroviak et al., 2014) and RNA-seq (Boroviak et al., 2015). We conclude that the *RGd2* allele faithfully reports endogenous *Rex1* transcription and accordingly that GFP expression coincides with naive pluripotency *in vivo* (Boroviak et al., 2014) .

Release of ES cells from 2i triggers progression towards multi-lineage specification

We monitored the early phase of ES cell transition *in vitro* after withdrawal from 2i in serum-free N2B27 medium on gelatin-coated plastic (Fig 2A). We started from ES cells in 2i alone because LIF delays the onset of differentiation (Dunn et al., 2014). Plating ES cells directly in N2B27 at low density (<10000 cells cm^{-2}) results primarily in neural differentiation (Ying et al., 2003). However, when cells were plated at an intermediate density (15000 cells cm^{-2}) and maintained in 2i for 24 h prior to withdrawal, numerous Brachyury (*T*) positive cells also appeared, indicative of non-neural specification (Fig 2B). The latter conditions were used throughout this study.

Oct4 protein expression did not change substantially for 48h after release from 2i (Fig 2B). Rare cells expressing low levels of Sox1 were first detected at 48h (Fig 2B; panels k, q). By 72h clusters of bright Sox1-positive cells that lacked Oct4 emerged (Fig 2B, panels o, r-t). Occasional Sox1/Oct4-double-positive cells were found outside these clusters (Fig 2B; arrowheads in r-t). T-expressing cells were first detected at 72h, mostly in dense clumps that were mutually exclusive with Sox1-positive clusters (Fig 2B, m-p). T-positive cells at this stage were also positive for Oct4 (Fig 2B; arrowheads in u-w), consistent with transient co-expression in early

primitive streak and during directed *in vitro* differentiation (Thomson et al., 2011, Hoffman et al., 2013).

Oct4 mRNA downregulation started after 25h (Fig 2C). In contrast, transcripts for naive transcription factors (TFs) (Dunn et al., 2014, Martello and Smith, 2014) declined within the first 25 h. Downregulation of Nanog, Esrrb and Tfcp2l1 initiated before Rex1 and Klf2. Concurrently, transcripts for early post-implantation epiblast markers Fgf5, Otx2, Oct6 (*Pou3f1*) (Acampora et al., 2016, Pelton et al., 2002), were upregulated (Fig 2D). mRNAs for naive TFs were eliminated by 48h (Fig 2C). Similar results were observed with multiple ES cell lines (Fig S1A, B).

These results indicate that upon release from self-renewal in defined conditions ES cells are driven by autocrine signals to progress from the naive state to multi-lineage specification in an orderly sequence. First naive TFs are extinguished and markers diagnostic of post-implantation epiblast are induced. Subsequently lineage specific markers emerge and Oct4 is downregulated.

Pluripotency factors display individual downregulation kinetics upon 2i withdrawal

To follow the kinetics of transition following release from 2i, we monitored the RGd2 reporter by flow cytometry (Fig 3A). GFP was expressed unimodally with a log normal distribution in 2i. This tight peak persisted throughout the first 16h after withdrawal, although mean fluorescence rose slightly, possibly due to a transient increase in Rex1 mRNA at earlier time points (Figs 2C, S1A). By 25h GFP intensity became heterogeneous with many cells shifted to lower expression. This profile suggests that Rex1 is downregulated with different kinetics in individual cells. By 48h the majority of cells extinguished GFP. Treatment with the protein synthesis inhibitor cycloheximide confirmed that the half-life of GFPd2 is slightly under 2 hours in both 2i- and N2B27-cultured ES cells (Fig S2A, B, C). Therefore, observed changes in GFP levels upon 2i withdrawal should track Rex1 transcription.

We compared the expression of the RGd2 reporter with the naïve TF Nanog and with Otx2, a TF that is upregulated in the peri-implantation epiblast (Acampora et al., 2013, Acampora et al., 2016) (Fig 3B). Quantification of fluorescence intensities for Nanog and Otx2 in single cells across the 25h time course showed that in 2i almost all cells expressed Nanog at high or intermediate (mid) levels (Fig 3C, D). Otx2 was expressed at low to intermediate (mid) levels in many cells but absent in 23%. 16h post-2i withdrawal GFP remained ubiquitous, Nanog became undetectable in most cells (72%), and Otx2 was upregulated (Fig 3B, C, D). Nanog and Otx2 were co-expressed at mid or high levels in only 15% of cells. By 25h GFP intensity was heterogeneous, consistent with the flow cytometry profile. Otx2 was expressed in almost all cells, although the proportion of Otx2-high cells was lower than at 16h, consistent with a relative decrease in Otx2 transcript levels (Fig 2D). Nanog persisted in only 7% of cells (Fig 3D) most of which were in the Rex1-high fraction

(Fig 3E, F). The overall pattern of Nanog and Otx2 expression during ES cell progression mirrors dynamics in the embryo. Nanog is co-expressed with low levels of Otx2 in the naïve pre-implantation epiblast, but as Nanog is down-regulated in the peri-implantation epiblast, Otx2 is up-regulated in mutually exclusive fashion (Acampora et al., 2016).

A second naive pluripotency factor, Tfcp2l1 (Martello et al., 2013, Ye et al., 2013), was already undetectable in most of the population by 16h (Fig S2D), concomitant with rapid decrease in transcripts (Fig 2C). We also examined Tfe3, a bHLH transcription factor that supports naïve pluripotency (Betschinger et al., 2013). Nuclear level of Tfe3 was also already reduced by 16h relative to 2i (Fig S2E, F, G), Tfe3 becoming mostly cytoplasmic. In contrast, Oct4 and Sox2 proteins exhibited homogeneous expression throughout the first 25 h after 2i withdrawal (Fig 3G).

These results reveal that TFs associated with pluripotency display individual expression behaviour as ES cells transition from the ground state. RGd2 downregulation appears to reflect aggregate loss of naive TFs against a background of persistent Oct4 and Sox2 expression.

Exit from the ground state occurs asynchronously
To determine the time of exit from the ground state, entire cultures or subpopulations sorted on the basis of RGd2 expression at selected time points were re-plated at single cell density in serum/L and 2i/L. Resulting colonies were stained for alkaline phosphatase (AP) activity (Fig 4A).

Serum/L permits proliferation of both naive ES cells and differentiating progeny (Marks et al., 2012). Thus, colonies in serum/L reflect plating efficiency and differentiation propensity. 2i-cultured ES cells and cells from inhibitor-withdrawn cultures generated similar numbers of colonies, indicating equivalent plating capacity (Fig 4B). However, the proportions of colony types varied. From 2i-cells, around 60% of colonies were wholly undifferentiated, with most of the remainder were mixed and a few completely differentiated. This heterogeneity is typical of ES cells plated in serum/L (Wray et al., 2010). The degree of differentiation increased with duration of 2i withdrawal; only 10% of colonies formed after 16h were undifferentiated while over 20% were wholly differentiated. From 48h-cultures 95% of colonies were differentiated (Wray et al., 2010). Thus ES cells become increasingly predisposed to differentiation as the 2i withdrawal period is prolonged.

In 2i/L self-renewal is optimal, but differentiating cells are eliminated and only naive cells form colonies. Strikingly, the clonogenic efficiency of 16h-cells in 2i/L was equivalent to that of ground state ES cells (Fig 4B). Thus, the increased propensity for differentiation detected in serum/L is not matched by loss of self-renewal potential. However, after 25h of 2i withdrawal, clonogenicity in 2i/L was significantly reduced, and by 48h had fallen to 10% of the starting level. Therefore up to 16h after

2i withdrawal, self-renewal potential remains intact, despite the reduction in expression of some naive TFs, induction of post-implantation epiblast markers, and increased tendency to differentiate in serum/L. Between 16h and 25h self-renewal capacity is partially lost whereas by 48h exit from the naive state is almost complete across the culture. Thus exit from the naïve state proceeds gradually in the ES cell population over an extended period (≤48 h).

Downregulation of Rex1 tracks loss of ES cell self-renewal potential

To determine whether the gradual loss of ES cell identity at the population level is recapitulated at the single cell level we exploited flow cytometry to fractionate cells based on RGd2 expression. We sorted 4 subpopulations 25h post 2i withdrawal, and then replated (Fig 4C, D). In serum/L, colony numbers were relatively constant, although the proportion of undifferentiated colonies declined with decreasing GFP. In 2i/L marked differences in total colony numbers were evident (Fig 4D). The GFP-high fraction exhibited equivalent clonogenicity to 2i cells (Fig 4B, D), indicating complete retention of naive status. However, subpopulations with lower GFP levels produced progressively fewer colonies. The number of colonies formed from the GFP-low fraction was only 15% of the number from 2i or GFP-high cells. Thus the great majority of this subpopulation has departed the ES cell state (Fig 4D). These data demonstrate that by 25h the population has become functionally heterogeneous. Therefore, ES cells transition asynchronously.

To examine how closely exit from the naive state and downregulation of Rex1 coincide, we sorted cultures 20h post 2i withdrawal, when the first GFP-low cells appear, into GFP-high and GFP-low subpopulations using the same gates as for 25h-cultures. (Fig S3A). Clonogenic efficiency of GFP-high cells in 2i/L was equivalent to ground state ES cells, but was reduced was 4-fold lower for GFP-low cells (Figs 4B, S3A). Thus the earliest cells that we could obtain after Rex1 downregulation have mostly exited the ES cell state. These data suggest that in individual cells the transition occurs at, or slightly after, loss of Rex1 expression.

We examined whether cell cycle dictates the kinetics of Rex1 downregulation. We stained ES cells with the DNA-binding dye Hoechst and isolated subpopulations in G1, S and G2/M by flow cytometry (Fig S3B). We plated these cells along with stained but unsorted controls directly in N2B27 at 3×10^4 cells cm-^{2}, which approximates the density at the time of 2i withdrawal in our standard protocol. All populations displayed a similar heterogeneous GFP distribution 25h after plating, although the G1 starting subpopulation showed a marginally narrower range and a slightly lower mean intensity (Fig S3C). We conclude that the kinetics of Rex1 downregulation is largely independent of initial cell cycle phase.

For subsequent analyses we selected and defined cell populations as follows: 2i-cells represent the ground state; 16h and 25h-H cells are in a reversible phase preceding the extinction of ES cell character; and 25h-L cells are the primary

products of exit from naïve pluripotency (Fig 4E). Flow cytometry and colony assays confirmed the reproducibility of this system (Figs 4F, S3D, S3E). Colonies formed from 16h and 25h-H populations after replating in 2i/L re-expressed naïve pluripotency markers at the same level as 2i cells, and down-regulated Otx2, Oct6 and Fgf5 (Fig S4A, B), demonstrating that the ground state was re-established. Immunoblotting after 2i withdrawal showed progressive downregulation of Nanog, Esrrb and Klf4 proteins and decreasing GFP, while Oct4 was constant (Fig 4G). The difference between 25h-H and 25h-L cells is of particular note; Nanog and Esrrb proteins are almost undetectable in 25h-L cells and Klf4 is greatly diminished. These three factors are pivotal members of the ES cell gene regulatory circuitry (Dunn et al., 2014) and their elimination together with the absence of Tfcp2l1 and nuclear Tfe3 (Fig S2) is expected to be sufficient for elimination of self-renewal in 25h-L cells.

To assess further the variation between 25h-cells we performed single cell RT-qPCR for selected genes (Fig 4H). This analysis confirmed that general pluripotency factors remained constant or showed modest changes, whereas naive TFs and post-implantation markers in general showed reciprocal expression. Notably, 2i cells were devoid of Fgf5, Oct6 (Pou3f1), Sox3 and Sox4 transcripts that are up-regulated in the post-implantation epiblast (Boroviak et al., 2015, Pelton et al., 2002, Acampora et al., 2016). The 25h-H cells showed variable upregulation of these 4 markers and downregulation of no more than 3 of the naive TFs. In contrast, in 25h-L cells all the post-implantation epiblast markers were co-expressed and at least 4 of the naive TFs were downregulated. These results suggest that decay of ES cell identity correlates with cumulative loss of naive TFs and concomitant cumulative upregulation of factors associated with early post-implantation epiblast. In the reversible 25h-H population, these factors are expressed in various combinations without an evident hierarchy.

Transcriptional changes during transition from naive pluripotency

To examine global expression dynamics we carried out microarray profiling on using 3 biological replicates. We also performed RNA-seq on independently derived RGd2 ES cell lines. We found a total of 8810 genes in the microarray that were differentially expressed between at least two subpopulations (Table S3). Consistent with the activation of MEK/ERK and GSK3 upon 2i withdrawal, we observed changes in the expression of components of the pathway and transcriptional targets. Activation of MEK/ERK is reflected in the upregulation of immediate ERK response genes, such as *Egr1, Fos, myc, c-Jun* (Murphy et al., 2004) and of negative-feedback regulators *Spry2* and ERK phosphatases *Dusp 4* and *6* (Fig 5A). mRNAs for canonical Wnt targets, *T, axin2, cdx1* and *cdx2* are detected at low levels in 2i, consistent with inhibition of Gsk3 (Marucci et al., 2014) (Fig 5A). Expression is reduced upon 2i withdrawal, indicating reduction of β-catenin dependent transcription during transition from the ground state. Lef1 is upregulated, however, suggesting increased potential for Wnt-stimulated transcriptional regulation after exit.

KEGG pathway enrichment analysis revealed that highly upregulated genes in 16h-cells were associated with cell cycle, cytoskeleton, steroid synthesis and cell adhesion; and in 25h-L cells with ribosome biogenesis, RNA processing, DNA replication, nucleotide metabolism, proteosome, VEGF and MAPK signalling. Most downregulated in 16h- and 25h-low cells were genes with functions in lysosomes, oxidative phosphorylation (OxPhos), glycolysis, glycosylation and glycan degradation. (Fig 5B). An overall decrease in transcripts encoding components of mitochondrial respiratory complexes was confirmed by RNA-seq (Fig S5A). The changes encompassed all five mitochondrial enzyme complexes that mediate electron transport and ATP synthesis. To investigate metabolic consequences, we measured oxygen consumption rate (OCR) and extracellular acidification rate (ECAR), indicators of mitochondrial respiration and glycolysis, respectively. Naïve ES cells exhibited higher basal and maximal OCR and also ECAR than the 25h populations, indicating higher levels of both mitochondrial respiration and glycolysis (Fig 5C). Within the 25h population, 25h-H cells exhibited higher OCR and ECAR than 25h-L cells (Fig 5C), indicating that the switch in metabolism is not a direct response to inhibitor withdrawal but associated with a developmental transition. A reduction in mitochondrial respiration between naive and primed pluripotent stem cells has been reported in mouse and human (Takashima et al., 2014, Zhou et al., 2012, Guo et al., 2016), and also been proposed to occur in utero (Zhou et al., 2012). Our analyses indicate that metabolic resetting begins during transition from naïve pluripotency and initially involves reduction in both oxidative phosphorylation and glycolysis.

To benchmark developmental progression, we curated panels of markers for the following categories: general pluripotency; naive pluripotency; post-implantation epiblast; lineage specification. We then examined expression in our microarray (Fig 5D) and RNA-seq (Figs S5B, D) datasets. Most naive pluripotency TFs were downregulated in reversible cells, and almost absent in 25h-L cells, in accordance with decay of ES cell identity. Exceptions were Rex1, Klf5, Fbxo15 and Nr0b1 which were maintained in reversible cells and reduced but not eliminated in 25h-L cells (Figs 5D, S5B). None of this latter group of factors are components of the core gene regulatory circuitry of naïve pluripotency (Dunn et al., 2014). Among the general pluripotency markers, Oct4 and Sall4 remained constant while Sox2 was mildly downregulated in 25h-L cells, similar to its downregulation in the epiblast from E4.5 to E5.5 (Boroviak et al., 2015). Other pluripotency factors including Lin28a, Zfp281, Zic2, Utf1 exhibited upregulation (Figs 5D, S5B). Consistent with increased expression, Lin28a, Zfp281 and Zic2 are reported to drive transition from naïve to primed pluripotency (Zhang et al., 2016, Fidalgo et al., 2016, Luo et al., 2015, Betschinger et al., 2013).

To assess concordance with protein levels we performed mass spectrometric analysis via stable isotope labelling of amino acids in culture (SILAC). These data confirmed that relative nuclear protein levels of TFs associated with naive and

general pluripotency correlated with respective transcripts, except for Rex1 and Klf5 whose protein levels were not reduced in 25h-L cells, despite decreasing transcript levels (Fig S5C).

Factors that are upregulated in the post-implantation epiblast (Boroviak et al., 2015) (Lef1, Pou3f1(Oct6), Dnmt3a/b, Foxd3, Sox3, Fgf5, Cdh2, Otx2) were progressively induced upon 2i withdrawal (Figs 5D, S5B). A large panel of factors associated with commitment to germ line, neuroectoderm, endoderm or mesoderm lineages remained near background levels (RPKM< 10) and showed no up-regulation beyond levels expressed in naïve ES cells (Fig 5D, S5D). Of note, Gata4 and Gata6 were not induced in 25h-L cells, excluding primitive endoderm specification as an alternative path. These results establish that ES cell exit from naive pluripotency is temporally segregated from upregulation of lineage determination factors.

Comparison of ES cell progression with in vivo epiblast, EpiLCs and EpiSCs

We compared the RNA-seq data from our in vitro populations (Table S4) to data from embryo samples acquired by a small sample RNA-seq protocol (Boroviak et al., 2015). We isolated expression changes that occur between pre- and post-implantation epiblast (E4.5 and E5.5), and asked to what extent these are reflected in the *in vitro* transition. Out of 608 genes that are differentially expressed between E4.5 and E5.5 epiblast and robustly detected in one or both of the samples (FPKM≥10), more than half (366 of 608) exhibited differential expression during ES cell transition, with the direction of change conserved (Fig 6, Table S5). Several functional groups could be identified within the shared up- and down-regulated gene sets (Fig 6A). Besides transcription factors and epigenetic regulators with established functions in the epiblast, the common group included genes associated with extracellular matrix (ECM), cell adhesion, motility, shape, metabolism and autophagy. Transcripts for ECM components Fibronectin (Fn1), Laminin isoforms, Lamc1 and Lama1, Laminin-linker protein Nid1, Spp1 (Osteopontin), Sparc, Lgals3, Alpl were downregulated, while Col18a was upregulated (Fig 6B), indicating major reconfiguration of ECM production. Tissue non-specific alkaline phosphatase (Alpl), widely used as a surrogate marker for ES cells, modifies the ECM by dephosphorylating ECM molecules such as Osteopontin (Spp1) (Narisawa, 2015).

