Supplementary Figures

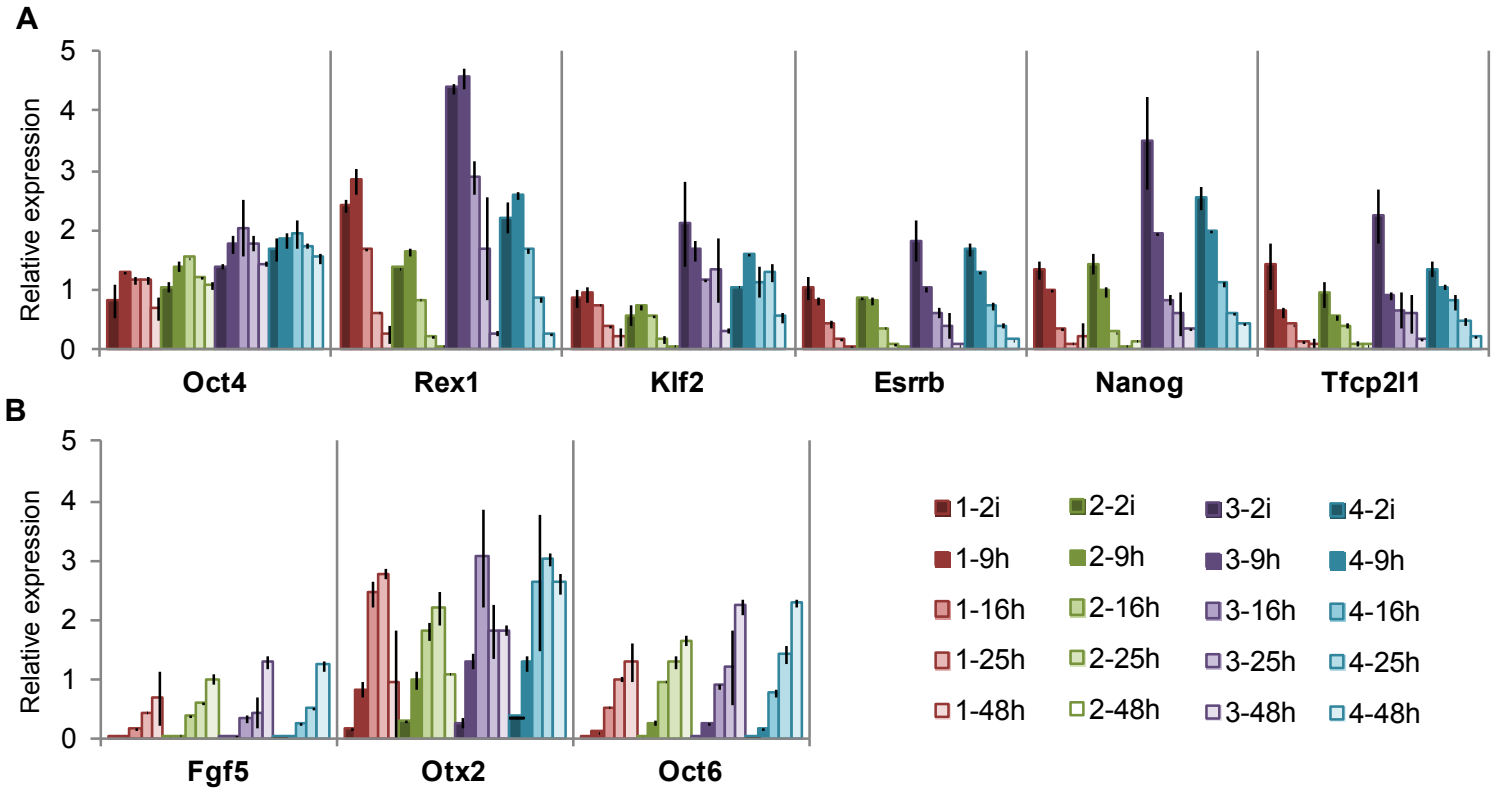
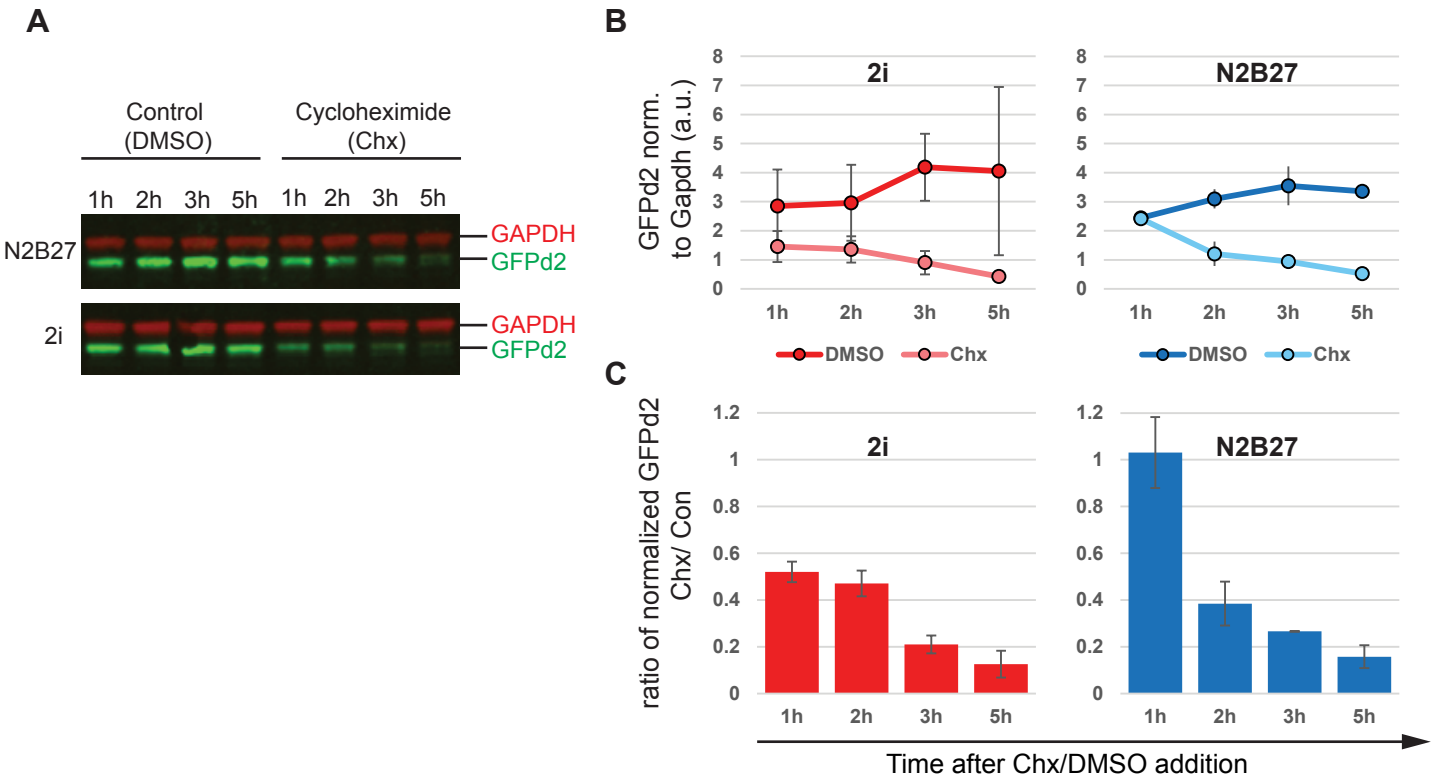