LIF receptor components, Lifr and Ilst6, signal transducers Jak3 and Stat3, and the transcriptional targets Ccnd3, Klf4 and Tfcp2l1 were downregulated in vitro and in vivo, indicating that diminished LIF signalling is a common feature in transition from naive pluripotency (Fig 6B). BMP4 and Nodal inhibitors, Lefty1 and Lefty2 were also downregulated together with altered expression of positive and negative regulators of Wnt signalling. These events highlight a changing signalling context.

We noted changes in enzymes that regulate metabolism and autophagy. Threonine dehydrogenase (Tdh) was downregulated. This is a vital enzyme for mouse ES cell survival that converts threonine into acetyl co-A and glycine, feeding the TCA cycle

and purine synthesis, (Wang et al., 2009). *Pim2*, one of the most highly upregulated genes in both settings (Fig 6C), is a kinase that promotes glycolysis (Yu et al., 2013) and mTOR signalling (Lu et al., 2013, Zhang et al., 2015). mTOR pathway inhibitor Ddit4 (Redd1), and mTOR activator Rragd were both upregulated. mTOR signalling is activated during ES cell differentiation (Betschinger et al., 2013), and it has been reported that mTOR inhibition induces a diapause-like state of arrested development in the mouse blastocyst (Bulut-Karslioglu et al., 2016). Pim2, Ddit4 and Rragd are candidates that might contribute to complex mTOR regulation during this transitional period. Activated mTOR suppresses autophagy by phosphorylating and inhibiting Ulk1 and Atg13, two factors that are required for autophagy initiation, and through phosphorylation-dependent cytoplasmic sequestration of Tfeb, a TF that orchestrates expression of genes involved in lysosome function and autophagy (Kim and Guan, 2015, Napolitano and Ballabio, 2016). These three mTOR targets, Ulk1, Atg13 and Tfeb, along with several autophagosome- associated factors were downregulated (Fig 6B), suggesting transcriptional and post-transcriptional suppression of autophagy as the cells transition from naive pluripotency in vitro and in vivo.

We found that 190 of the 608 genes that show differential expression during epiblast transition do not change expression between 2i ES cells and 25h-L cells, while 52 showed differential expression in the opposite direction to the embryo (Table S5). This latter subset included ERK target Egr1 and factors that regulate cell proliferation (Atf3, Tef, Trp53, Tada3, Klf6, Ccng1), apoptosis (Apaf1, Bid), cell adhesion and morphology (Krt18, Cdh2, Fez1, Lamb1, Tns3, Amotl1) as well as signalling pathway components, such as Notch3, Rbpj and Nodal co-receptor Tdgf1 (Cripto). Contrasting expression of these genes might reflect differences in the developmental snapshots sampled in vitro and in vivo, and/or the absence of paracrine signalling cues in our minimal culture system. Nonetheless, overall the transcriptome analyses support the idea that loss of Rex1 expression in a simple and defined ES cell culture system mimics several features of the developmental transition from pre- to post-implantation pluripotency.

We also compared 25h-L cells to post-implantation epiblast-like cells (EpiLC), a transient intermediate generated during *in vitro* germ cell differentiation by plating ES cells from 2i/L into Fgf2, Activin and 1% KSR for 48h (Hayashi et al., 2011, Buecker et al., 2014). Differential gene expression analysis showed that EpiLCs were overall similar to 25h-L cells, although a number of genes distinguish them (Table S6). We also generated EpiLC from RGd2 ES cells and measured reporter expression by flow cytometry. We found that a subpopulation of EpiLCs (23%) express Rex1 at naive ES cell levels (Figs S6A, B), indicating that EpiLC populations are heterogeneous and contain a substantial fraction of undifferentiated ES cells.

We additionally undertook a comparison with published data on EpiSC, which are related to gastrula stage epiblast (Kojima et al., 2014). Marker expression (Fig S6C, D) shows that 25h-L cells are related to E5.5 epiblast and are distinct from EpiSC. These data confirm that ES cells do not transition directly into EpiSC (Hayashi et al., 2011, Smith, in press).

Acquisition of DNA methylation during transition from naïve pluripotency

Genome-wide DNA methylation increases substantially between E4.5 and E5.5 in utero (Auclair et al., 2014). Expression of *de novo* DNA methyltransferases Dnmt3a and Dnmt3b is markedly upregulated both in ES cells and in the epiblast during transition (Figs 7A, 5D, S5B). In contrast, Prdm14, which represses Dnmt3a/b and promotes Tet activity on target genes (Yamaji et al., 2013, Okashita et al., 2014, Ficz et al., 2013), is downregulated. Whole genome bisulfite sequencing (WGBS) revealed an increase in total CG methylation across the genome upon 2i withdrawal (Fig 7B). Average genome methylation tended to increase in small increments between 2i, 16h and 25h-H, with a more pronounced and statistically significant increase at 25h-L. The increase was similar across gene bodies, exons, introns, intergenic regions, satellites and retrotransposon sequences (LINEs, SINEs, LTRs, IAPs), while methylation of CpG islands (CGIs) was not generally altered (Fig S5A, B). Promoters that contain CGIs remained refractory to DNA methylation upon 2i withdrawal, while non-CGI promoters exhibited increased methylation similar to the genome average (Fig 7C).

Whole genome methylation data are not available for E5.5 epiblast. We therefore compared our profiles with published data on E6.5 post-implantation epiblast (Seisenberger et al., 2012). This analysis showed that the CG methylation level of 25h-L cells is intermediate between naïve ES cells and E6.5 post-implantation epiblast across all genomic regions (Figs 7, S7). This is in line with reduced representation bisulfite sequencing (RRBS) data on embryos showing a marked increase in methylation between E4.5 and E5.5 with a further increase at E6.5 (Auclair et al., 2014).

To investigate how DNA methylation might relate to gene expression changes during exit from the naïve state we examined enhancers. We identified "naïve enhancers" from published ChIP-seq datasets (Buecker et al., 2014) as regions displaying the general enhancer mark H3K4me1, together with the active enhancer marks H3K27Ac and p300 in naïve ES cells (Visel et al., 2009, Rada-Iglesias et al., 2011, Heintzman et al., 2007, Visel et al., 2013, Blow et al., 2010). In 2i these enhancers were lowly methylated, but they gained methylation progressively upon 2i withdrawal (Fig 7D). In contrast ES cell "super enhancers" (SEs) defined in serum-cultured ES cells (Whyte et al., 2013, Hnisz et al., 2013) exhibited low and relatively constant methylation levels in 2i, 16h and 25h-H cells, with a small increase in 25h-L cells but below the genome average (Fig 7D). These observations indicate that on exit from

naïve pluripotency naïve enhancers are methylated, indicative of decommissioning, while SEs that may be linked to general pluripotency-associated transcription are protected from methylation.

We compared our WGBS data to published RRBS data from E4.5 and E5.5 epiblasts (Auclair et al., 2014), and Rex1-sorted subpopulations of serum-cultured RGd2 ES cells (Singer et al., 2014) (Table M). This analysis is limited to promoters and CGIs that are covered in all datasets since RRBS enriches for genomic regions with high CpG content (Meissner et al., 2005). The vast majority of CGIs and CG-rich promoters in epiblast samples exhibited less than 5% methylation (Fig S8A, Table S9). This was matched in 2i and 25h-L cells. In contrast both Rex1-high and Rex1-low fractions of serum-cultured ES cells exhibited higher levels of methylation. Therefore, we did not include serum ES cell samples in further analyses. Most promoters gained less than 5% methylation during ES cell and epiblast transitions (Fig S8B). We asked if those that do gain methylation are conserved between in vivo and in vitro settings. We isolated 2000 promoters with the highest methylation increases (Table S10). Interestingly, these promoters had higher methylation relative to all other promoters in naïve ES cells and epiblast (Fig S8C). The majority were associated with lowly-expressed or non-expressed genes (Table S10). In the common group of 1288 promoters were only 52 associated genes that are expressed in both in ES cell populations and the epiblast (Fig S8D). A subset of these genes exhibited downregulation both in ES cell and epiblast progression. Among these are *Tdh, Lefty1, Tcl1 and Prdm14*.

WGBS analysis also showed no genome-wide correlation between promoter methylation and gene expression changes. Nonetheless, we noted increased methylation in a subset of naïve pluripotency gene promoters including *Nanog*, *Nr0b1*, and *Dppa3,* in addition to *Prdm14* and *Tcl1* as noted above (Fig 7E, S8E). Promoters of other naïve and general pluripotency factors did not gain methylation, showing that pluripotency-associated genes acquire methylation with different kinetics. Thus, we conclude that promoter methylation is not a major driver of transcriptional changes during exit from the naïve state, although, it might contribute to repression of a subset of genes that are of potential functional significance for the transition.

Discussion

The RGd2 reporter provides near-real time detection of exit from naive pluripotency and enables isolation of the first cells to change functional identity. Loss of Rex1 expression marks a progression in pluripotent status that precedes a decline in Oct4 level or acquisition of lineage-specific gene expression. Our results characterize a defined *in vitro* monolayer differentiation system in the absence of uterine or extraembryonic cues. In these simple conditions autocrine signals are sufficient to drive transcriptomic, metabolic and methylome changes that are broadly convergent

with peri-implantation epiblast. These findings indicate that the gene regulatory circuitry of ES cells has an innate capacity to orchestrate a profound developmental transition.

At the onset of this transition, the molecular network that sustains naïve pluripotency is dismantled (Buecker et al., 2014, Kalkan and Smith, 2014, Leeb et al., 2014). Apparently co-incident with acute downregulation of the critical naive TFs, post-implantation epiblast markers are up-regulated (Acampora et al., 2016, Boroviak et al., 2015). Increased differentiation when transferred to serum suggests enhanced sensitivity to inductive cues before cells have fully extinguished ES cell identity. However, for as long as Rex1 is expressed, cells retain in full the ability to regain the ground state. Such reactivation of self-renewal, despite marked reduction in the levels of functionally important naïve TFs, is consistent with evidence that the mouse ES cell state is founded on a highly flexible transcription factor circuitry (Dunn et al., 2014, Martello and Smith, 2014, Young, 2011, Niwa, 2014). We surmise that loss of Rex1 reflects a cumulative reduction of the suite of factors below a critical threshold. From this point the naive TF network cannot be reactivated and is subsumed by an emerging new circuitry. The apparent gradual loss of self-renewal gleaned from whole population analyses arises from asynchronous single cell dynamics and at the level of individual cells the exit from ES cell identity may be precipitate.

Rex1 transcription is considered to be directly regulated by several naïve TFs (Chen et al., 2008, Kim et al., 2008) which can explain how the RGd2 reporter serves as a sensor of the overall activity of the naïve transcription factor circuitry. Nevertheless, 10-15% of the Rex1-low cells at 25h can be restored to ground state self-renewal. This may be explained in part by incomplete efficiency of flow sorting, but also suggests that Rex1 downregulation might be separated from exit in a minority of cells. The higher incidence of reversion for Rex1-Low cells at 20h is consistent with silencing of Rex1 slightly preceding loss of ES cell identity. Reversibility of Rex1 reporter expression has been reported in serum (Toyooka et al., 2008), where ES cells are continuously exposed to conflicting pro- and anti-differentiation stimuli that may perturb developmental progression. Even in those conditions, however, it is apparent that with more stringent categorisation of Rex1-negative cells, reversion frequency is low (Nakai-Futatsugi and Niwa, 2016). Furthermore, Rex1-negative cells from serum culture tend to be eliminated from blastocyst chimaeras, indicating limited identity with resident epiblast (Alexandrova et al., 2016, Toyooka et al., 2008). Nonetheless, the connection between down-regulation of naïve factors and loss of Rex1-GFP may not be invariant. Indeed rare Nanog-positive/GFP-negative cells are detected at 25h.

Consistent with loss of functional ES cell character, the 25h-L population show significant transcriptome variance from their naïve predecessors. They are clearly distinct from EpiSC, and converge towards EpiLC and E5.5 epiblast. It will be of

interest to determine to what extent micro-environmental modulations, such as substrate composition, may increase the veracity of the ES cell model.

The dynamic and global increase in DNA methylation as ES cells leave the naïve state generate a methylome intermediate between naïve and primed pluripotent compartments. We detected profound increases in the methylation levels of naïve ES cell enhancers and in a minority of promoters. However, there was no overall correlation between promoter methylation and gene expression, as also observed when ES cell cultures were switched between 2i and serum (Ficz et al., 2013, Habibi et al., 2013). Thus increased promoter methylation does not appear to be a major driver of initial progression from naïve pluripotency. Indeed most pluripotency gene promoters were spared from methylation, although *Nanog* and *Prdm14* were prominent exceptions that gained methylation in 25h-L cells.

We have proposed that downregulation of naïve pluripotency factors elicits differentiation competence via an intermediate phase of "formative pluripotency" (Kalkan and Smith, 2014, Smith, in press). This is postulated as a period of competence acquisition for multi-lineage specification. *In vivo* the formative phase corresponds to peri- and immediate post-implantation epiblast (E4.75-5.75), before cells exhibit expression of lineage specification factors. Notably during this period epiblast cells acquire competence for germ cell induction (Ohinata et al., 2009, Hayashi et al., 2011). Our results indicate that ES cells that downregulate Rex1 and depart naïve pluripotency show transcriptome and methylome features that may be anticipated for immediate post-implantation epiblast. Thus, these Rex1-low cells represent a snapshot of the nascent formative phase, undergoing rewiring of the gene regulatory network and remodelling of the epigenome. The datasets we provide constitute a resource for inspecting an overlooked phase of pluripotency. It will be of future interest to dissect in detail the molecular dynamics and drivers of transition in this defined and simple system and also to determine whether the formative phase may be suspended as a stem cell state in culture, as achieved for naïve ES cells and primed EpiSCs.

Materials and Methods

ES cell lines and culture

ES cell lines carrying the RGd2 reporter were derived from embryos using previously described protocols (Nichols et al., 2009) . For routine maintenance ES cells were plated at 2-3 x 10^4 cells cm^{-2} in 2i on 0.1% gelatine-coated dishes (Sigma, G1890) and passaged every 3 days following dissociation with Accutase (PAA, L11-007). 2i consists of N2B27 (NDiff N2B27 base medium, Stem Cell Sciences Ltd, Cat. SCS-SF- NB-02) or homemade N2B27, supplemented with PD0325901 (1 µM) and CHIR99021 (3 µM). LIF prepared in house was added only when indicated.

Immunofluorescence (IF) staining of ES cells and image quantification

Cells were fixed for 10 min with 4%PFA at RT, followed by permeabilization and blocking in PBS containing 0.1% TritonX-100 and 3% donkey serum. Cells were incubated with primary antibodies (Table 7) in blocking solution overnight at 4C. Alexa Fluor-conjugated donkey secondary antibodies (Molecular Probes) were used at 1:1000. Images were obtained using Leica SP5 confocal microscope for Fig S2 and Leica 4000B standard fluorescent microscope for the rest. Mean IF and DAPI intensity per cell was quantified with Volocity (Fig 3) and Cell Profiler (Fig S2). For Fig 3 cells were ordered according to increasing mean IF intensity in DAPI-positive particles and then consecutive 25, 50, 25 percentiles of positive cells were labelled as low, mid and high for a particular marker.

Monolayer differentiation, flow cytometry, cell sorting and clonogenicity assays

Cells were plated at 1.5 x 10^4 cells cm^{-2} in 2i and medium was replaced with N2B27 or fresh 2i after 20-24 h. Prior to sorting cells were washed, pelleted and resuspended in the respective culture medium. For dissociation of 2i-cells 2i inhibitors were added into Accutase. ToPro-3 (Invitrogen) was added at a concentration of 0.05 nM to label membrane compromised cells. Cells were sorted in MoFlo flow sorter (Beckman Coulter). From 2i -and 16h-cultures all ToPro-3-negative cells, and to obtain Rex1-high and –low fractions from 25h-cultures, cells with highest and lowest 15% GFP expression were collected. For clonogenicity assays, 500-800 cells were plated on 6-well dishes in L/S or 2i/L, coated with 0.1% gelatine or laminin (Sigma, L2020), respectively. On day 4 (L/S) or 6 (2i/L), plates were fixed and stained for AP (Sigma, 86R-1KT). Plates were scanned using a Cell Celector (Aviso) and scored manually. Colony formation efficiency for a given population was determined by dividing the average number of colonies formed in 2i/L by that in L/S. Flow cytometry was performed using a Dako Cytomation CyAn ADP high-performance cytometer and results were analyzed with Summit.

RNA extraction, cDNA synthesis and qPCR

Total RNA was isolated using RNeasy mini kit (Qiagen). cDNA was synthesized using SuperScript III (Invitrogen) and oligo-dT primers. qRT-PCR was performed with TaqMan Gene Expression assays (Thermo Scientific) using probes listed in Table S7.

Single cell RT-qPCR

Cells were sorted using a G1 enrichment strategy, based on forward scatter (FS) and side scatter (SC) gating. Single cells were sorted into 96-well plates containing CellsDirect One-Step qRT-PCR master mix (Invitrogen, 11753-100) for cDNA synthesis and pre-amplification. Fluidigm assays were performed according to the manufacturer's protocols using EvaGreen detection at the Genomics Core Facility at the European Molecular Biology Laboratory in Heidelberg, Germany. Primer sets used are listed in Table S7.

Microarray, RNA-sequencing, DNA methylome and proteome analyses

Processing of ES cell samples and data analyses are described in the Supplementary file. Affymetrix and RNA-seq data are in ArrayExpress repository under accessions E-MTAB-5304 and E-MTAB-5305.

References

ACAMPORA, D., DI GIOVANNANTONIO, L. G. & SIMEONE, A. 2013. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development,* 140**,** 43-55.

ACAMPORA, D., OMODEI, D., PETROSINO, G., GAROFALO, A., SAVARESE, M., NIGRO, V., DI GIOVANNANTONIO, L. G., MERCADANTE, V. & SIMEONE, A. 2016. Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast. *Cell Rep,* 15**,** 2651-64.

ALEXANDROVA, S., KALKAN, T., HUMPHREYS, P., RIDDELL, A., SCOGNAMIGLIO, R., TRUMPP, A. & NICHOLS, J. 2016. Selection and dynamics of embryonic stem cell integration into early mouse embryos. *Development,* 143**,** 24-34.

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.

BEDZHOV, I. & ZERNICKA-GOETZ, M. 2014. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell,* 156**,** 1032-44.

BETSCHINGER, J., NICHOLS, J., DIETMANN, S., CORRIN, P. D., PADDISON, P. J. & SMITH, A. 2013. Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell,* 153**,** 335-47.