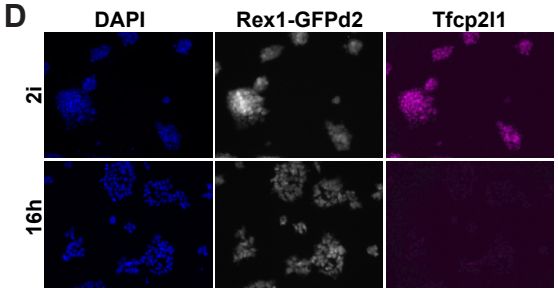
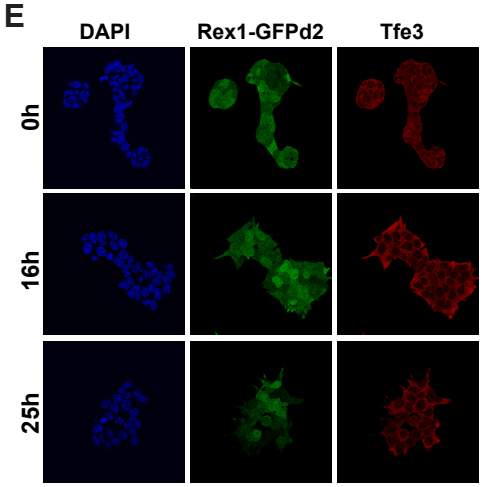