BLOW, M. J., MCCULLEY, D. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., BRISTOW, J., REN, B., BLACK, B. L., RUBIN, E. M., VISEL, A. & PENNACCHIO, L. A. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet,* 42**,** 806-10.

BOROVIAK, T., LOOS, R., BERTONE, P., SMITH, A. & NICHOLS, J. 2014. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol,* 16**,** 516-28.

BOROVIAK, T., LOOS, R., LOMBARD, P., OKAHARA, J., BEHR, R., SASAKI, E., NICHOLS, J., SMITH, A. & BERTONE, P. 2015. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev Cell,* 35**,** 366-82.

BOROVIAK, T. & NICHOLS, J. 2014. The birth of embryonic pluripotency. *Philos Trans R Soc Lond B Biol Sci,* 369.

BROOK, F. A. & GARDNER, R. L. 1997. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A,* 94**,** 5709-12.

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**,** 838-53.

BULUT-KARSLIOGLU, A., BIECHELE, S., JIN, H., MACRAE, T. A., HEJNA, M., GERTSENSTEIN, M., SONG, J. S. & RAMALHO-SANTOS, M. 2016. Inhibition of mTOR induces a paused pluripotent state. *Nature,* 540**,** 119-123.

CHEN, X., XU, H., YUAN, P., FANG, F., HUSS, M., VEGA, V. B., WONG, E., ORLOV, Y. L., ZHANG, W., JIANG, J., LOH, Y. H., YEO, H. C., YEO, Z. X., NARANG, V., GOVINDARAJAN, K. R., LEONG, B., SHAHAB, A., RUAN, Y., BOURQUE, G., SUNG, W. K., CLARKE, N. D., WEI, C. L. & NG, H. H. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell,* 133**,** 1106-17.

COCKBURN, K. & ROSSANT, J. 2010. Making the blastocyst: lessons from the mouse. *J Clin Invest,* 120**,** 995-1003.

DAVIES, O. R., LIN, C. Y., RADZISHEUSKAYA, A., ZHOU, X., TAUBE, J., BLIN, G., WATERHOUSE, A., SMITH, A. J. & LOWELL, S. 2013. Tcf15 primes pluripotent cells for differentiation. *Cell Rep,* 3**,** 472-84.

DUNN, S. J., MARTELLO, G., YORDANOV, B., EMMOTT, S. & SMITH, A. G. 2014. Defining an essential transcription factor program for naive pluripotency. *Science,* 344**,** 1156-60.

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**,** 351-9.

FIDALGO, M., HUANG, X., GUALLAR, D., SANCHEZ-PRIEGO, C., VALDES, V. J., SAUNDERS, A., DING, J., WU, W. S., CLAVEL, C. & WANG, J. 2016. Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent States. *Cell Stem Cell,* 19**,** 355-69.

GUO, G., VON MEYENN, F., SANTOS, F., CHEN, Y., REIK, W., BERTONE, P., SMITH, A. & NICHOLS, J. 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports,* 6**,** 437-46.

HABIBI, E., BRINKMAN, A. B., ARAND, J., KROEZE, L. I., KERSTENS, H. H., 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**,** 360-9.

HACKETT, J. A. & SURANI, M. A. 2014. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell,* 15**,** 416-30.

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**,** 519-32.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet,* 39**,** 311-8.

HNISZ, D., ABRAHAM, B. J., LEE, T. I., LAU, A., SAINT-ANDRE, V., SIGOVA, A. A., HOKE, H. A. & YOUNG, R. A. 2013. Super-enhancers in the control of cell identity and disease. *Cell,* 155**,** 934-47.

HOFFMAN, J. A., WU, C. I. & MERRILL, B. J. 2013. Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development,* 140**,** 1665-75.

KALKAN, T. & SMITH, A. 2014. Mapping the route from naive pluripotency to lineage specification. *Philos Trans R Soc Lond B Biol Sci,* 369.

KIM, J., CHU, J., SHEN, X., WANG, J. & ORKIN, S. H. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell,* 132**,** 1049-61.

KIM, J. D., KIM, H., EKRAM, M. B., YU, S., FAULK, C. & KIM, J. 2011. Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. *Hum Mol Genet,* 20**,** 1353-62.

KIM, Y. C. & GUAN, K. L. 2015. mTOR: a pharmacologic target for autophagy regulation. *J Clin Invest,* 125**,** 25-32.

KOJIMA, Y., KAUFMAN-FRANCIS, K., STUDDERT, J. B., STEINER, K. A., POWER, M. D., LOEBEL, D. A., JONES, V., HOR, A., DE ALENCASTRO, G., LOGAN, G. J., TEBER, E. T., TAM, O. H., STUTZ, M. D., ALEXANDER, I. E., PICKETT, H. A. & TAM, P. P. 2014. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell,* 14**,** 107-20.

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**,** 517-32.

LEE, H. J., HORE, T. A. & REIK, W. 2014. Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell,* 14**,** 710-9.

LEEB, M., DIETMANN, S., PARAMOR, M., NIWA, H. & SMITH, A. 2014. Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell,* 14**,** 385-93.

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. *Nat Struct Mol Biol,* 20**,** 311-6.

LIU, P., DOU, X., LIU, C., WANG, L., XING, C., PENG, G., CHEN, J., YU, F., QIAO, Y., SONG, L., WU, Y., YUE, C., LI, J., HAN, J. D., TANG, K. & JING, N. 2015. Histone deacetylation promotes mouse neural induction by restricting Nodal-dependent mesendoderm fate. *Nat Commun,* 6**,** 6830.

LU, J., ZAVOROTINSKAYA, T., DAI, Y., NIU, X. H., CASTILLO, J., SIM, J., YU, J., WANG, Y., LANGOWSKI, J. L., HOLASH, J., SHANNON, K. & GARCIA, P. D. 2013. Pim2 is required for maintaining multiple myeloma cell growth through modulating TSC2 phosphorylation. *Blood,* 122**,** 1610-20.

LUO, Z., GAO, X., LIN, C., SMITH, E. R., MARSHALL, S. A., SWANSON, S. K., FLORENS, L., WASHBURN, M. P. & SHILATIFARD, A. 2015. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol Cell,* 57**,** 685-94.

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**,** 590-604.

MARTELLO, G., BERTONE, P. & SMITH, A. 2013. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J,* 32**,** 2561-74.

MARTELLO, G. & SMITH, A. 2014. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol,* 30**,** 647-75.

MARUCCI, L., PEDONE, E., DI VICINO, U., SANUY-ESCRIBANO, B., ISALAN, M. & COSMA, M. P. 2014. beta-catenin fluctuates in mouse ESCs and is essential for Nanog-mediated reprogramming of somatic cells to pluripotency. *Cell Rep,* 8**,** 1686-96.

MASUI, S., OHTSUKA, S., YAGI, R., TAKAHASHI, K., KO, M. S. & NIWA, H. 2008. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev Biol,* 8**,** 45.

MEISSNER, A., GNIRKE, A., BELL, G. W., RAMSAHOYE, B., LANDER, E. S. & JAENISCH, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res,* 33**,** 5868-77.

MURPHY, L. O., MACKEIGAN, J. P. & BLENIS, J. 2004. A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration. *Mol Cell Biol,* 24**,** 144-53.

NAKAI-FUTATSUGI, Y. & NIWA, H. 2016. Zscan4 Is Activated after Telomere Shortening in Mouse Embryonic Stem Cells. *Stem Cell Reports,* 6**,** 483-95.

NAPOLITANO, G. & BALLABIO, A. 2016. TFEB at a glance. *J Cell Sci.*

NARISAWA, S. 2015. Genetically Modified Mice for Studying TNAP Function. *Subcell Biochem,* 76**,** 45-57.

NICHOLS, J., JONES, K., PHILLIPS, J. M., NEWLAND, S. A., ROODE, M., MANSFIELD, W., SMITH, A. & COOKE, A. 2009. Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. *Nat Med,* 15**,** 814-8.

NICHOLS, J. & SMITH, A. 2009. Naive and primed pluripotent states. *Cell Stem Cell,* 4**,** 487-92.

NICHOLS, J. & SMITH, A. 2012. Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect Biol,* 4**,** a008128.

NIWA, H. 2014. The pluripotency transcription factor network at work in reprogramming. *Curr Opin Genet Dev,* 28**,** 25-31.

OHINATA, Y., OHTA, H., SHIGETA, M., YAMANAKA, K., WAKAYAMA, T. & SAITOU, M. 2009. A signaling principle for the specification of the germ cell lineage in mice. *Cell,* 137**,** 571-84.

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**,** 269-80.

OSORNO, R., TSAKIRIDIS, A., WONG, F., CAMBRAY, N., ECONOMOU, C., WILKIE, R., BLIN, G., SCOTTING, P. J., CHAMBERS, I. & WILSON, V. 2012. The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development,* 139**,** 2288-98.

PELTON, T. A., SHARMA, S., SCHULZ, T. C., RATHJEN, J. & RATHJEN, P. D. 2002. Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J Cell Sci,* 115**,** 329-39.

RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature,* 470**,** 279-83.

RESPUELA, P., NIKOLIC, M., TAN, M., FROMMOLT, P., ZHAO, Y., WYSOCKA, J. & RADA-IGLESIAS, A. 2016. Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. *Cell Stem Cell,* 18**,** 118-33.

REZENDE, N. C., LEE, M. Y., MONETTE, S., MARK, W., LU, A. & GUDAS, L. J. 2011. Rex1 (Zfp42) null mice show impaired testicular function, abnormal testis morphology, and aberrant gene expression. *Dev Biol,* 356**,** 370-82.

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**,** 849-62.

SINGER, Z. S., YONG, J., TISCHLER, J., HACKETT, J. A., ALTINOK, A., SURANI, M. A., CAI, L. & ELOWITZ, M. B. 2014. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol Cell,* 55**,** 319-31.

SMITH, A. in press. Formative pluripotency: the executive phase in a developmental continuum. *Development.*

SMITH, Z. D., CHAN, M. M., MIKKELSEN, T. S., GU, H., GNIRKE, A., REGEV, A. & MEISSNER, A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature,* 484**,** 339-44.

SOLTER, D., SKREB, N. & DAMJANOV, I. 1970. Extrauterine growth of mouse egg cylinders results in malignant teratoma. *Nature,* 227**,** 503-504.

TAKASHIMA, Y., GUO, G., LOOS, R., NICHOLS, J., FICZ, G., KRUEGER, F., OXLEY, D., SANTOS, F., CLARKE, J., MANSFIELD, W., REIK, W., BERTONE, P. & SMITH, A. 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell,* 158**,** 1254-69.

TAM, P. P. & ZHOU, S. X. 1996. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol,* 178**,** 124-32.

THOMSON, M., LIU, S. J., ZOU, L. N., SMITH, Z., MEISSNER, A. & RAMANATHAN, S. 2011. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell,* 145**,** 875-89.

TOYOOKA, Y., SHIMOSATO, D., MURAKAMI, K., TAKAHASHI, K. & NIWA, H. 2008. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development,* 135**,** 909-18.

VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457**,** 854-8.

VISEL, A., TAHER, L., GIRGIS, H., MAY, D., GOLONZHKA, O., HOCH, R. V., MCKINSEY, G. L., PATTABIRAMAN, K., SILBERBERG, S. N., BLOW, M. J., HANSEN, D. V., NORD, A. S., AKIYAMA, J. A., HOLT, A., HOSSEINI, R., PHOUANENAVONG, S., PLAJZER-FRICK, I., SHOUKRY, M., AFZAL, V., KAPLAN, T., KRIEGSTEIN, A. R., RUBIN, E. M., OVCHARENKO, I., PENNACCHIO, L. A. & RUBENSTEIN, J. L. 2013. A high-resolution enhancer atlas of the developing telencephalon. *Cell,* 152**,** 895-908.

WANG, J., ALEXANDER, P., WU, L., HAMMER, R., CLEAVER, O. & MCKNIGHT, S. L. 2009. Dependence of mouse embryonic stem cells on threonine catabolism. *Science,* 325**,** 435-9.

WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

WRAY, J., KALKAN, T., GOMEZ-LOPEZ, S., ECKARDT, D., COOK, A., KEMLER, R. & SMITH, A. 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3

repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol,* 13**,** 838-45.

WRAY, J., KALKAN, T. & SMITH, A. G. 2010. The ground state of pluripotency. *Biochem Soc Trans,* 38**,** 1027-32.

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**,** 368-82.

YANG, S. H., KALKAN, T., MORISSROE, C., MARKS, H., STUNNENBERG, H., SMITH, A. & SHARROCKS, A. D. 2014. Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. *Cell Rep,* 7**,** 1968-81.

YE, S., LI, P., TONG, C. & YING, Q. L. 2013. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J,* 32**,** 2548-2560.

YING, Q. L., STAVRIDIS, M., GRIFFITHS, D., LI, M. & SMITH, A. 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol,* 21**,** 183-6.

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**,** 519-23.

YOUNG, R. A. 2011. Control of the embryonic stem cell state. *Cell,* 144**,** 940-54.

YU, Z., ZHAO, X., HUANG, L., ZHANG, T., YANG, F., XIE, L., SONG, S., MIAO, P., ZHAO, L., SUN, X., LIU, J. & HUANG, G. 2013. Proviral insertion in murine lymphomas 2 (PIM2) oncogene phosphorylates pyruvate kinase M2 (PKM2) and promotes glycolysis in cancer cells. *J Biol Chem,* 288**,** 35406-16.

ZHANG, J., RATANASIRINTRAWOOT, S., CHANDRASEKARAN, S., WU, Z., FICARRO, S. B., YU, C., ROSS, C. A., CACCHIARELLI, D., XIA, Q., SELIGSON, M., SHINODA, G., XIE, W., CAHAN, P., WANG, L., NG, S. C., TINTARA, S., TRAPNELL, C., ONDER, T., LOH, Y. H., MIKKELSEN, T., SLIZ, P., TEITELL, M. A., ASARA, J. M., MARTO, J. A., LI, H., COLLINS, J. J. & DALEY, G. Q. 2016. LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell,* 19**,** 66-80.

ZHANG, X. H., YU, H. L., WANG, F. J., HAN, Y. L. & YANG, W. L. 2015. Pim-2 Modulates Aerobic Glycolysis and Energy Production during the Development of Colorectal Tumors. *Int J Med Sci,* 12**,** 487-93.

ZHOU, W., CHOI, M., MARGINEANTU, D., MARGARETHA, L., HESSON, J., CAVANAUGH, C., BLAU, C. A., HORWITZ, M. S., HOCKENBERY, D., WARE, C. & RUOHOLA-BAKER, H. 2012. HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J,* 31**,** 2103-16.
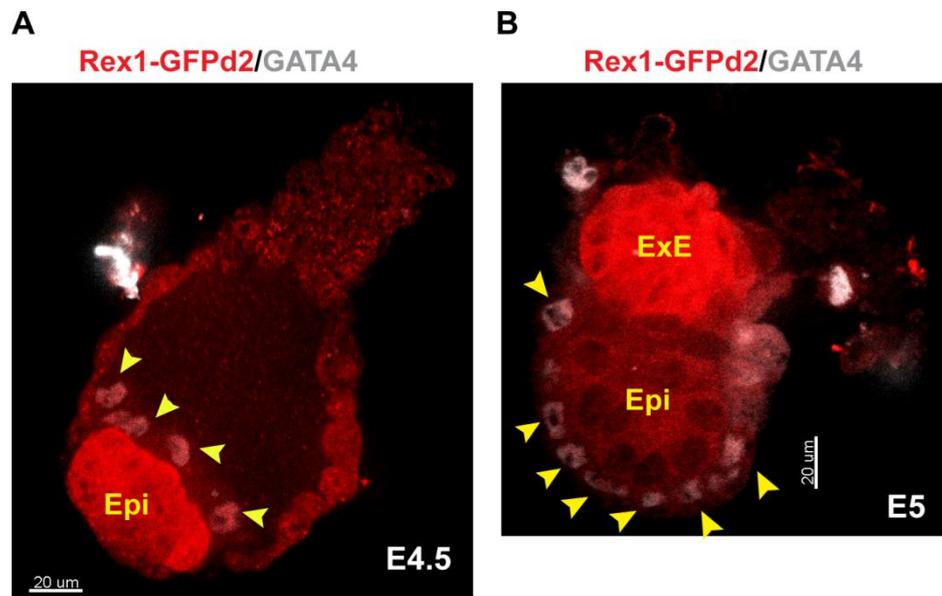
Figures

Fig. 1. Expression of the RGd2 reporter before and after implantation IF staining for GFP (Rex1GFPd2) (red) and Gata4 (gray) **(A)** E4.5 **(B)** E5. Arrowheads show GATA4-positive nuclei. Scale bar is 20µm. Extraembryonic ectoderm (ExE). Epiblast (Epi).
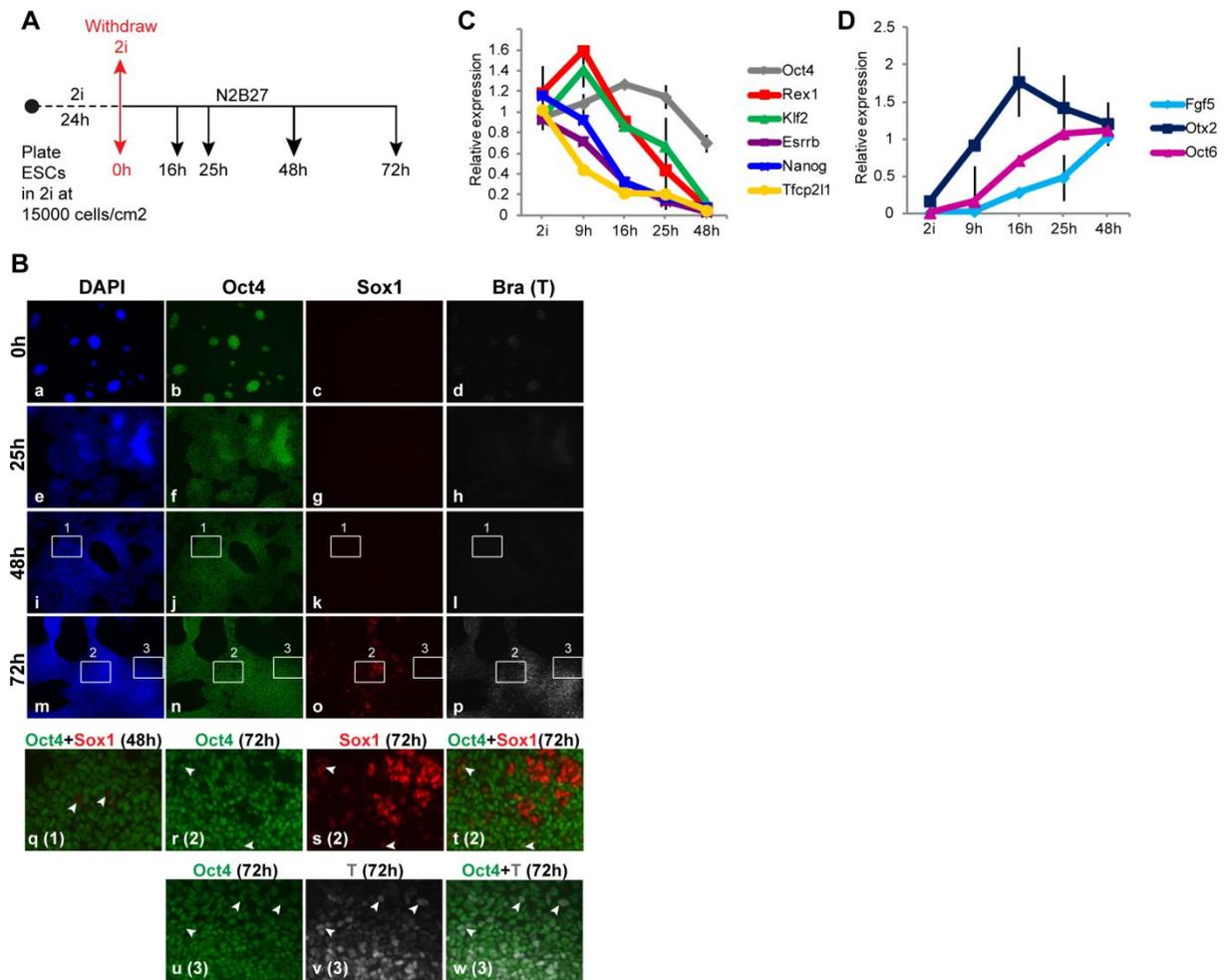
Fig. 2. Multilineage specification of ES cells upon release from 2i
(A) Protocol for monolayer differentiation of naive ES cells in N2B27 by withdrawal of 2i. **(B)** IF staining for Oct4, Sox1and Bra (*T*). Lower panels (**q-w**) show enlarged insets from 48h and 72h, with respective inset number in parentheses. RT-qPCR for selected **(C)** pluripotency **(D)** early post-implantation epiblast markers. Expression levels are shown as fold change relative to naive ES cells in panel C and 48h samples in panel D. GAPDH was used for normalization. Error bars depict standard deviation (sd) from 2 biological replicates.

A

B

DAPI Rex1-GFPd2 Nanog Otx2

0h   a   b   c   d

16h e   f   g   h

25h i   j   k   l

C

Otx2 IF intensity (a.u.)

0h

16h

25h

Nanog
IF intensity (a.u.)

D

cell number

0h

16h

25h

Nanog   Nanog   Nanog   Nanog
high   mid   low   neg

Otx2 high   Otx2 low
Otx2 mid   Otx2 neg

E

Rex1-GFPd2
IF intensity (a.u.)

25h

Nanog
IF intensity (a.u.)

F

cell number

25h

Rex1   Rex1   Rex1   Rex1
high   mid   low   neg

Nanog high
Nanog mid
Nanog low
Nanog neg

G

DAPI   Oct4   Sox2

2i

25h
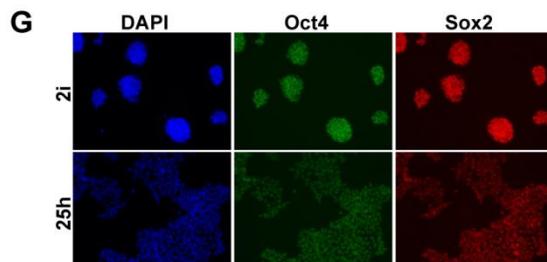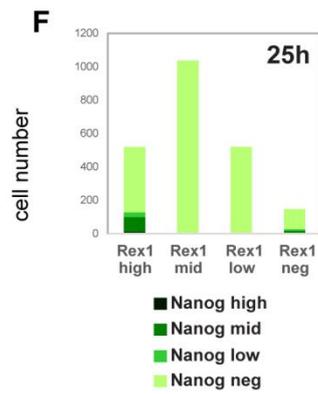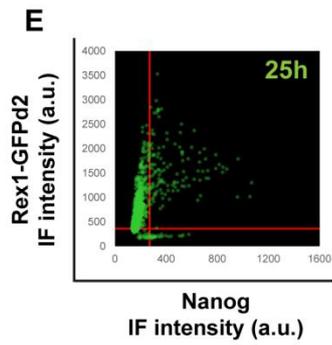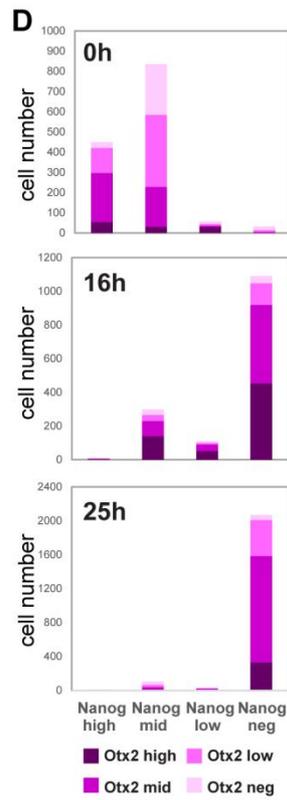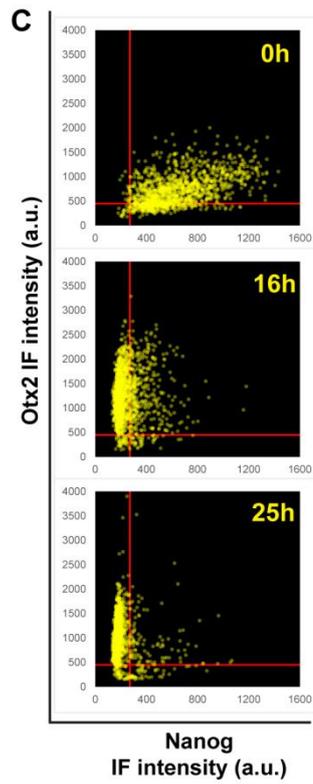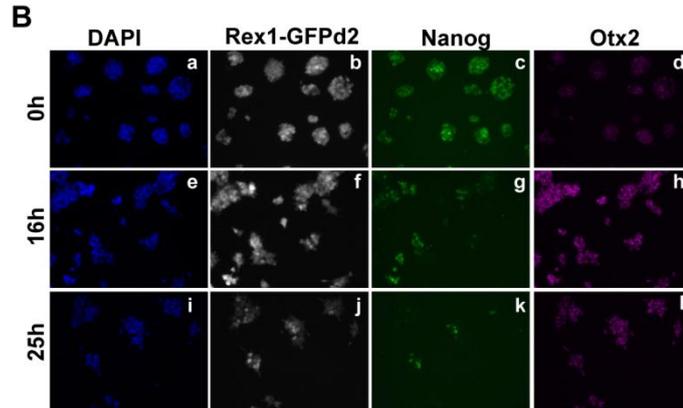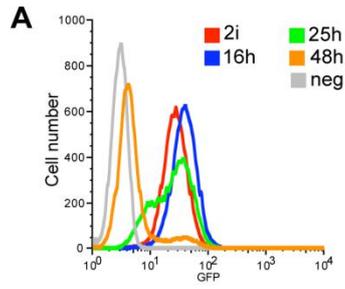
Fig. 3. Expression of transcription factors during transition of ES cells
(A) GFP flow cytometry profile at indicated time points post-2i withdrawal (10000 cells per time-point). Wild-type ES cells were used as negative control (neg). **(B)** IF staining for GFP, Nanog, Otx2 (a-l). **(C)** Otx2 vs Nanog fluorescence intensity per cell in arbitrary units (a.u) as quantified by Volocity. X and Y intercepts of the red lines mark the cut-off for Nanog- and Otx2-negative cells, respectively. **(D)** Distribution of Otx2 expression in Nanog-subpopulations. **(E)** GFP vs Nanog fluorescence intensity per cell. X and Y intercepts of the red lines mark the cut-off for Nanog- and GFP-negative cells, respectively **(F)** Distribution of Nanog expression in Rex1-subpopulations at 25h **(G)** IF staining for Oct4 and Sox2

Figure 4. Downregulation of Rex1 tracks exit from the naive state

(A) Protocol for sorting and colony assays **(B)** Clonogenicity of cells from 2i and differentiating subpopulations sorted at indicated time points, plated in serum/LIF (Serum/L) or 2i supplemented with LIF (2i/L). SD from 2 technical replicates **(C)** Sorting of 25h-cultures into 4 subpopulations based on GFP levels by flow cytometry. Lower plot shows the GFP profiles of post-sort subpopulations. **(D)** Clonogenicity of 4 subpopulations shown in C. SD from 2 technical replicates **(E)** Diagram summarizing phases of transition from the naive state **(F)** Clonogenicity of the indicated subpopulations. SD from 3 biological replicates each with 2 technical

replicates **(G)** Immunoblot on total cell lysates from sorted subpopulations. β-tubulin and Gapdh are loading controls. **(H)** Expression of selected general (red) and naive (orange) pluripotency and early post-implantation epiblast (blue) markers in single cells measured by Fluidigm system**.** Scale bar represents Log2 transformed expression value.

A

MEK/ERK targets: Myc, Spry2, Dusp4, Dusp6, Egr1, Fos, Jun

Wnt/b-catenin targets: Cdx2, Cdx1, T, Axin2

Wnt/b-catenin transcriptional effectors: Ctnnb1, Tcf7, Tcf7l1 (Tcf3), Tcf7l2 (Tcf4), Lef1

Log Expression: 6 8 12
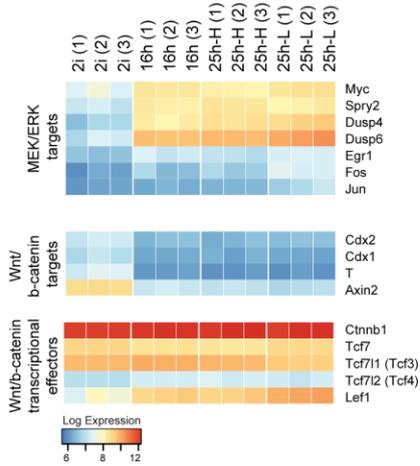
B

UP in 16h vs 2i
Cell adhesion molecules (CAMs)
Leukocyte transendothelial migration
Steroid biosynthesis
Cell cycle
Pathways in cancer
Regulation of actin cytoskeleton
Focal adhesion

UP in 25h-L vs 2i
Ribosome biogenesis in eukaryotes
Spliceosome
RNA transport
mRNA surveillance pathway
DNA replication
Nucleotide excision repair
Proteasome
MAPK signaling pathway
VEGF signaling pathway
Purine metabolism
Pyrimidine metabolism

DOWN in 16h vs 2i
Lysosome
Metabolic pathways
Other glycan degradation
Amino sugar and nucleotide sugar metabolism
Oxidative phosphorylation
Glycosaminoglycan degradation
Collecting duct acid secretion
Sphingolipid metabolism
N-Glycan biosynthesis
Glycosphingolipid biosynthesis - globo series
Vitamin B6 metabolism
Mucin type O-Glycan biosynthesis
Starch and sucrose metabolism

DOWN in 25h-L vs 2i
Lysosome
Metabolic pathways
Oxidative phosphorylation
Valine, leucine and isoleucine degradation
Amino sugar and nucleotide sugar metabolism
Peroxisome
Propanoate metabolism
Cardiac muscle contraction
Regulation of autophagy
Glycosphingolipid biosynthesis - globo series
Glycolysis / Gluconeogenesis
Protein digestion and absorption
Other glycan degradation
Glycosaminoglycan degradation

C

D

Pou5f1, Sall4, Sox2, Zic2, Zic3, Zfp281, Lin28a, Utf1, Nanog, Nr0b1, Zfp42, Klf5, Ifitm3, Klf2, Nr5a2, Esrrb, Tfcp2l1, Tcl1, Tbx3, Elf3, Klf4, Hoxa1, Eomes, Nanos3, Prdm1, Pou3f3, Hoxb1, Ascl1, Gata4, Otx1, Gata6, Zic1, Foxa2, T, Cdx2, Cdx1, Gsc, Sox7, Meox1, Sox17, Pax3, Foxf1, Kdr, Dppa3, Lhx1, Cxcr4, Nes, Mixl1, Neurog1, Pou3f2, Neurog2, Dnd1, Cer1, Pax6, Ddx4, Sox3, Sox1, Foxd3, Fgf5, Dazl, Cdh2, Pou3f1, Lef1, Otx2, Dnmt3b, Dnmt3a

Log Expression: 6 8 12

Fig. 5. Transcriptional changes in ES cells during progression from naive pluripotency.

(A) Expression of MEK/ERK and Wnt/b-catenin transcriptional effectors and targets from 3 independent replicates measured by microarray profiling. Scale represents Log2 transformed expression value **(B)** Enriched KEGG pathway categories in the differentially expressed gene sets ranked according to p-value (p<0.05) **(C)** OCR and ECAR levels of 2i vs 25h populations (left) and sorted 25h-H and 25h-L subpopulations with unsorted whole population (right) (SD from 6 technical replicates) **(D)** Expression of general (red) and naive (orange) pluripotency, early post-implantation (blue) and lineage-priming factors (black) detected by microarray. Scale represents Log2 transformed expression value.

A

	DOWN	UP
Transcription factors	Zfp42, Tcl1, Kl2, Klf4, Ncoa3, Esrrb, Tfcp2l1,Tbx3, Prdm14, Dppa5a, Dppa2, Nr0b1,Nanog, Fbx015, Eif2s2, Mybl2, Cdx1, Tfeb, Elf3, Tead4, Tfap2	Sox3, Otx2, Sall2, Pou3f1 Usf1, Limd1, Limd2, Sox12, Crip2, Nfxl1,Tead2, Hes6
Epigenetic regulators	Tet2, Zfp57, Zfp819, Arid5b, H2AFγ, Cdyl2	Dnmt3a, Dnm3b, Mael, Phf10, Zmynd8
Secreted proteins	Pdgfa, Vegfc, Bmp4, Ctgf, Il34, Htra1 *ECM* Lama1, Lamc1, Fn1, Nid1, Spp1, Sparc, Lgals3	Enpp2, Rspo1, Mdk *ECM* Col18a
Surface proteins	Cd38, Gpa33, Slc6a8, Anxa2, Mme *Surface receptors* Susd2, Epha4, Lpar1, Lifr, Ilst6, Zp3, Ooep, Folr1, Cd63, Ngfr, Rell1, Gpr160, Adrb3, Syngr1, Gprc5a *Cell adhesion molecules* Jam2, Pecam1, Icam1, Gjb3/5, Cd9	Tspan4/7, Emb, Hpn, Tmem51/54/98/237 *Surface receptors* Lrp1, Lrp2, Fzd2, Gprc5c, Smo Unc5b, Sema6a, Sort1, Ptprf *Cell adhesion molecules* Cd276, Cldn7
Cell shape and motility	Rfx2, Rreb1, Llgl2, Cobl, Vangl1, Arpc1b, Ctgf, Ooep, Kifc3, Hspb1, Lpar1, Flnb, Ckb, Slc6a8, Capns1, Tubb2b, Myh13, Anxa2, Sept1, Stmn3	Shroom2, Dab2, Vangl2, Trio, Pfn1 Kif1a, Tubb6, Ina, Enpp2, Bcar1 Rab25, Ptpn14, Myh10, Hpn, Arhgef16, Sema6a, Efs, Dbn1
Metabolism	Pfkp, Sod2, Gpx1, Tdh, Manba, Fbxo6, Renbp, Hexb, Ctsc, Laptm5, Ppap2a, Echdc2, Ephx2, Fabp3, Acaa2,Pigl, Pigp	Pim2, Ldha, Stard4, Galnt3, Parp8
mTOR and autophagy	Atg13, Tbc1d25, Map1lc3b, Capns1, Ulk1,Gabarapl2, Wdr45	Pim2, Ddit4l, Igtp
Wnt	Dkk1, Porcn, Notum, Vangl1	Vangl2, Rspo1, Fzd2
LIF/ Stat	Lifr, Ilst6, Jak3, Stat3	
Tgfb	Lefty1, Lefty2, BMP4	
RNA processing	Apobec2, Rpp25, Ddx58 Lact2b , Grsf1	