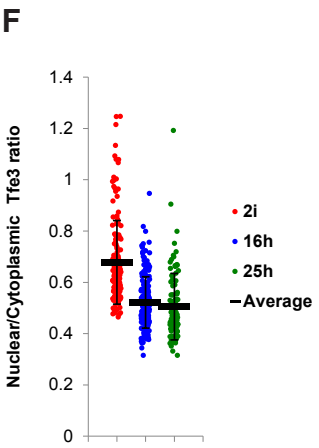
Fig. S1. Gene expression changes in ES cells upon release from 2i

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.









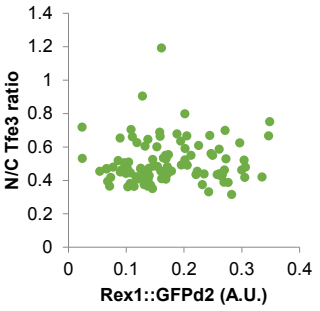
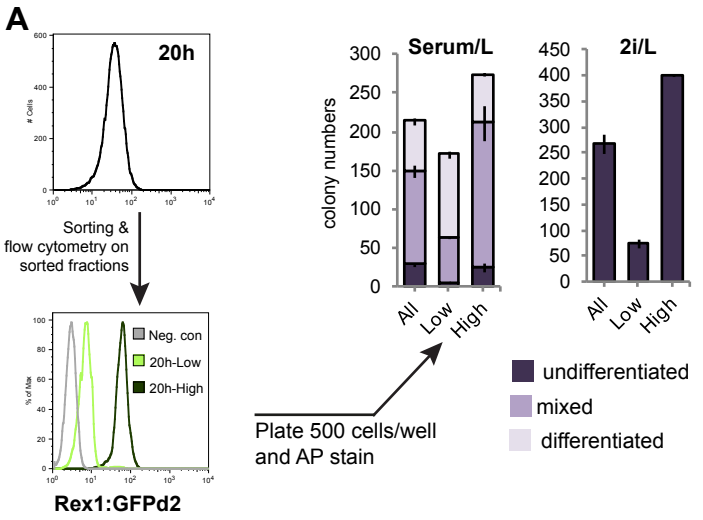
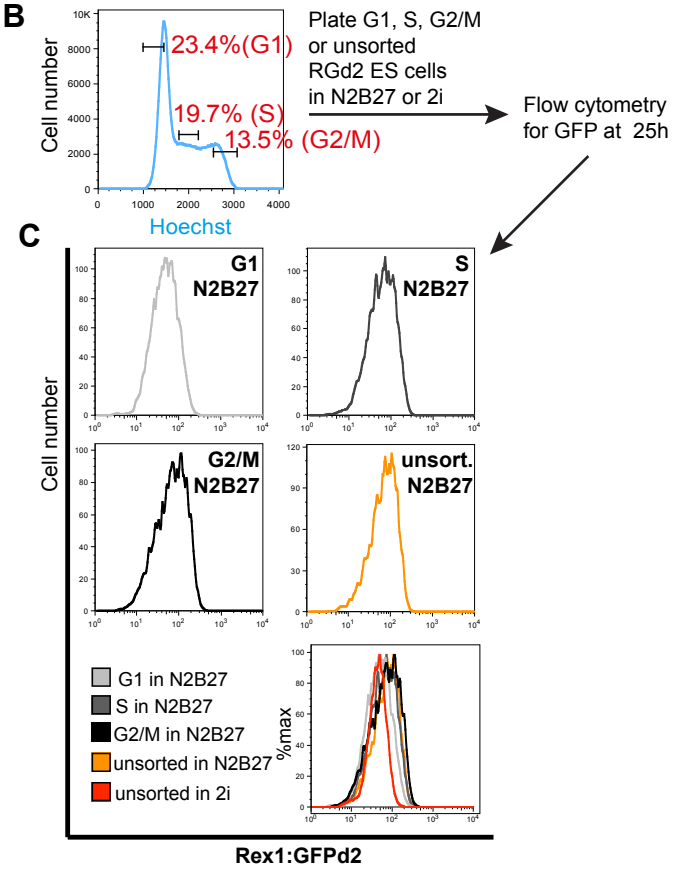
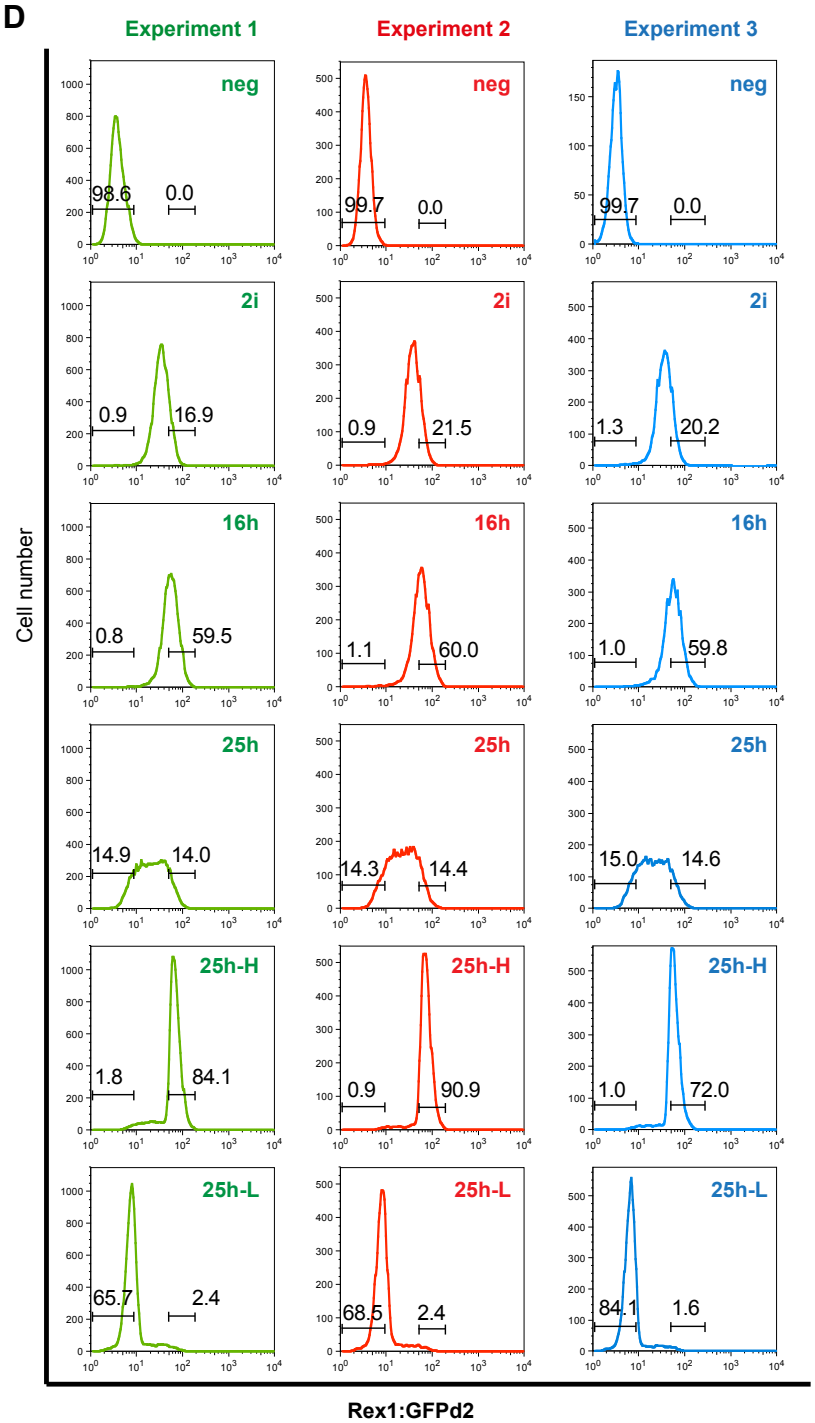


Fig. S2. Half-life of GFPd2 and IF staining for Tfcp2l1 and Tfe3

(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.