B
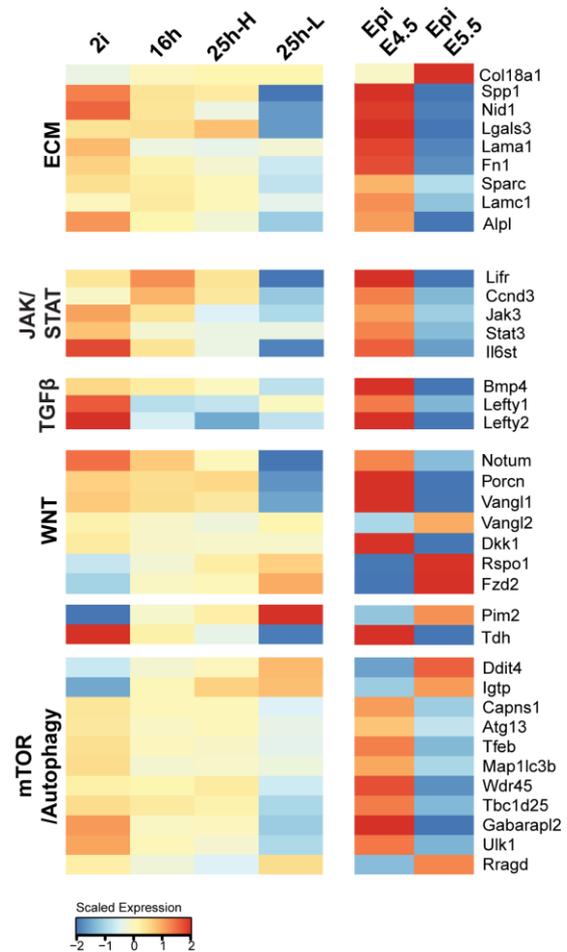
Fig. 6. Comparison of transcriptional changes during pluripotency progression in ES cells and in the embryo
(A) Functional grouping of genes that show similar regulation in ES cells and in the embryo **(B)** Expression of genes from selected pathways
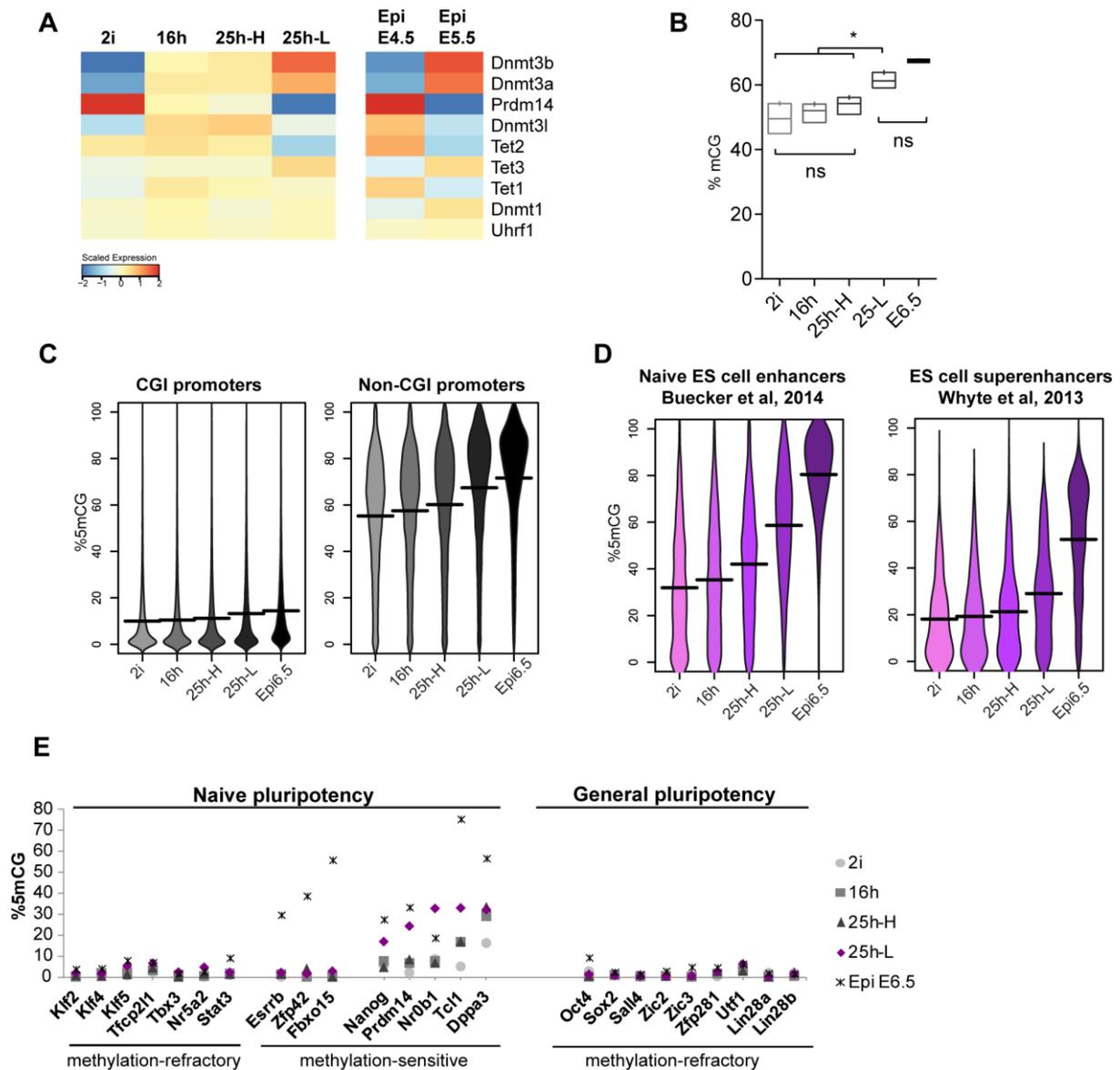
Fig. 7. Acquisition of DNA methylation during transition from naïve pluripotency.

(A) mRNA expression of factors that modulate DNA methylation **(B)** Global genomic methylation in CG context (mCG) in 2kb tiles. ns - non significant, * denotes significance (One way multiple comparisons ANOVA corrected with Tukey's test, p<0.05) **(C)** Percentage of mCG in the promoters (-1000 to +500 of TSS) of expressed genes (RPKM≥10 from RNA-seq), **(D)** naïve enhancers and super enhancers, **(E)** promoters of pluripotency genes.

Supplementary Figures
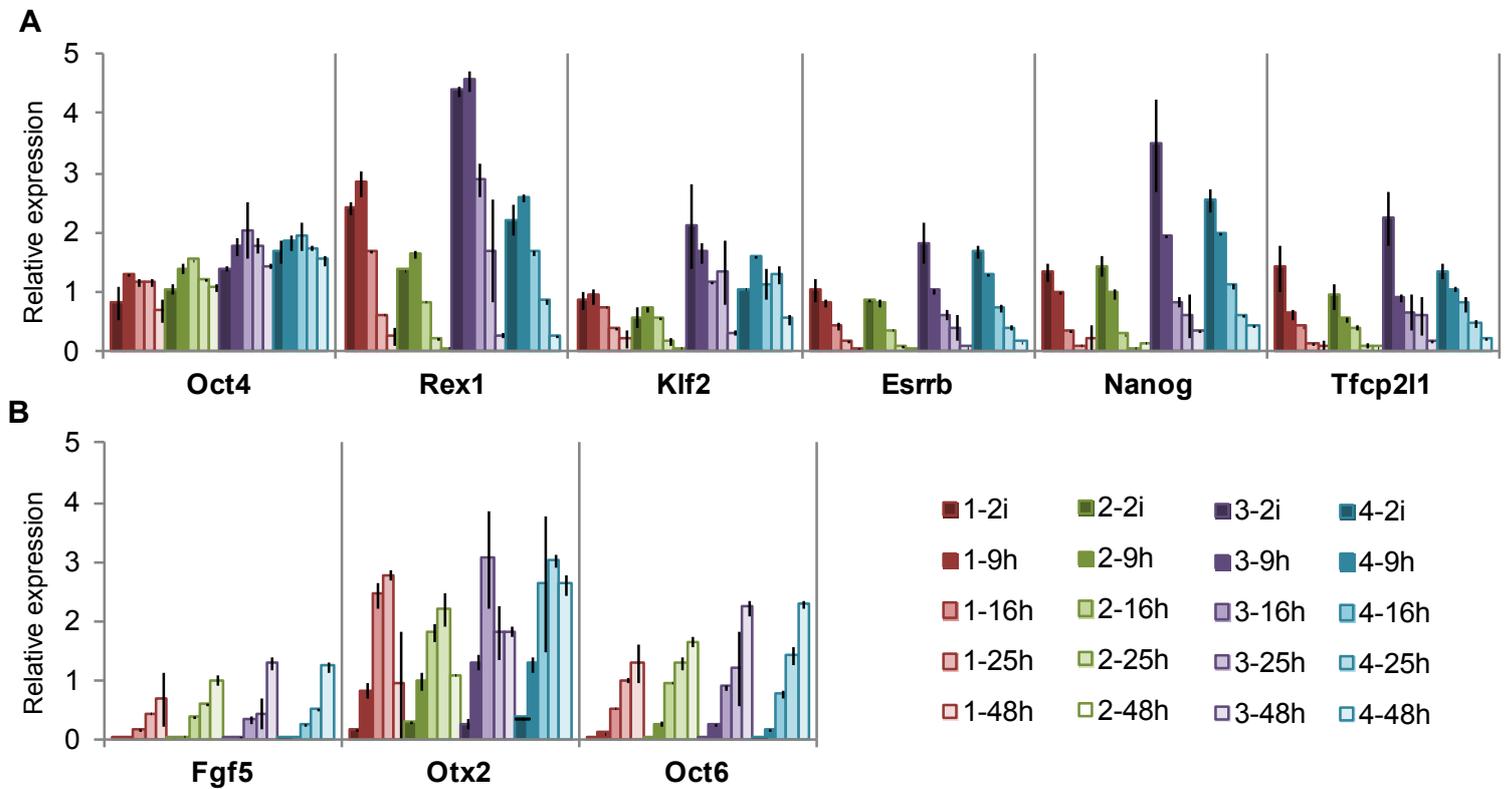
Fig. S1. Gene expression changes in ES cells upon release from 2i

Expression of **(A)** pluripotency **(B)** early post-implantation epiblast markers in 4 different ES cell lines measured by RT-qPCR (SD from 2 biological replicates). (1) E14vC (male, wt), (2) RGd2 1903.4 (male, RGd2 knock-in), (3) 129 (female, wt) (4) RGd2 1903.3 (female, RGd2 knock-in). GAPDH was used for normalization cDNA amount. Expression levels are represented as fold changes relative to 2i sample from RGd2 1903.4 ES cells for pluripotency markers and 48h sample for post-implantation epiblast markers.
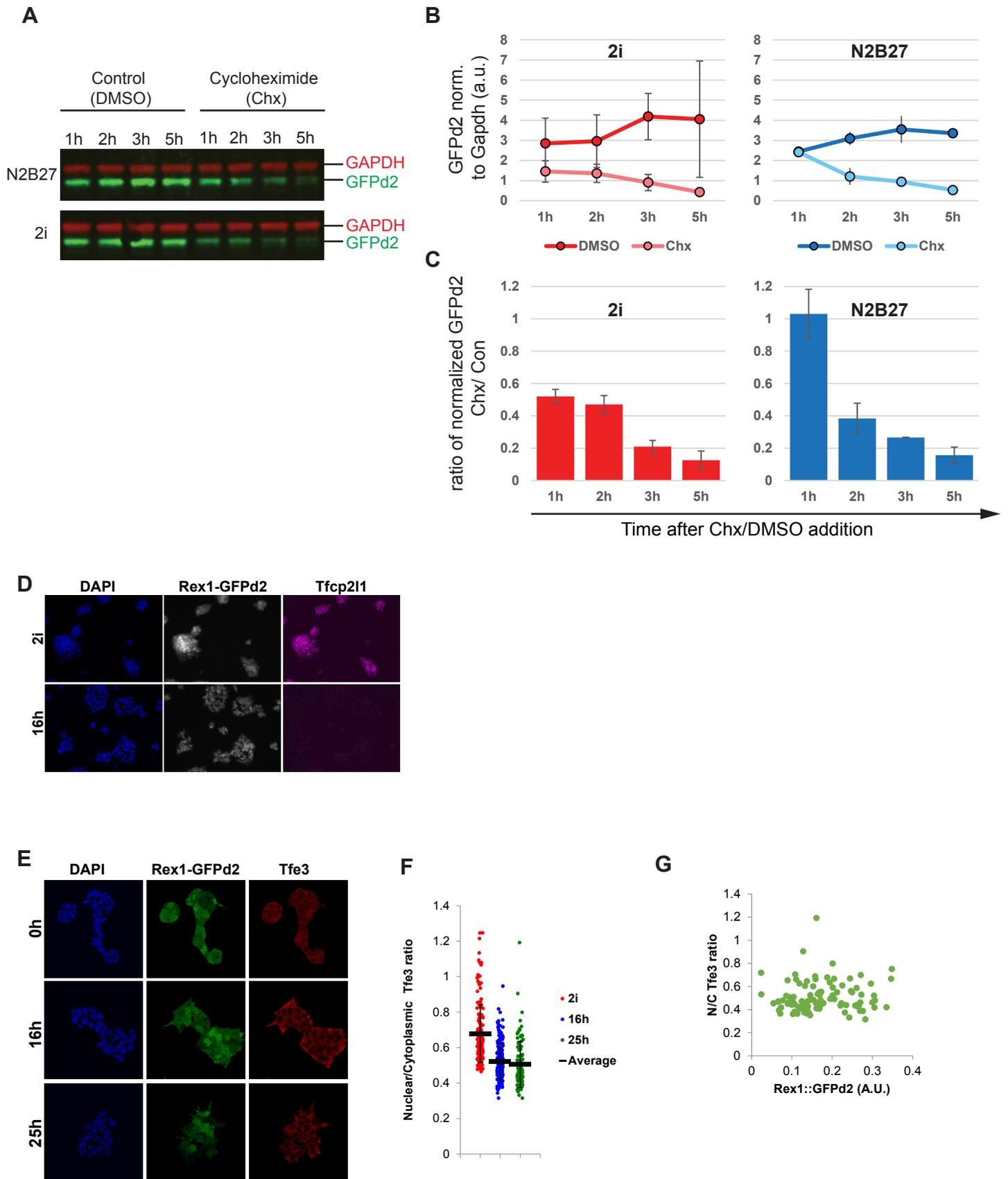
Fig. S2. Half-life of GFPd2 and IF staining for Tfcp2l1 and Tfe3

(A) Western blot for GFP following Cycloheximide (Chx) or DMSO (control) treatment of ES cells in 2i and N2B27 using Odessey imaging system. GAPDH was used as loading control. "h" indicates hours after addition of Chx or DMSO. **(B)** GFPd2 levels normalized to GAPDH after quantification on Odessey (SD from 2 biological replicates) **(C)** Ratio of normalized GFPd2 in chx-treated vs control samples. (SD from 2 biological replicates). **(D)** IF staining for GFP and Tfcp2l1 **(E)** GFP and Tfe3 (**F**) Nuclear to cytoplasmic ratio (N/C) of Tfe3 in single cells quantified using Cell Profiler (~ 150 cells/sample).Black bars show the mean. **(G)** N/C of Tfe3 vs. GFP intensity in single cells in the 25h population.
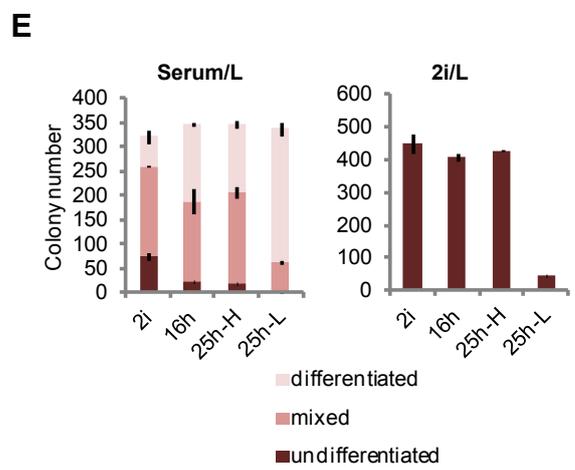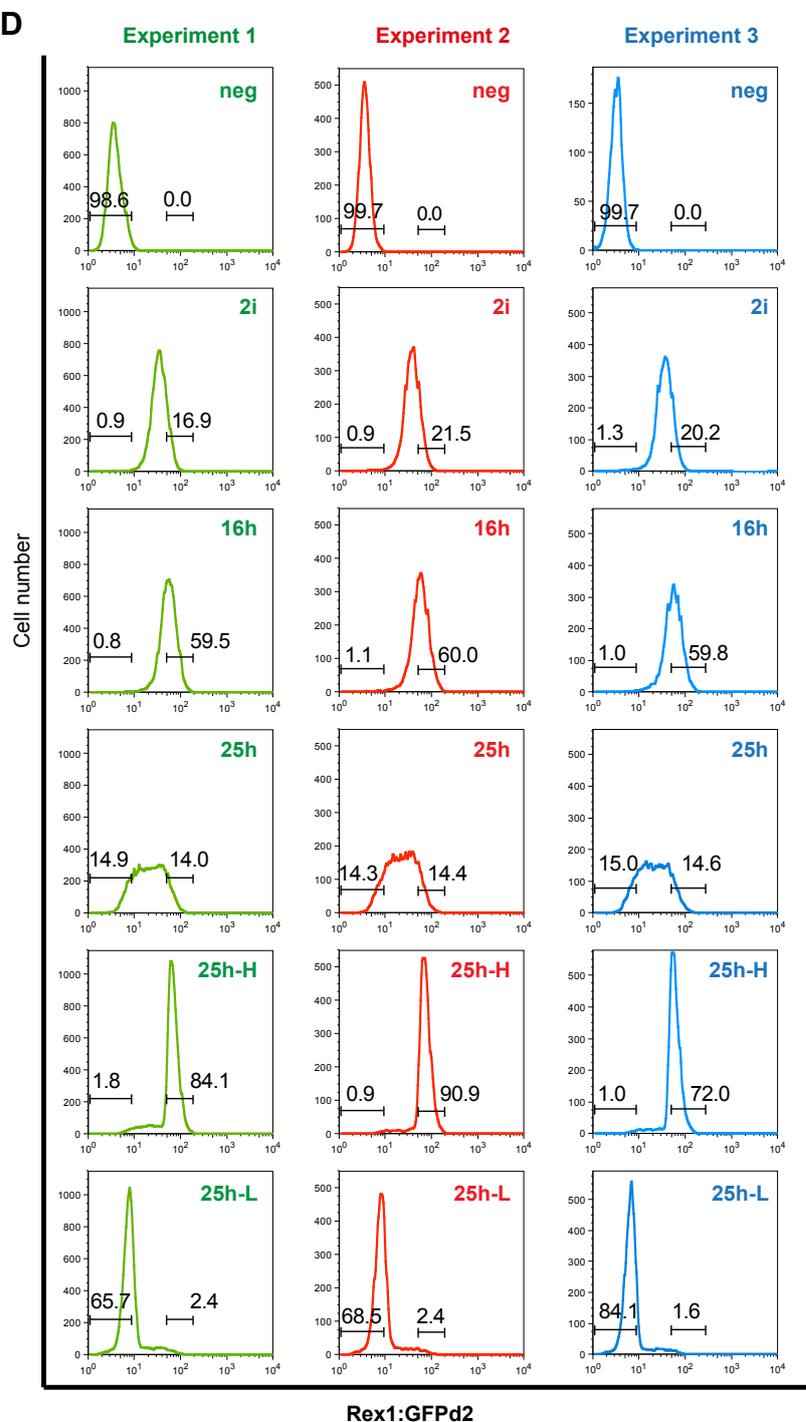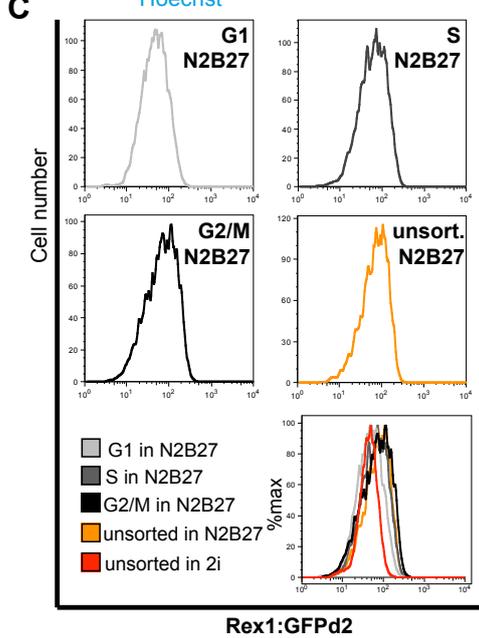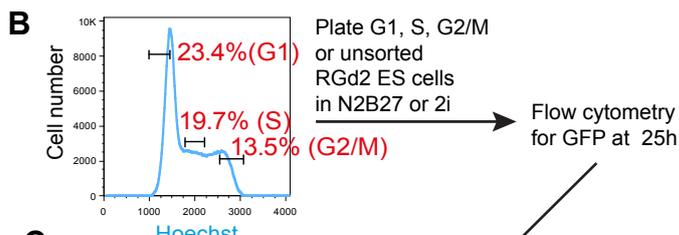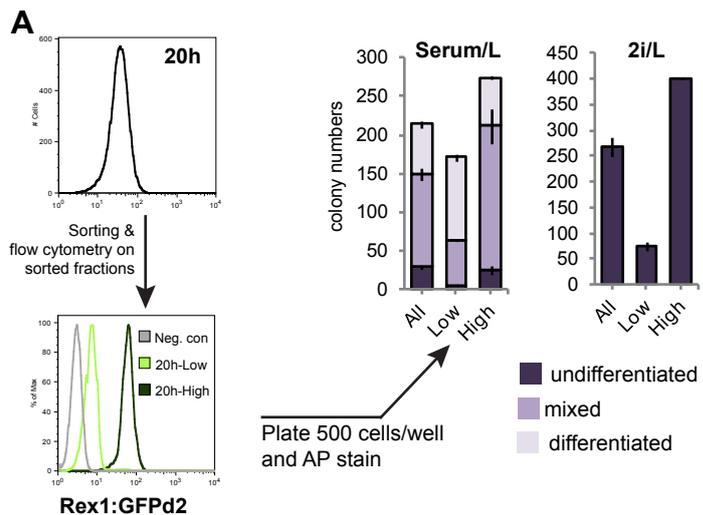
Fig. S3. Downregulation of Rex1 tracks exit from the naive state (related to Fig. 4.)