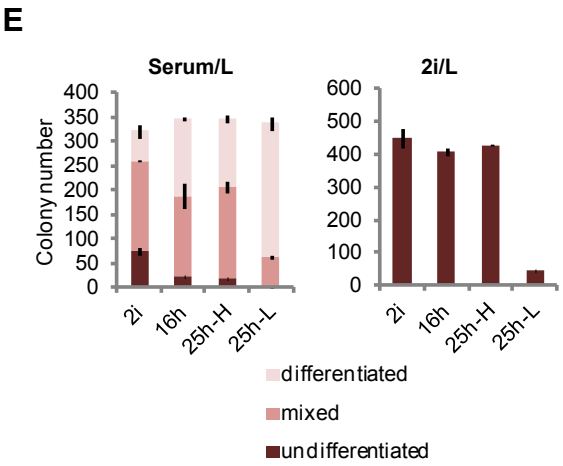


Fig. S3. Downregulation of Rex1 tracks exit from the naive state (related to Fig. 4.) **(A)** GFP profiles and clonogenicity of sorted Rex1-GFPd2-High/Low subpopulations and total population (All) from 20h cultures. GFP profiles of sorted subpopulations are shown in the lower histogram. A wild type ES cell line was used as negative control (Neg). Lines on the bars represent standard deviation (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) .Lines on the bars represent standard deviation from 2 biological replicates each with 2 technical replicates.

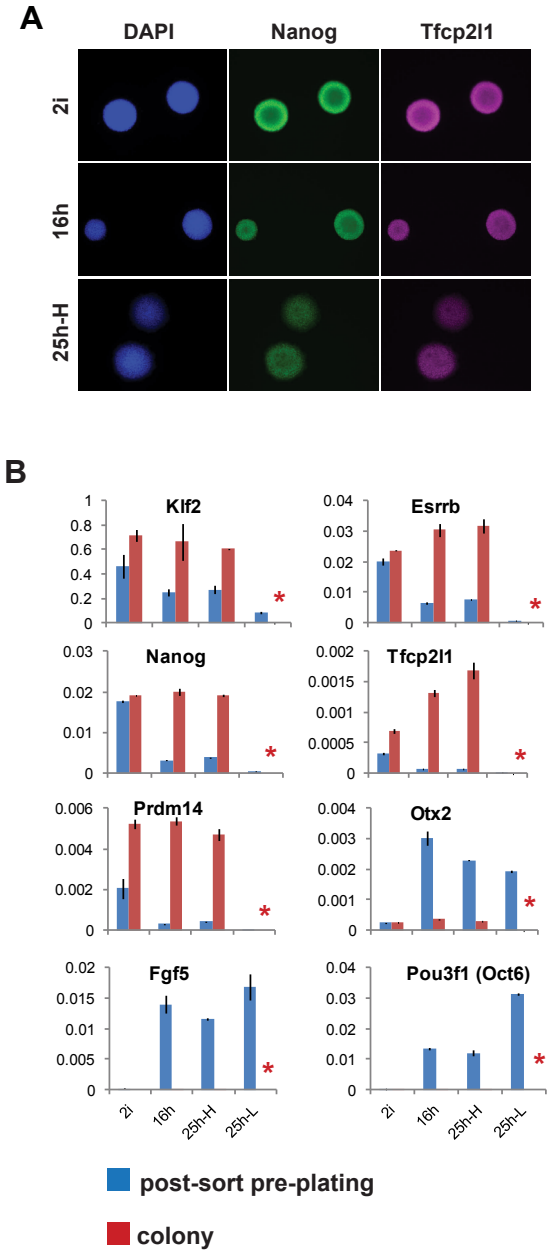


Fig S4. Re-establishment of naive state gene expression in reverted cells (Related to Fig 4)

(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.

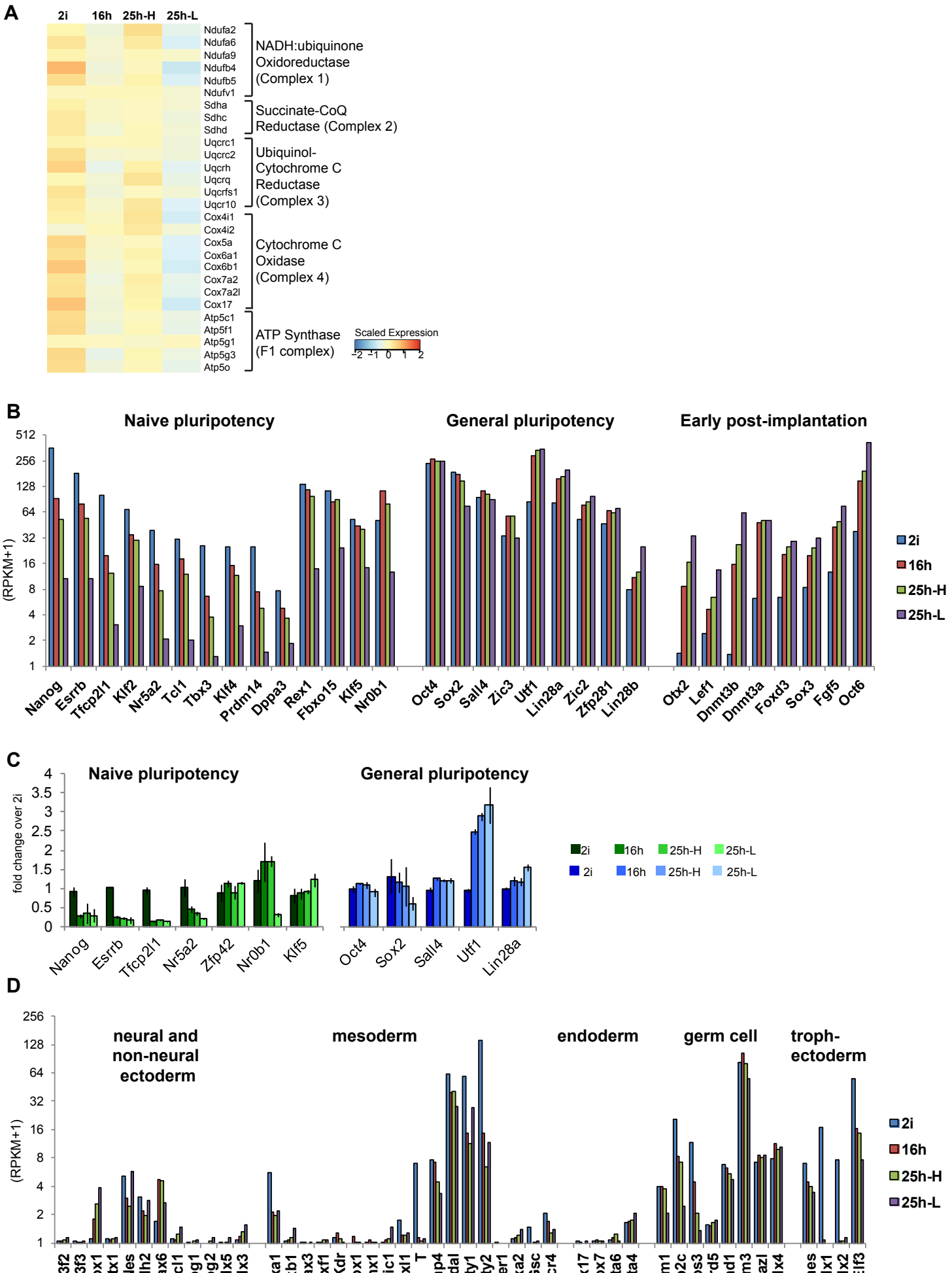
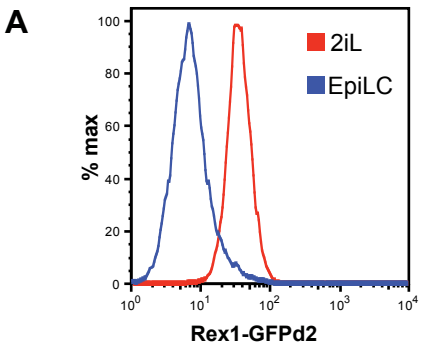
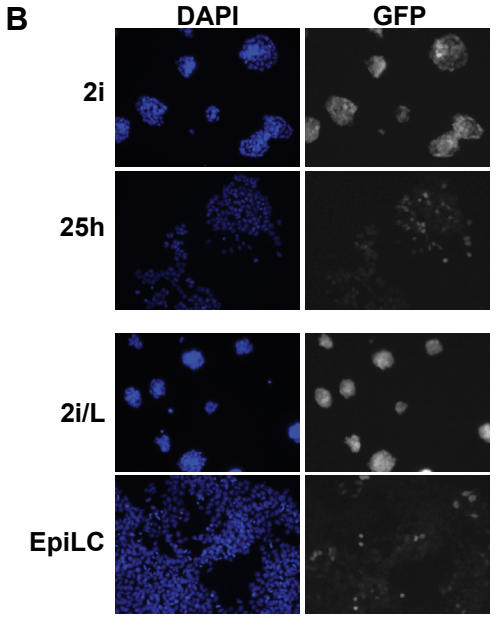
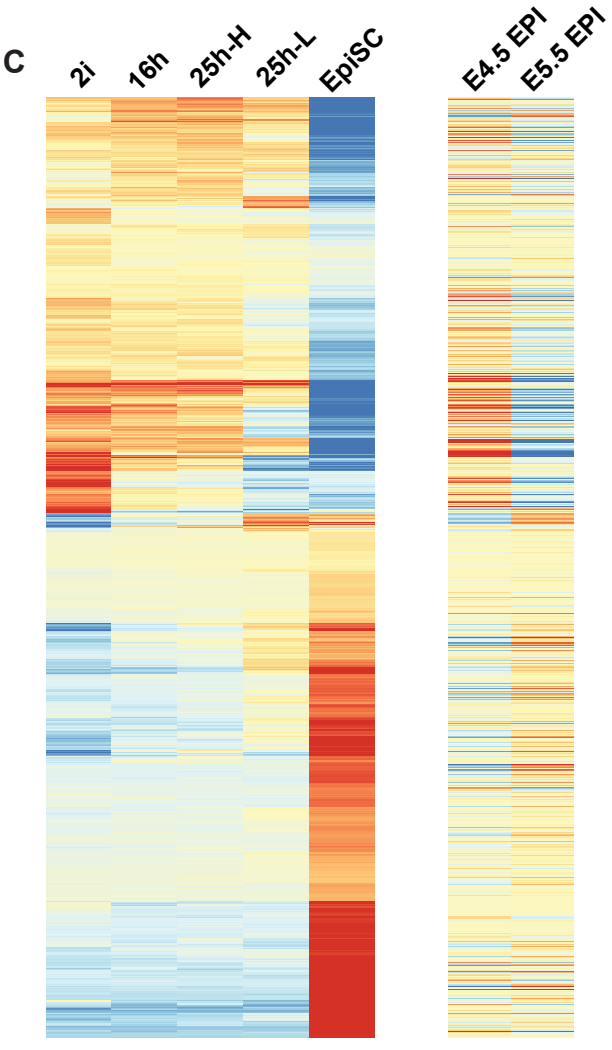