(A) GFP profiles and clonogenicity of sorted Rex1-GFPd2-High/Low subpopulations and total population (All) from 20h cultures. GFP profiles of sorted subpopulations are shown in the lower histogram. A wild type ES cell line was used as negative control (Neg) (SD from 2 technical replicates) **(B)** Hoechst profile of ES cells cultured in 2i. Gates for sorting of G1, S, and G2/M subpopulations and respective percentage of cells are displayed on the histogram. **(C)** GFP profiles of sorted populations and unsorted Hoechst-stained ES cells at 25 h after plating. **(D)** GFP profiles of whole and sorted populations from 3 independent experiments. All populations including 2i and 16h cultures were stained with ToPro 3 as a dead cell indicator prior to sorting, and ToPro-negative cells were isolated for subsequent analysis. Gates encompassing the highest and the lowest ~15% GFP-expressing cells in 25h cultures are shown as black bars and percentages of cells falling into these gates are shown. Sorted 25h-H and 25h-L subpopulations were reanalyzed by flow cytometry to determine purity and the respective profiles are shown in the bottom histograms **(E)** Clogenicity of sorted subpopulations from RGd2 ES cell line (1903.4) (SD from 2 biological replicates each with 2 technical replicate).
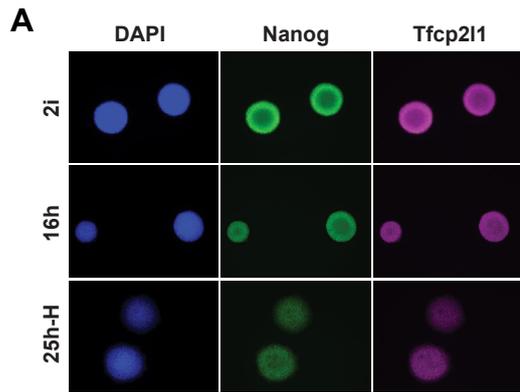
A

B

■ post-sort pre-plating
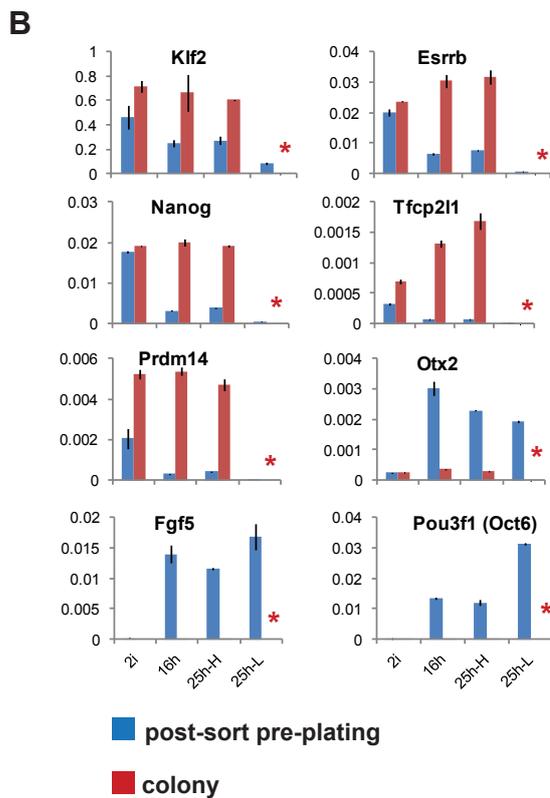
■ colony

Fig S4. Re-establishment of naive state gene expression in reverted cells (Related to Fig 4)

(A) IF staining for Nanog and Tfcp2l1 **(B)** RT-qPCR on colonies from 2i, 16h and 25h-L cells 6 days after re-plating in 2i/L at clonal density. RT-qPCR was performed on samples right after sorting and on pooled colonies 5 days after re-plating (sd from 2 technical replicates). Red asterisk indicates absence of colony samples from 25h-L population due to loss of reversion ability.
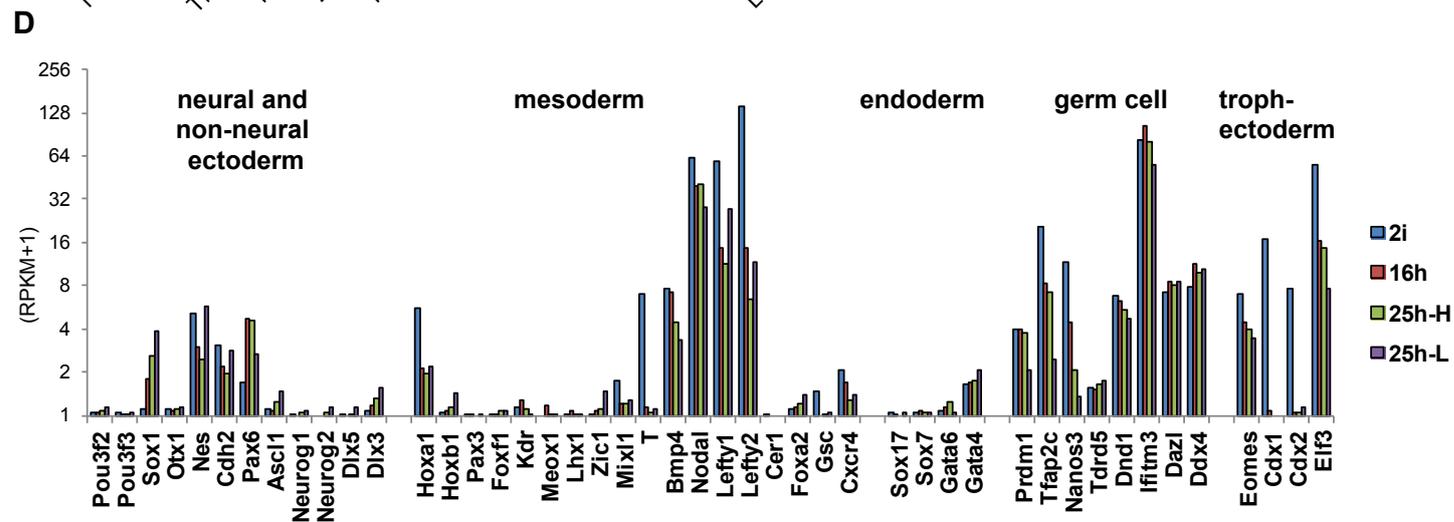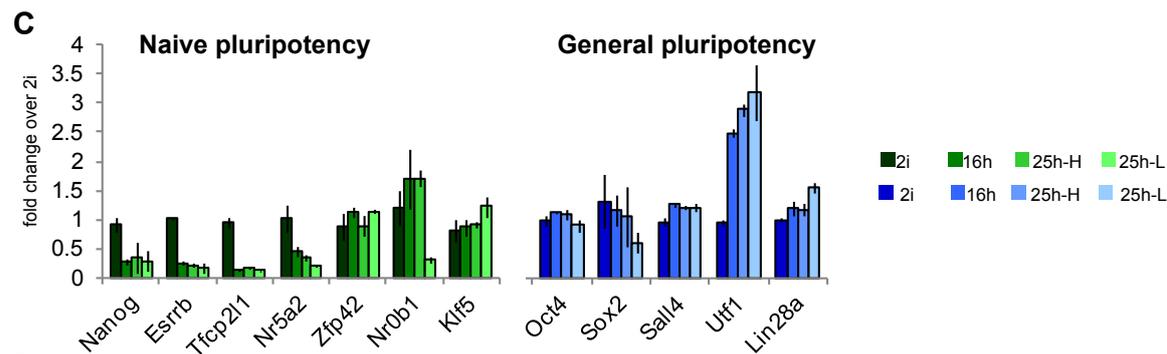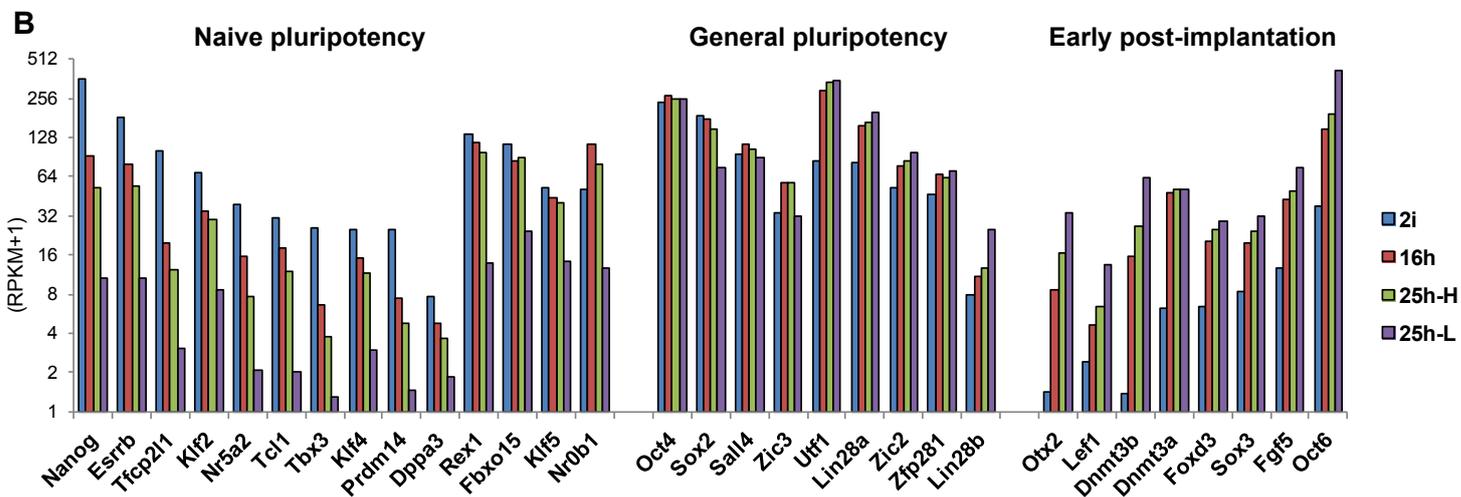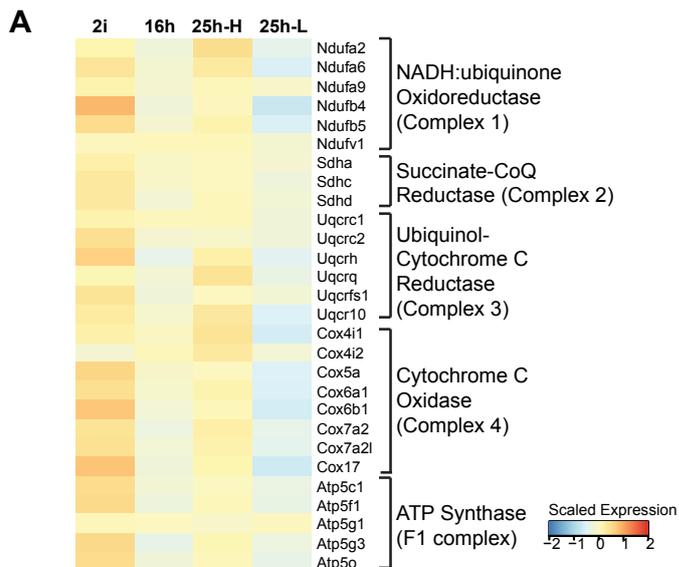
Fig. S5. Changes in mRNA and protein levels of selected genes during the transition (related to Fig. 5)

(A) Expression of mitochondrial ETC complex subunits from RNA-seq. **(B)** Selected pluripotency and post-implantation epiblast markers measured by RNA-seq. **(C)** Relative nuclear protein levels in transiting populations measured by mass spectroscopy, displayed as fold change over levels in 2i. Error bars indicate s.d. from 3 biological replicates, except for Nr5a2, Rex1 (*Zfp42*), Nr0b1, Sox2 and Tfcp2l1, which were not detected in all replicates of the 25h-L fraction, most likely due to reduced levels**. (D)** Transcript levels of lineage markers measured by RNA-seq.
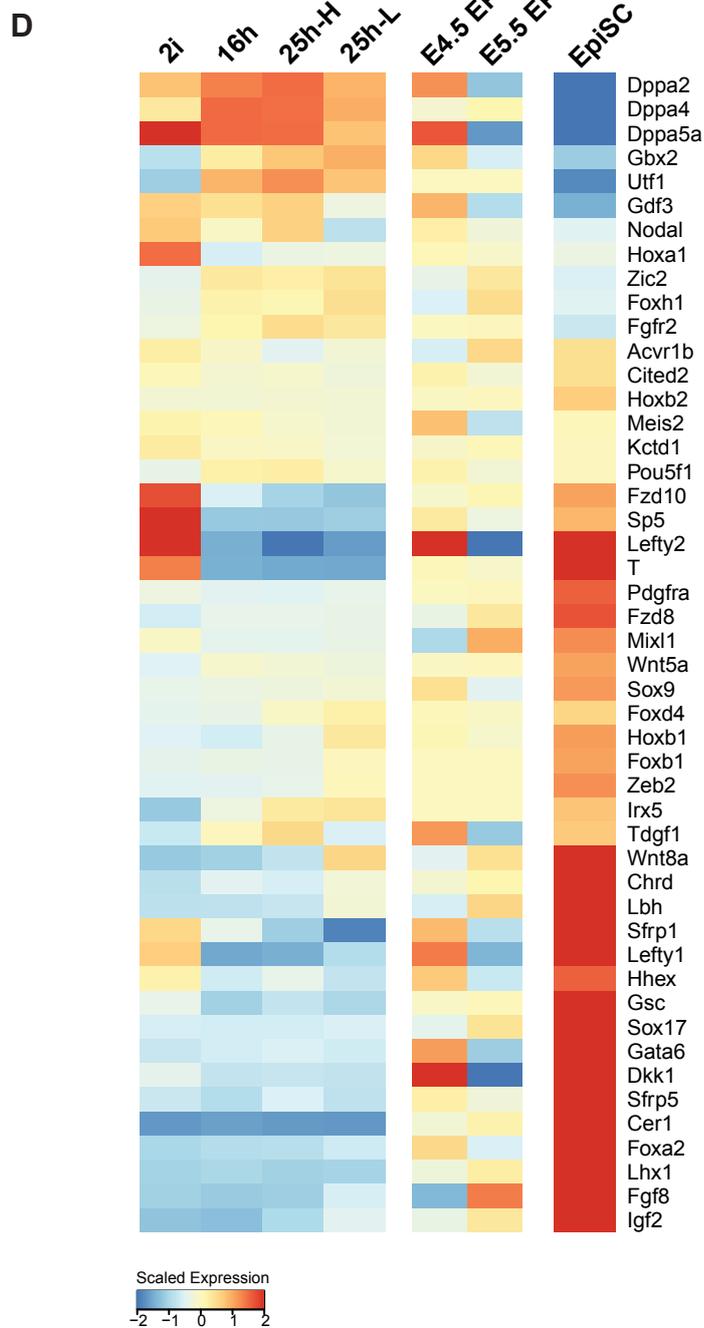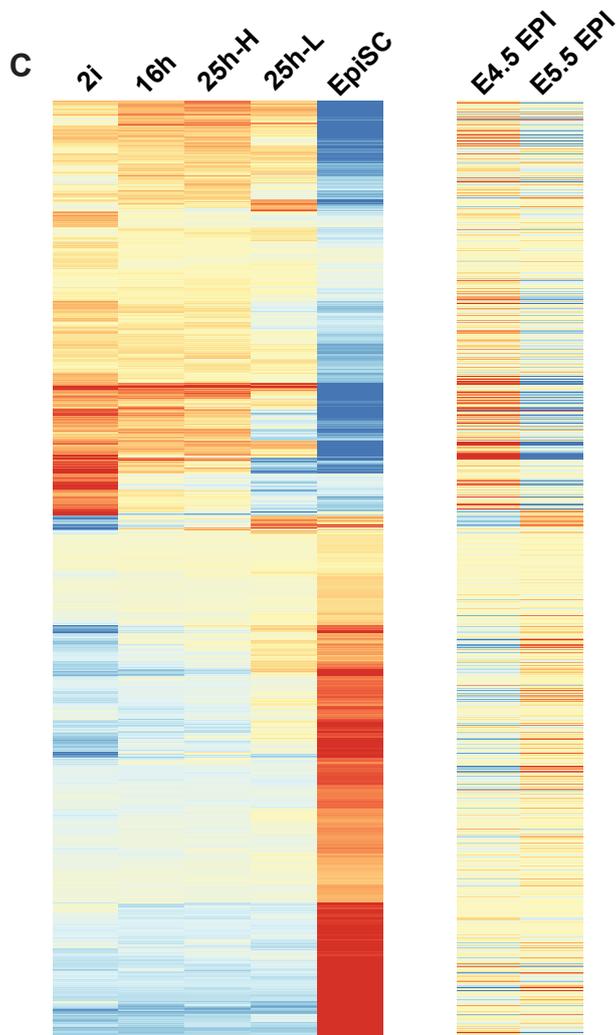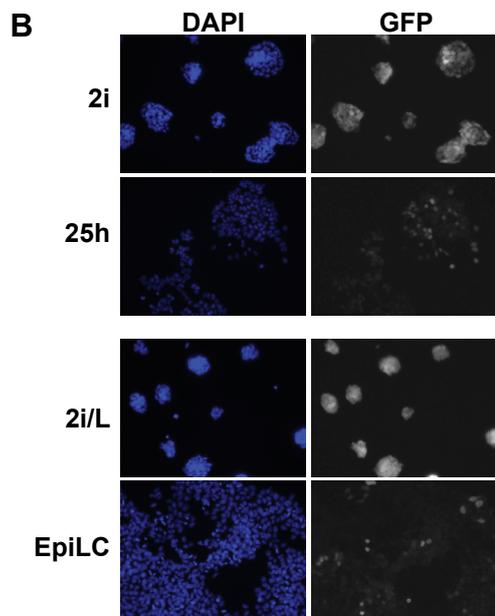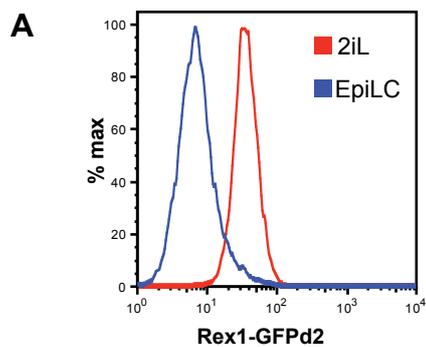
Fig. S6. Comparison of Rex1 populations with EpiLCs and EpiSCs