Fig. S5. Changes in mRNA and protein levels of selected genes during the transition (related to Fig. 5)

(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.







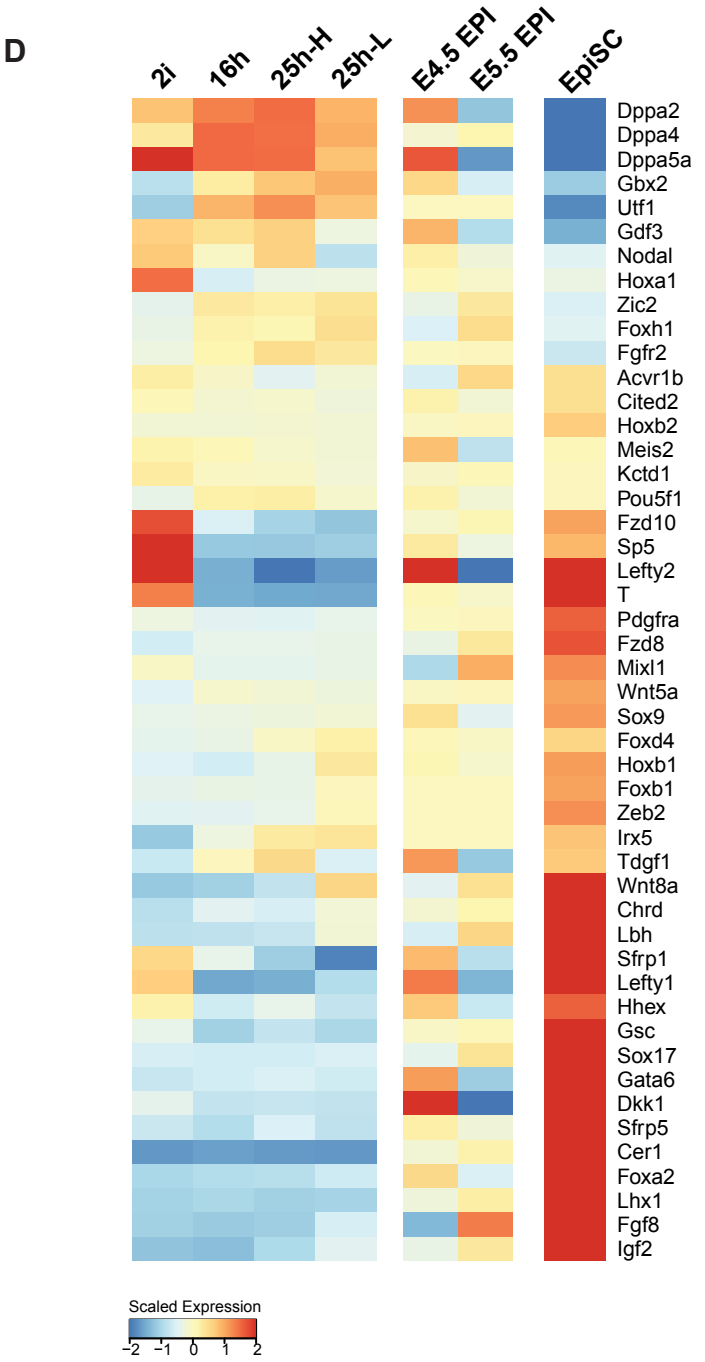


Fig. S6. Comparison of Rex1 populations with EpiLCs and EpiSCs

(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).

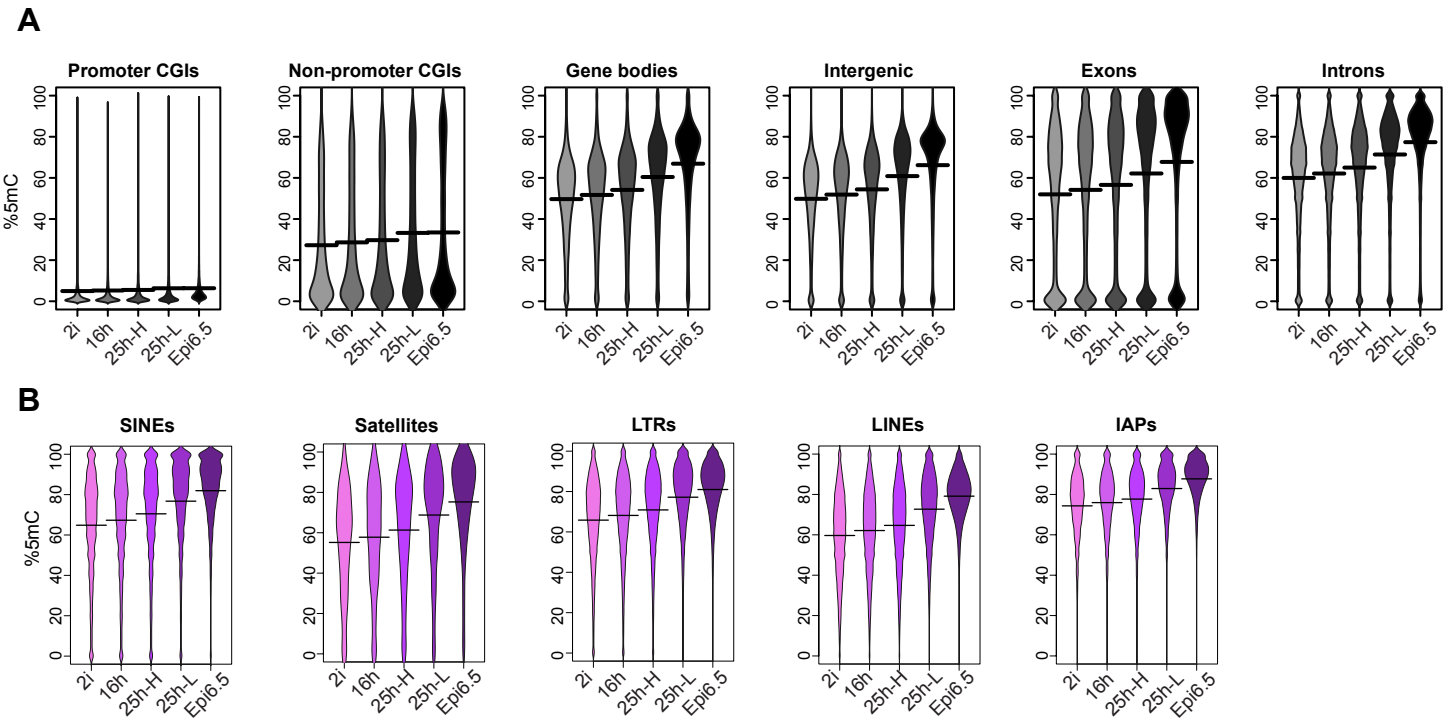
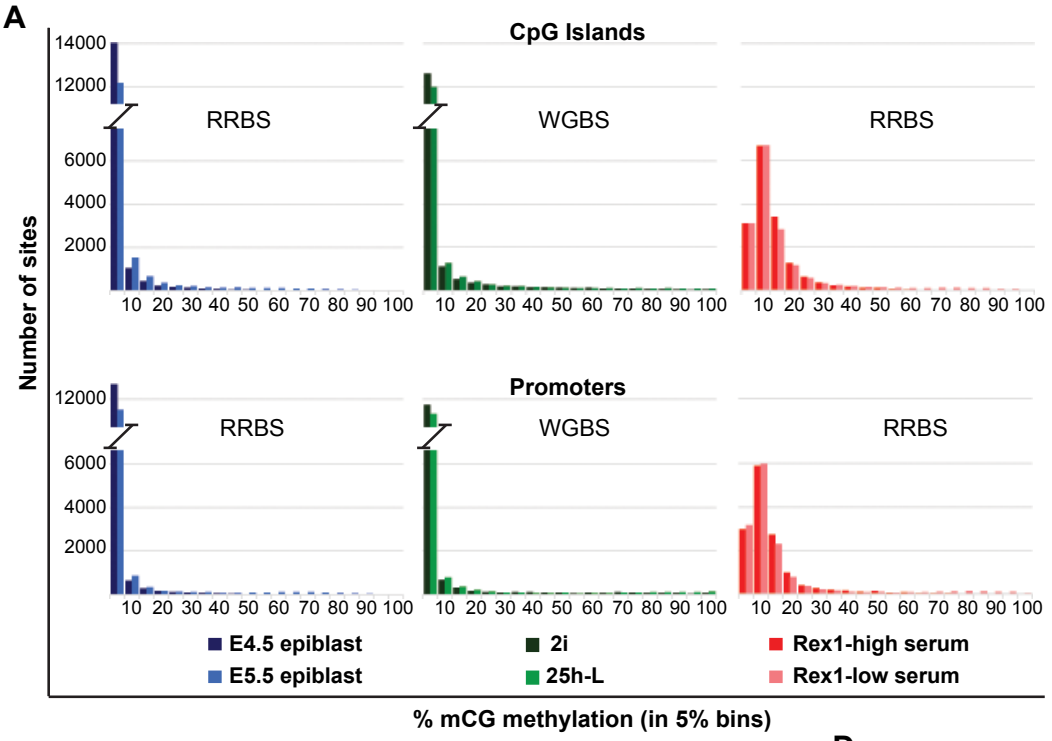
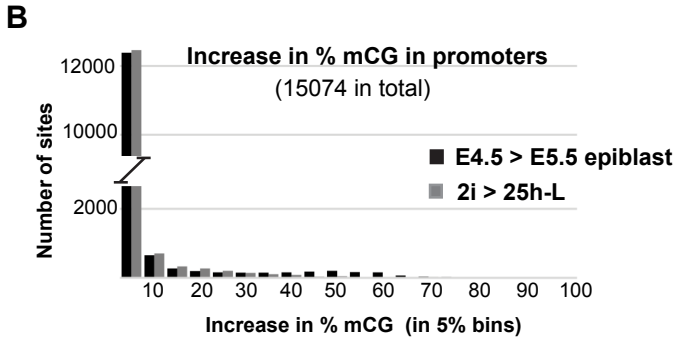
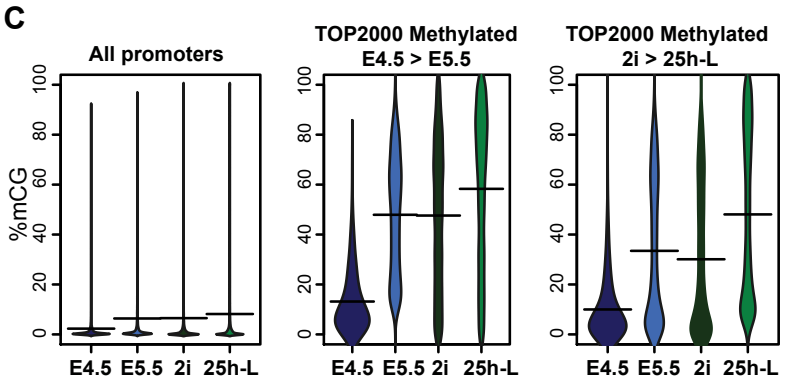
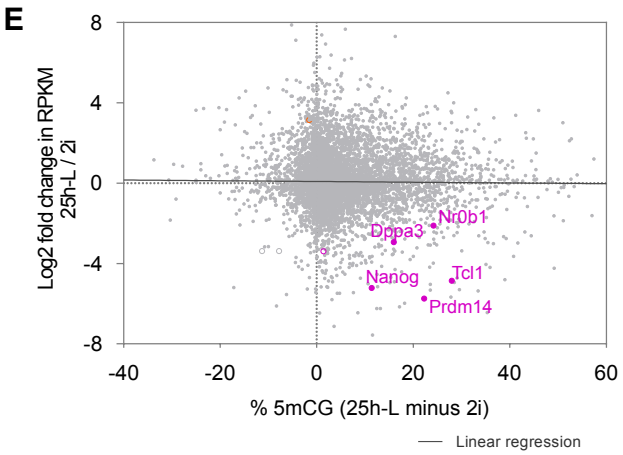