(A) GFP profiles of EpiLCs generated from RGd2 ES cells and RGd2 ES cells cultured in 2i/L. **(B)** IF staining for GFP in 25h cultures and EpiLCs generated from RGd2 ES cells**. (C)** Expression of the differentially expressed gene set between EpiSCs and ES cells (from Kojima et, al 2014). **(D)** Expression of selected EpiSC- or Epiblast-specific genes (Kojima et, al 2014).
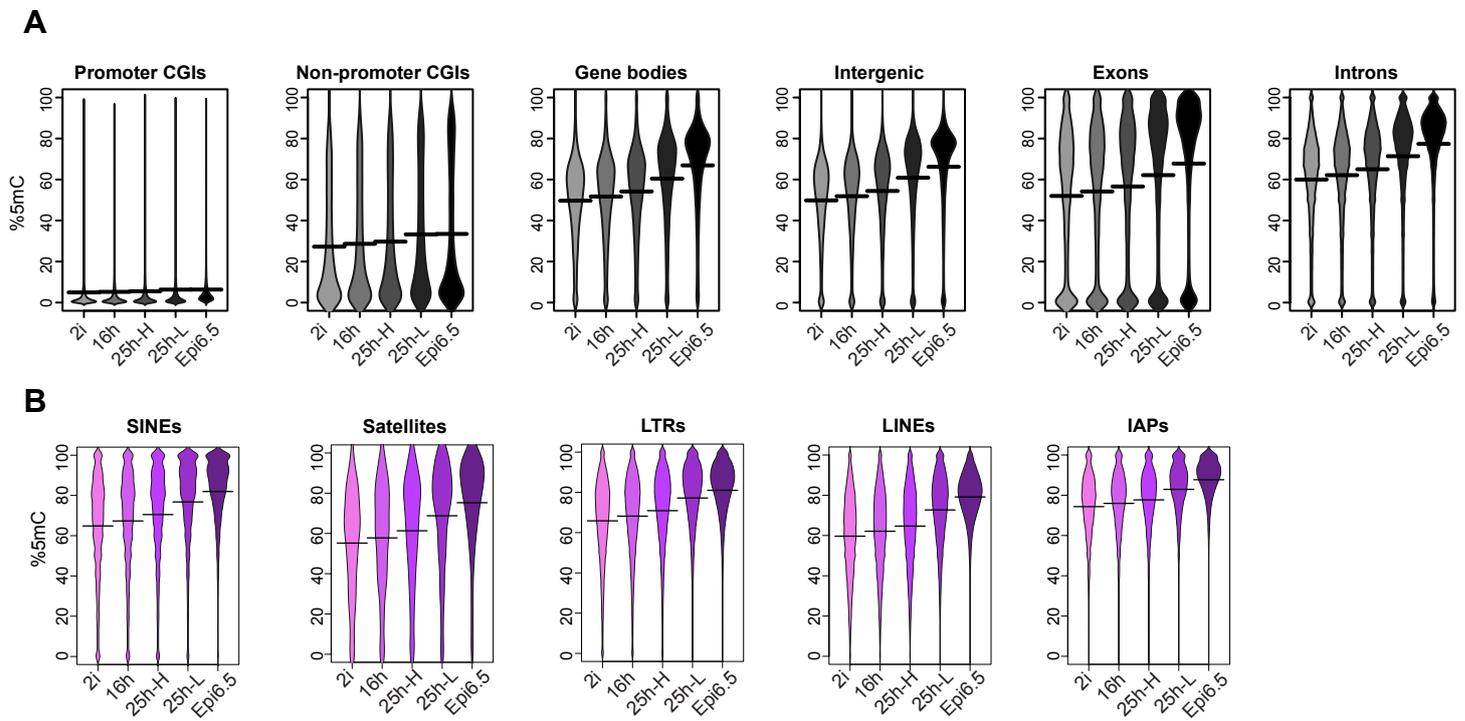
A

B

Fig. S7. Changes in DNA methylation during progression from naïve pluripotency.
Methylated cytosine levels in the CG context (mCG) **(A)** genomic features and **(B)** classes of
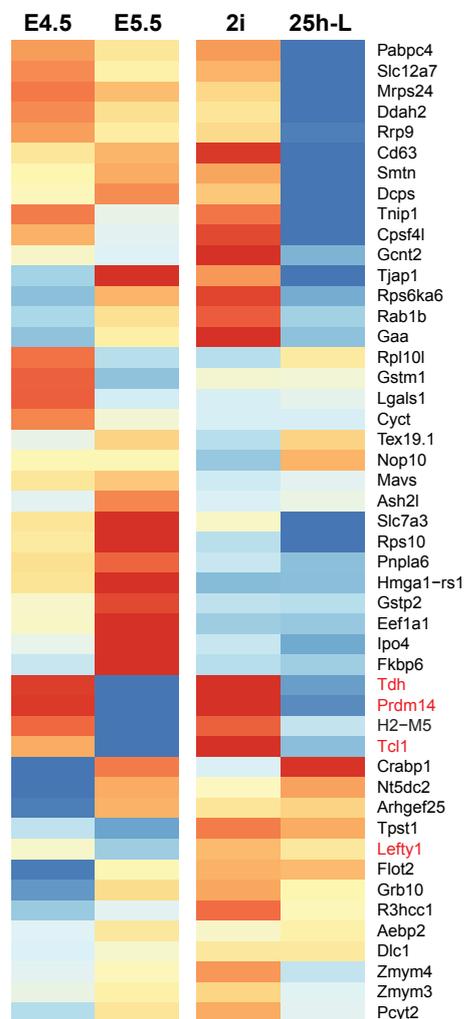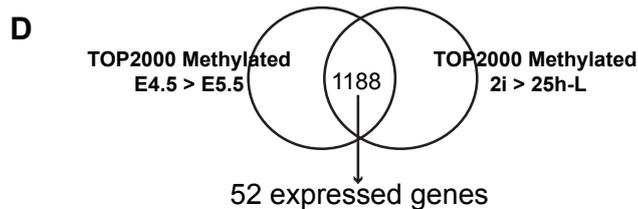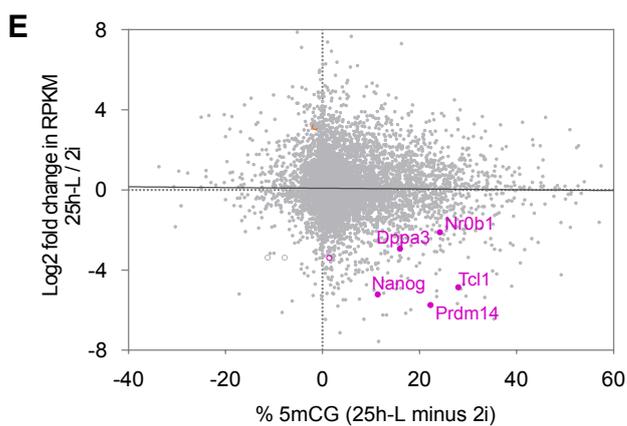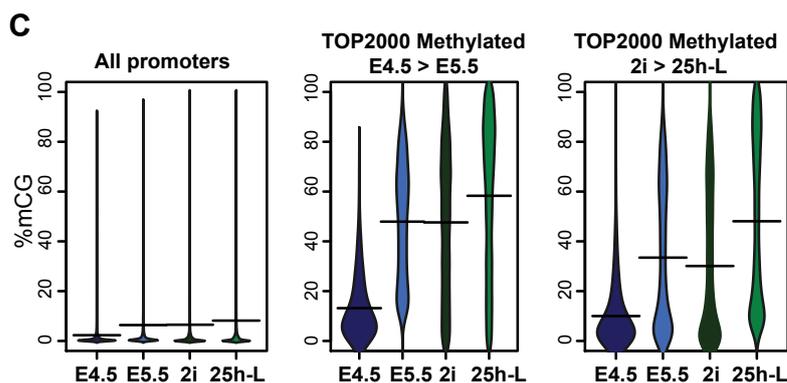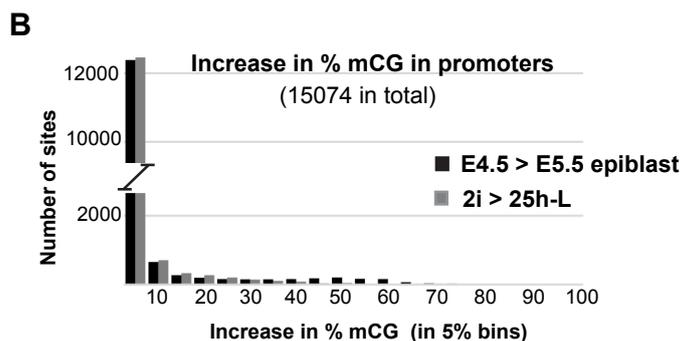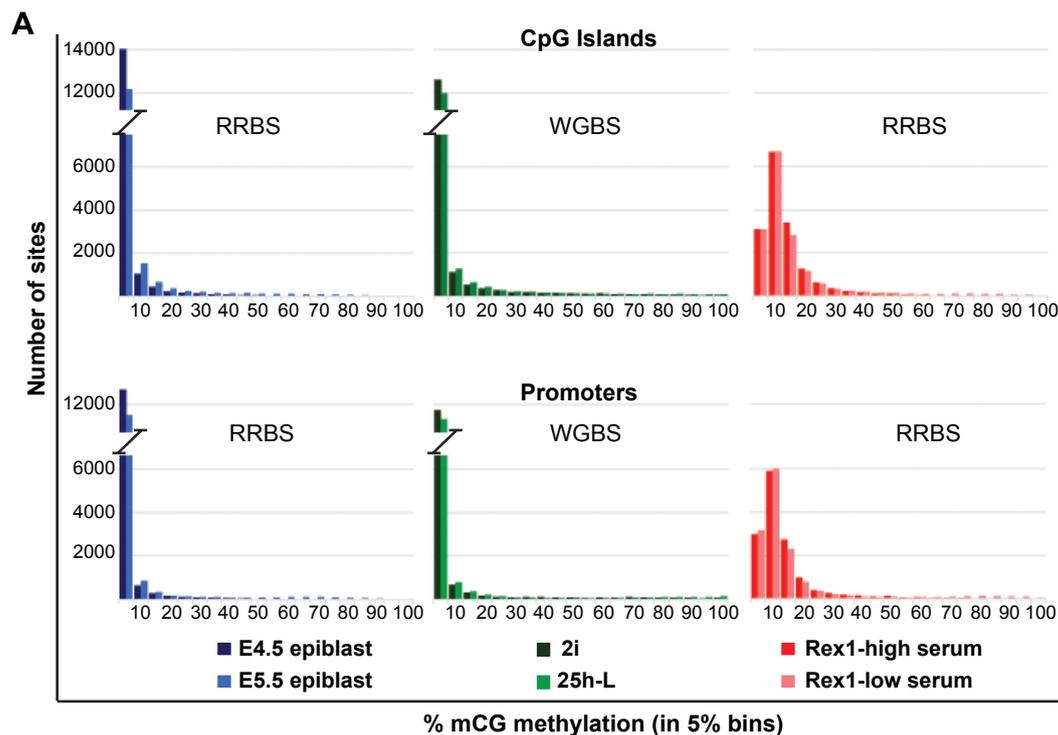DNA repeats.

Fig S8. Comparison of changes in DNA methylation during ES cell and epiblast progression

(A) Histograms of methylation levels in CpGs (mCG) across common CpG islands and promoters in epiblast samples (Auclair et al, 2014), Rex1-sorted serum ES cells (Singer et al, 2015) and 2i/25h-L populations (this study).**(B)** Histogram of the methylation increase in promoters (-1000 to +500bp of TSS) between E4.5/E5.5 epiblasts and 2i/25h-L cells. **(C)** Methylation of all promoters and the TOP2000 that exhibit highest methylation increase during the transition **(D)** Expression of genes associated with highly methylated promoters **(E)** Scatter plot of Log2 transformed fold change in mRNA levels vs difference in % promoter mCG of genes between 2i and 25h-L cells. Black line shows linear regression.

Supplementary Materials and Methods

Mouse colony establishment and immunostaining of embryos

Mice were maintained as described previously (Nichols et al., 2009a). RGd2.c6 ES cells carrying a GFPd2-IRES-Blastcidin expression cassette between the translation start and stop codons of one of the *Zfp42* (Rex1) alleles (Wray et al., 2011) were injected into E3.5 C57Bl/6 blastocysts. Offspring were assessed for chimaerism by coat colour. Three male chimaeras with a high degree of coat colour contribution were bred with wild-type 129 females. The resulting offspring that genotyped positive for the Rex1-GFPd2 allele was back-crossed to wild-type 129 animals once more. Following this, heterozygous offspring were crossed to generate homozygous mice. Homozygous mice were then bred to generate a stock of mice homozygous for RGd2 reporter. Immunostaining was performed as described previously (Nichols et al., 2009b) using antibodies listed in Table S7. Embryos were imaged on a Leica TCS SP5 confocal microscope.

Immunoblots

ES cells were lysed in 1xPBS with 1%TritonX-100, 0.1%SDS, protease and protein inhibitors (Roche) and sonicated briefly in the Bioruptor (Diagenode) to shear the gDNA. Primary antibodies and dilutions used are listed in Table S7. HRP-conjugated secondary antibodies/ECL reagent (GE Healthcare Life Sciences) were used for Fig 4G. For quantitative western blot in Fig S2A, IRDye secondary antibodies (Licor) were used and signal intensities were quantified by Odessey (Licor).

Cycloheximide (Chx) treatment

ES cells were subjected to the standard differentiation protocol (Fig 4A). 6h after medium change to 2i or N2B27, Chx (Sigma, C4859) 100µg/ml or DMSO was added.

Detection of OCR and ECAR by extracellular flux analysis

Naïve ES cells and transiting populations were dissociated and counted using Vicell. 250000 cells were plated per well of XF24 cell culture microplates (Agilent Bioscience) that were coated with 40µl of Cell Tak (88µg/ml) (Corning, 354240). Sea Horse XF Base medium was supplemented with 2mm L-Glutamine, 1mM Sodium Pyruvate and 10mm Glucose. 1h after plating cells were subjected to Mito Stress Assay using Seahorse XF^{e}24 Analyzer, according to manufacturer's protocol (Agilent Biosciences). The drug concentrations used were the following: Oligomycin; 1µM, FCCP; 1µM, Rotenone/Antimycin A; 0.5µM.

Microarray processing

RNA samples were processed for microarray hybridization according to the GeneChip whole-transcript sense target labeling assay (Affymetrix). Briefly, 2 µg of each sample was depleted of ribosomal RNA (RiboMinus, Invitrogen). Double-stranded cDNA was synthesized using random hexamers tagged with a 5′ T7 primer, and the products were amplified with T7 RNA polymerase to generate antisense cRNA. Reverse transcription was performed on the cRNA template using SuperScript III to yield ssDNA, substituting dUTPs for dTTPs, and the cRNA was subsequently degraded via RNase H digestion. cDNA products were then nicked with uracil DNA glycosylase (UDG) and apurinic/apyrimidinic endonuclease 1 (APE 1) at sites of first-strand dUTP incorporation, followed by biotin labeling with terminal deoxynucleotidyl transferase (TdT). Affymetrix Mouse Exon Array 1.0 ST arrays were hybridized for 16 h at 45°C, washed, stained with streptavidin-phycoerythrin (SAPE) conjugate on an Affymetrix fluidics station and scanned.

Microarray data analysis

Affymetrix Mouse Exon Array 1.0 ST arrays were processed in the xps system for R/Bioconductor. Background correction and quantile normalization was performed with the Robust Multi-chip Average (RMA) method (Irizarry et al., 2003) and transcripts were summarized by median polish, considering all probesets on the array remapped to Ensembl annotation. Where a gene was represented by multiple splice variants, the transcript model having the maximal value was taken as the dominant isoform. Differential expression was computed on log$_2$-transformed expression values with limma (Ritchie et al., 2015). Statistical significance was determined by an empirical Bayes moderated t-test and p-values were adjusted for multiple testing using the FDR metric (Benjamini and Hochberg, 1995). Hierarchical clustering was performed with the hclust algorithm in R using Ward's method. Clusters were extracted and ranked by time points of predominant expression. GO category and KEGG pathway enrichment analysis was applied to differentially expressed gene sets

with the GOStats (Falcon and Gentleman, 2007) and Signaling Pathway Impact Analysis (SPIA (Tarca et al., 2009)) packages for Bioconductor.