Fig. S7. Changes in DNA methylation during progression from naïve pluripotency. Methylated cytosine levels in the CG context (mCG) **(A)** genomic features and **(B)** classes of DNA repeats.









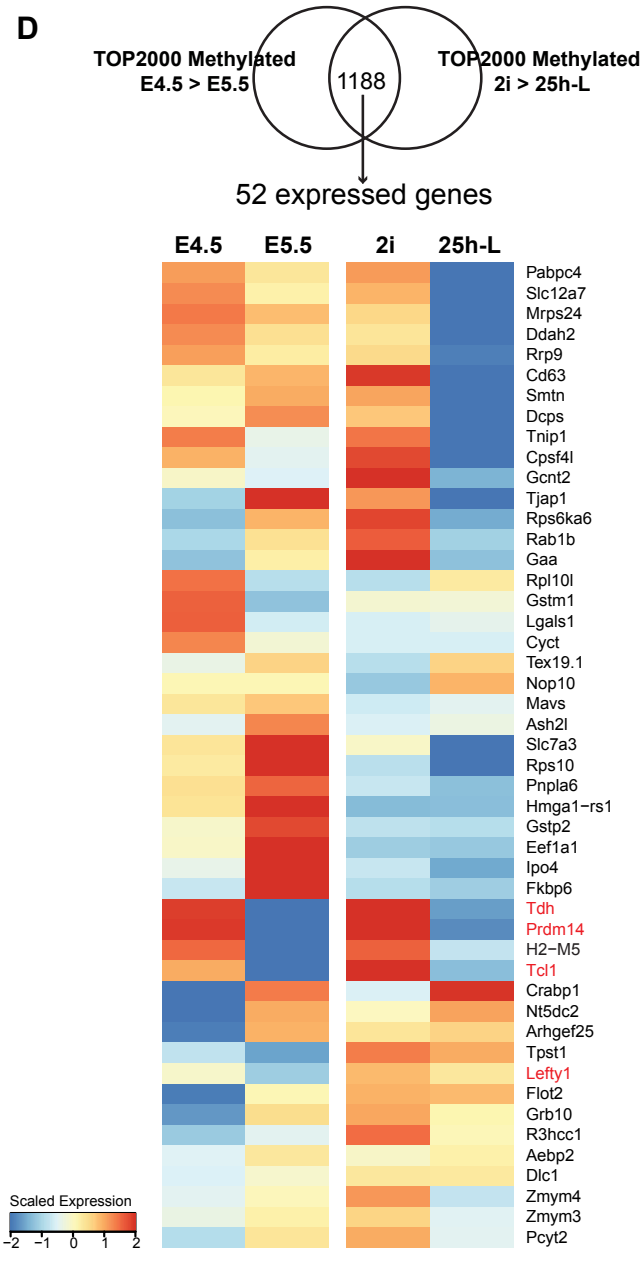


Fig S8. Comparison of changes in DNA methylation during ES cell and epiblast progression

(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 DNA 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 top 2000 genes (TOP2000) that exhibit highest DNA methylation increase in their promoters during the transition **(D)** Expression profile of genes associated with highly methylated promoters **(E)** Scatter plot of Log2 transformed fold change in mRNA levels versus difference in % mCG in the associated promoters between 2i and 25h-L cells. Black line shows linear regression.

Supplementary tables

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 H3K4me1, 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

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).

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 Technologies) 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 Technologies). The drug concentrations used in the assay 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. Heatmaps display RPKM values scaled to the mean expression of each gene over all samples.

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 passages.

Subfractionation and protein extraction for mass spectrometry

For proteome extraction $2.4x10^7$ 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 $1x10^7$ 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.