Transcriptome sequencing

RGd2-C6 and RGd2-1903.4 ES cells were subjected to RNA-seq using both mRNA-directed and ribosomal RNA depletion strategies. Two rounds of poly(A) selection (Oligotex mRNA Mini Kit, Qiagen) was applied to RGd2-C6 cells and libraries were prepared as previously described (Marks et al., 2012). 5 µg total RNA from RGd2-1903.4 cells was processed with Ribo-Zero capture probes (Illumina) and libraries were produced from 100ng of rRNA-depleted RNA using NEXTflex Directional RNA-Seq Kit V2 (Bioo Scientific) with 12 cycles of PCR amplification. Libraries were sequenced in the Illumina platform in single-end mode.

RNA-seq data analysis
Additional RNA-seq data from published studies were retrieved from the European Nucleotide Archive (ENA (Silvester et al., 2015)). EpiLC data were obtained from accession SRP040451 (Buecker et al., 2014) and EpiSC data from SRP041756 (Factor et al., 2014). Transcriptome data from early mouse embryos were obtained from ERP007120 (Boroviak et al., 2015). Sequencing reads were aligned to mouse genome build GRCm38/mm10 with STAR (Dobin et al., 2013) using the two-pass method for novel splice detection (Engstrom et al., 2013). Read alignment was guided by GENCODE M9 (Mudge and Harrow, 2015)mouse genome annotation from Ensembl release 84 (Yates et al., 2016) and splice junction donor/acceptor overlap settings were tailored to the corresponding read length of each dataset. Transcripts were quantified with htseq-count (Anders et al., 2015)based on annotation from Ensembl 84. Libraries were corrected for total read counts using size factors computed by the Bioconductor package DESeq (Anders and Huber, 2010). Principal components were computed by singular value decomposition with the prcomp function in the R stats package from variance-stabilized count data. Differential expression between EpiLC, EpiSC and ESC sample groups was assessed with DESeq.

BS-seq library preparation and methylome analysis

gDNA was isolated using Gentra Puregene Cell Kit (Qiagen). BS-seq libraries were prepared according to a previously published protocol, using NEXTflex BS-seq barcode adapters (Bioo Scientific) (Ficz et al., 2013). Paired-end 100 bp next generation sequencing was performed on an Illumina HiSeq system at the facility at the Wellcome Trust Sanger Institute.

Sequence processing and data analysis
Raw sequence reads were trimmed to remove both poor quality calls and adapters using Trim

Galore! (v0.3.5, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed sequences were quality checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and mapped to the mouse NCBI37 genome build) using Bismark in paired-end mode (v0.12.3, default parameters). CG methylation of genomic features was analysed in SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) with the integrated bisulfite analysis pipeline by averaging the individual methylation levels of CpGs, each probe covered by at least 3 CpGs or 5 CpGs in the case of repeats. Extreme outlier probes were excluded from further analysis as they likely represent mapping artefacts. Promoters were defined as the region -1 kb to +500 bp of the transcription start site, apart from promoters of selected pluripotency genes (Dppa3, Esrrb, Sall4, Zic3, Utf1 and Nr5a2), which were extracted manually at the overlap of PolII sites with H3K4me3 peaks using published datasets. CGI coordinates were obtained from Illingworth et al., 2010 and ES superenhancer coordinates from Whyte et al., 2013. Naïve mES enhancer coordinates were extracted for this study from overlapping H3K4me1, H3K27ac and p300 ChIP peaks (Buecker et al, 2014) and listed in Table S8. Repeat annotations were extracted from the UCSC RepeatMasker track (mm9 build). Exons and introns were defined with Ensembl-derived coordinates integrated in SeqMonk. E6.5 embryo dataset was used from Seisenberger et al, 2012.

Protein analysis by Mass Spectrometry

Cell culture
Heavy-SILAC-labelled ES cells (Arg6/Lys6) were obtained by culturing cells in arginine- and lysine-free DMEM/F12 (Dundee Cell Products) complemented with B27 (Gibco), in-house prepared N2, 0.1 mM 2-mercaptoethanol (Sigma), 2 mM L-glutamine, 148 mg/l heavy L-arginine, 92 mg/l heavy L-lysine (CK Gas Products) supplemented with 2i inhibitors for 3 days.

Subfractionation and protein extraction for mass spectrometry
For proteome extraction 2.4×10^7 cells from 2i, 16h, GFP-High and -Low fractions of 25h-cultures were resuspended in ice-cold fractionation buffer [0.25 M sucrose, 50 mM Tris-HCl, pH 7.9, 5 mM EDTA, 10 mM DTT, PhosSTOP Phosphatase Inhibitor Tablet (Roche), EDTA-free Protease Inhibitor Tablet (Roche)] at 1×10^7 cells ml^{-1}. Absence of cell lysis was checked using phase contrast microscopy before transfer into a pre-chilled cell disruption bomb (Parr, model 4639). Cell suspensions were incubated at 175 psi for 10 min on ice and then adiabatically decompressed via drop wise release from the vessel. Cell disruption was

assessed microscopically showing that almost all nuclei (95% - 100%) were released. Nuclei enriched fractions (S1) were obtained by centrifugation at 600 g for 10 min and snap frozen in liquid nitrogen before storage at -80°C. The remaining cell material (S2) was incubated with RIPA lysis buffer [10 mM Tris-HCl, pH 7.9, 30 mM NaCl, 5 mM EDTA, 0.2% NP-40, 0.2% sodium deoxycholate, 0.2% sodium dodecylsulfate (SDS), EDTA-free protease inhibitor tablet and PhosSTOP Phosphatase inhibitor tablet (Roche)], for 10 min on ice before spinning at 2800 g and 4°C for 10 min to pellet cell debris. The supernatants were transferred to a new tube and proteins precipitated with 4 x volumes of ice-cold 80% acetone at -20°C overnight. S1 samples were thawed on ice and membranes disrupted by addition of 2x RIPA lysis buffer in volumes of 0.5 – 1 ml. Lysed S1 fractions were sonicated and proteins precipitated as described before. Protein pellets of S1 and S2 samples were washed with ice-cold water by vortexing rigorously followed by centrifugation at either 4000 g for S2 or maximum speed of a benchtop centrifuge for S1 for 30 min. The washing step was repeated once more and proteins from both fractions resolubilised in typically 50 µl of 8M Urea containing 500 mM TEAB. Protein concentrations were determined using Pierce BCA protein assay in a 96 well plate format according to the manufacturers' instructions and revealed yields of ~ 0.5 mg per sample. At that point extracted proteomes of the sorted fractions were mixed at a 1:1 ratio with the Arg6/Lys6 labelled ESC standard.

Protein digestion and peptide separation

Proteins were reduced with DTT (20 mM final) for 35 min at room temperature followed by alkylation with IAA (40 mM final) for another 35 min at room temperature in the dark. Samples were diluted 1:10 with water (0.8 M Urea, 50 mM TEAB) and trypsin (Worthington) digestion performed at an enzyme/substrate ratio of 1:40. Trypsin was added three times: first for 1 h, then overnight before another 1 h digest the next day, all at 37°C. Samples were checked for 1:1 SILAC pair formation by mass spectrometry and snap frozen on dry ice/ethanol before lypholisation. Freeze-dried peptides were separated by high pH reverse phase chromatography using a UPLC reverse-phase column (Waters, BEH C18, 2.1 x 150 mm, 1.7 mm) on a Waters nanoACQUITY UPLC system. 20mM ammonium-formate (pH10) was used as the hydrophilic mobile phase (solvent A) and 20mM ammonium formate/80% acetonitrile was the organic mobile phase (solvent B). A gradient was developed consisting of 10 min at 100% solvent A, 50 min gradient to 70% solvent B, 7 min at 100% B, 7 min at 100% A. After the initial loading peptides fractions (20) were collected every two minutes at a flow rate of 0.244 ml/min. Eluting peptides were lyphilised and stored at -80°C.

Liquid chromatography - mass spectrometry

Lyophilised peptides were re-suspended in 100 µl of 10% formic acid (FA), vortexed and centrifuged at 13000 rpm for 5 minutes. The supernatant was diluted 10 fold and 2 µl were then taken for mass spectrometric analysis. All LC-MS/MS experiments were performed using a Dionex Ultimate 3000 RSLC nanoUPLC (Thermo Fisher Scientific Inc, Waltham, MA, USA) system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). Separation of peptides was performed by reverse-phase chromatography at a flow rate of 300 nL min^{-1} and a Thermo Scientific reverse-phase nano Easy-spray column (Thermo Scientific PepMap C18, 2 µm particle size, 100A pore size, 75 µm i.d. x 50cm length). Peptides were loaded onto a pre-column (Thermo Scientific PepMap 100 C18, 5 µm particle size, 100A pore size, 300 µm i.d. x 5mm length) from the Ultimate 3000 autosampler with 0.1% FA for 3 min at a flow rate of 10 µL/min. After this period, the column valve was switched to allow elution of peptides from the pre-column onto the analytical column. Solvent A was 0.1% FA and solvent B was 80% acetonitrile/0.1% FA. The linear gradient employed was 2-40% B in 30 min (total run time including high organic wash and re-equilibration was 60 minutes).

The LC eluant was sprayed into the mass spectrometer by means of an Easy-spray source (Thermo Fisher Scientific Inc.). All m/z values of eluting ions were measured in an Orbitrap mass analyzer, set at a resolution of 70000. Data dependent scans (Top 20) were employed to automatically isolate and generate fragment ions by higher energy collisional dissociation (HCD) in the quadrupole mass analyser and measurement of the resulting fragment ions was performed in the Orbitrap analyser, set at a resolution of 17500. Peptide ions with charge states of 2+ and above were selected for fragmentation.

References

Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome biology*. 11:R106.

Anders, S., P.T. Pyl, and W. Huber. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31:166-169.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15-21.

Engstrom, P.G., T. Steijger, B. Sipos, G.R. Grant, A. Kahles, G. Ratsch, N. Goldman, T.J. Hubbard, J. Harrow, R. Guigo, P. Bertone, and R. Consortium. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature methods*. 10:1185-1191.

Factor, D.C., O. Corradin, G.E. Zentner, A. Saiakhova, L. Song, J.G. Chenoweth, R.D. McKay, G.E. Crawford, P.C. Scacheri, and P.J. Tesar. 2014. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. *Cell stem cell*. 14:854-863.

Falcon, S., and R. Gentleman. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics*. 23:257-258.

Irizarry, R.A., B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, and T.P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 4:249-264.

Marks, H., T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A.F. Stewart, A. Smith, and H.G. Stunnenberg. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. 149:590-604.

Mudge, J.M., and J. Harrow. 2015. Creating reference gene annotation for the mouse C57BL6/J genome assembly. *Mammalian genome : official journal of the International Mammalian Genome Society*. 26:366-378.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 43:e47.

Silvester, N., B. Alako, C. Amid, A. Cerdeno-Tarraga, I. Cleland, R. Gibson, N. Goodgame, P. Ten Hoopen, S. Kay, R. Leinonen, W. Li, X. Liu, R. Lopez, N. Pakseresht, S. Pallreddy, S. Plaister, R. Radhakrishnan, M. Rossello, A. Senf, D. Smirnov, A.L. Toribio, D. Vaughan, V. Zalunin, and G. Cochrane. 2015. Content discovery and retrieval services at the European Nucleotide Archive. *Nucleic acids research*. 43:D23-29.

Tarca, A.L., S. Draghici, P. Khatri, S.S. Hassan, P. Mittal, J.S. Kim, C.J. Kim, J.P. Kusanovic, and R. Romero. 2009. A novel signaling pathway impact analysis. *Bioinformatics*. 25:75-82.

Yates, A., W. Akanni, M.R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, L. Gil, C.G. Giron, L. Gordon, T. Hourlier, S.E. Hunt, S.H. Janacek, N. Johnson, T. Juettemann, S. Keenan, I. Lavidas, F.J. Martin, T. Maurel, W. McLaren, D.N. Murphy, R. Nag, M. Nuhn, A. Parker, M. Patricio, M. Pignatelli, M. Rahtz, H.S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S.P. Wilder, A. Zadissa, E. Birney, J. Harrow, M. Muffato, E. Perry, M. Ruffier, G. Spudich, S.J. Trevanion, F. Cunningham, B.L. Aken, D.R. Zerbino, and P. Flicek. 2016. Ensembl 2016. *Nucleic acids research*. 44:D710-716.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol,* 11**,** R106.

ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics,* 31**,** 166-9.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

ENGSTROM, P. G., STEIJGER, T., SIPOS, B., GRANT, G. R., KAHLES, A., RATSCH, G., GOLDMAN, N., HUBBARD, T. J., HARROW, J., GUIGO, R., BERTONE, P. & CONSORTIUM, R. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods,* 10**,** 1185-91.

FACTOR, D. C., CORRADIN, O., ZENTNER, G. E., SAIAKHOVA, A., SONG, L., CHENOWETH, J. G., MCKAY, R. D., CRAWFORD, G. E., SCACHERI, P. C. & TESAR, P. J. 2014. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. *Cell Stem Cell,* 14**,** 854-63.

FALCON, S. & GENTLEMAN, R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics,* 23**,** 257-8.

IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U. & SPEED, T. P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-64.

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**,** 590-604.

MUDGE, J. M. & HARROW, J. 2015. Creating reference gene annotation for the mouse C57BL6/J genome assembly. *Mamm Genome,* 26**,** 366-78.

NICHOLS, J., JONES, K., PHILLIPS, J. M., NEWLAND, S. A., ROODE, M., MANSFIELD, W., SMITH, A. & COOKE, A. 2009a. Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. *Nat Med,* 15**,** 814-8.

NICHOLS, J., SILVA, J., ROODE, M. & SMITH, A. 2009b. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development,* 136**,** 3215-22.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**,** e47.

SILVESTER, N., ALAKO, B., AMID, C., CERDENO-TARRAGA, A., CLELAND, I., GIBSON, R., GOODGAME, N., TEN HOOPEN, P., KAY, S., LEINONEN, R., LI, W., LIU, X., LOPEZ, R., PAKSERESHT, N., PALLREDDY, S., PLAISTER, S., RADHAKRISHNAN, R., ROSSELLO, M., SENF, A., SMIRNOV, D., TORIBIO, A. L., VAUGHAN, D., ZALUNIN, V. & COCHRANE, G. 2015. Content discovery and retrieval services at the European Nucleotide Archive. *Nucleic Acids Res,* 43**,** D23-9.

TARCA, A. L., DRAGHICI, S., KHATRI, P., HASSAN, S. S., MITTAL, P., KIM, J. S., KIM, C. J., KUSANOVIC, J. P. & ROMERO, R. 2009. A novel signaling pathway impact analysis. *Bioinformatics,* 25**,** 75-82.

WRAY, J., KALKAN, T., GOMEZ-LOPEZ, S., ECKARDT, D., COOK, A., KEMLER, R. & SMITH, A. 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol,* 13**,** 838-45.

YATES, A., AKANNI, W., AMODE, M. R., BARRELL, D., BILLIS, K., CARVALHO-SILVA, D., CUMMINS, C., CLAPHAM, P., FITZGERALD, S., GIL, L., GIRON, C. G., GORDON, L., HOURLIER, T., HUNT, S. E., JANACEK, S. H., JOHNSON, N., JUETTEMANN, T., KEENAN, S., LAVIDAS, I., MARTIN, F. J., MAUREL, T., MCLAREN, W., MURPHY, D. N., NAG, R., NUHN, M., PARKER, A., PATRICIO, M., PIGNATELLI, M., RAHTZ, M., RIAT, H. S., SHEPPARD, D., TAYLOR, K., THORMANN, A., VULLO, A., WILDER, S. P., ZADISSA, A., BIRNEY, E., HARROW, J., MUFFATO, M., PERRY, E., RUFFIER, M., SPUDICH, G., TREVANION, S. J., CUNNINGHAM, F., AKEN, B. L., ZERBINO, D. R. & FLICEK, P. 2016. Ensembl 2016. *Nucleic Acids Res,* 44**,** D710-6.

Table S1. Outcome of RGd2 heterozygous crosses

Genotype	No. of animals	% of total	% expected
WT	34	29.3	25
Het	62	53.4	50
Homo	20	17.2	25

Table S2. Derivation of ES cell lines from ICMs obtained from heterozygous crosses

No. of ICMs plated	No. of ES cell lines established	Sex of ES cell lines carrying RGd2 allele
28	Wt (4) Het (11) Homo (5)	Het (7M, 4F) Homo (2M, 3F)

Table S3. Differentially expressed genes in sorted Rex1-subpopulations identified by exon microarray profiling

Click here to Download Table S3

Table S4. RPKM counts of genes in sorted Rex1-subpopulations measured by RNA-seq.

Two datasets obtained from poly-A enriched (poly-A) and ribosome-depleted (total) RNA
are presented. (RPKM=Reads Per Kilobase of transcript per Million mapped reads).

Click here to Download Table S4

**Table S5. Comparative analysis of differentially expressed genes during in vivo and in vitro
transition from naïve pluripotency** (related to Fig 6)

Differentially expressed genes between E4.5 and E5.5 epiblast samples were identified and
filtered based on expression level (FPKM ≥ 10) (Boroviak, et al 2015). This set was
overlapped with differentially expressed genes between 2i and 25h-L cells.

Click here to Download Table S5

Table S6. Differentially expressed genes between EpiLCs and 25h-L cells.

RNA-seq dataset of EpiLCs from Buecker et al 2014 was used in the analysis.

Click here to Download Table S6

Table S7. qRT-PCR reagents, PCR primers and antibodies used in this study

Click here to Download Table S7

Table S8. Genomic coordinates of naïve enhancers used in Fig 7.

Regions co-occupied by H3K27Ac and p300 in 2i/L ESCs (Buecker et al, 2014) were designated as naïve enhancers.

Click here to Download Table S8

Table S9. Levels of CpG methylation in epiblast samples and ESC subpopulations from different culture conditions

Percentage of mCG in the promoters (-1000 to +500 of TSS) or CGIs in E4.5 and E5.5 epiblast (Auclair et al 2014), Rex1-sorted ESC subpopulations from serum cultures (Singer et al 2014) and 2i/25h-L cells (this study).

Click here to Download Table S9

Table S10. Changes in promoter methylation during in vitro and in vivo transition from naïve pluripotency.

2000 promoters that exhibit highest methylation gain during the transition were identified based on the difference of percentage mCG between E5.5 and E4.5 epiblast samples and between 25h-L and 2i cells. Expression levels from RNA-seq of corresponding genes are also presented.

Click here to Download Table S10