

## Research Article



# Comparative functional genomics identifies unique molecular features of EPSCs

Vikas Malik<sup>1,\*</sup>, Ruge Zang<sup>1,\*</sup>, Alejandro Fuentes-Iglesias<sup>2</sup>, Xin Huang<sup>1</sup>, Dan Li<sup>1</sup>, Miguel Fidalgo<sup>2</sup>, Hongwei Zhou<sup>1</sup>, Jianlong Wang<sup>1</sup>

**Extended pluripotent or expanded potential stem cells (EPSCs) possess superior developmental potential to embryonic stem cells (ESCs). However, the molecular underpinning of EPSC maintenance in vitro is not well defined. We comparatively studied transcriptome, chromatin accessibility, active histone modification marks, and relative proteomes of ESCs and the two well-established EPSC lines to probe the molecular foundation underlying EPSC developmental potential. Despite some overlapping transcriptomic and chromatin accessibility features, we defined sets of molecular signatures that distinguish EPSCs from ESCs in transcriptional and translational regulation as well as metabolic control. Interestingly, EPSCs show similar reliance on pluripotency factors Oct4, Sox2, and Nanog for self-renewal as ESCs. Our study provides a rich resource for dissecting the regulatory network that governs the developmental potency of EPSCs and exploring alternative strategies to capture totipotent stem cells in culture.**

DOI: 10.26508/lsa.202201608 | Received 15 July 2022 | Revised 19 July 2022 | Accepted 20 July 2022 | Published online 12 August 2022

## Introduction

Totipotent cells can give rise to the whole conceptus, including embryonic and extraembryonic tissues, whereas pluripotent cells can only give rise to embryonic tissues. Thus, totipotent cells have a superior developmental potential over pluripotent cells. However, despite the derivation of mouse embryonic stem cells (ESCs) under stable culture with maintained pluripotency more than four decades ago (Evans & Kaufman, 1981; Martin, 1981), the capture of totipotent cells in stable culture is challenging. The pioneering efforts from the Liu and Deng groups led to the in vitro stabilization of expanded potential stem cells ("L-EPSCs" hereafter) (Yang et al, 2017a) and extended pluripotent stem cells ("D-EPSCs" hereafter) (Yang et al, 2017b), collectively known as EPSCs, resembling the earlier cleavage stages of embryonic development.

EPSCs have been derived from various sources, including four- or eight-cell mouse embryos, human fibroblast-derived induced pluripotent stem cells (iPSCs), mouse and human ESCs (Yang et al, 2017a, 2017b), pig (Gao et al, 2019), and bovine (Zhao et al, 2021) early blastocysts. These cells can be further adapted to grow in feeder-free (Zheng et al, 2021) and xeno-free (Liu et al, 2021) conditions, thus opening new avenues for their molecular dissection and clinical applications. Compared with ESCs, EPSCs display superior developmental potential as they can generate both embryonic and extraembryonic tissues, including yolk sac and placenta (Yang et al, 2017a, 2017b). Furthermore, EPSCs can directly give rise to ESCs, trophoblast stem cells (TSCs), and extra-embryonic endoderm (XEN) cells using proper defined culture conditions (Yang et al, 2017a).

Although the developmental potential of EPSCs relative to their in vivo counterparts is still challenged (Posfai et al, 2021b), these cells have been nonetheless proven valuable with multiple benefits over ESCs. For example, EPSCs show superior directed differentiation potential to generate functional hepatocytes transcriptionally closer to the primary human hepatocytes than hESC-derived counterparts (Wang et al, 2020). Compared with ESCs, EPSCs can be easily derived from a non-permissive humanized mouse model (NOD-SCID IL2rg<sup>-/-</sup> strain [Du et al, 2019]). Attributing to its higher proliferation rate and better genetic and epigenetic stability, a single EPSC can give rise to an entire mouse via tetraploid complementation (Yang et al, 2017b). In addition, mouse EPSCs have superior developmental potency and robust germline competence compared with conventional ESCs, allowing for the rapid generation of gene-targeted EPSC-derived mouse models through tetraploid complementation (Li et al, 2019a). EPSCs of both mouse and human origins outperform pluripotent stem cells (PSCs) in interspecies chimera in mouse conceptuses (Yang et al, 2017b; Gao et al, 2019) and monkey embryos cultured ex vivo (Tan et al, 2021). Similarly, EPSCs combined with TSCs (Sozen et al, 2019) or EPSCs alone (even a single EPSC) could form blastoids, blastocyst-like structures that can further develop to post-implantation embryo structure and

<sup>1</sup>Department of Medicine, Columbia Center for Human Development, Columbia Stem Cell Initiative, Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA. <sup>2</sup>Department of Physiology, Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), Universidade de Santiago de Compostela (USC)-Health Research Institute (IDIS), Santiago de Compostela, Spain

Correspondence: [jw3925@cumc.columbia.edu](mailto:jw3925@cumc.columbia.edu)

\*Vikas Malik and Ruge Zang contributed equally to this work.



induce decidualization in vitro, although they fail to generate live pups (Li et al, 2019b). Therefore, further studies to understand and improve their developmental potency are much needed.

Previous studies reported the epigenome and single-cell transcriptome of EPSCs (Yang et al, 2017a, 2017b; Posfai et al, 2021b); however, a precise molecular makeup of EPSCs is still missing. Here, we reprogrammed mouse ESCs to D-EPSCs and L-EPSCs using the respective protocols (Yang et al, 2017a, 2017b) and systematically mapped transcriptome, chromatin accessibility, active enhancer and promoter marks, and proteomes of D/L-EPSCs relative to ESCs. We found that, despite similar reliance on key pluripotency factors Oct4, Sox2, and Nanog for their maintenance, EPSCs differ from ESCs in molecular features including expression of other pluripotency-associated (e.g., *Lin28a*, *Utf1*, *Esrrb*, *Nr5a2*, and *Myc*), DNA methylation (e.g., *Dnmt3a/b/l* and *Mettl4*), and gastrulation (e.g., *Eomes*, *Dusp4*, *Bmp4*, and *Left1*) related genes. We also uncovered differentially open chromatin genomic loci harboring DNA motifs of RAR-RXR and Zfp281 in L- and D-EPSCs, respectively. In addition, our proteomics data revealed the differences in specific translational and metabolic regulation in ESCs, D-EPSCs, and L-EPSCs. Together, our study provides a rich resource for further dissecting the regulatory network governing the unique developmental potential of EPSCs.

## Results

### Transcriptome comparison reveals discernible gene expression changes between EPSCs and ESCs

To dissect the molecular features of EPSCs, we first converted ESCs (cultured in 2i/leukemia inhibitory factor [LIF] medium) to D-EPSCs and L-EPSCs (Fig 1A) following previously published protocols (Yang et al, 2017a, 2017b). We observed compact EPSC colonies with smooth edges with and without a feeder layer (Fig 1B), with L-EPSCs forming slightly flat colonies without feeder layers, consistent with a previous study (Posfai et al, 2021b). Next, we performed bulk RNA-seq to examine how a transcriptome shift is induced after switching ESCs into either EPSC conditions. Replicates correlated well, and the transcriptomic profiles of both EPSC lines were closer to each other than to the ESCs (Fig 1C), and principal component analysis (PCA) showed global gene expression variability in the three cell states (Fig 1D). Differential gene expression analysis revealed that the transcriptomes of ESCs show much larger gene expression differences with D-EPSCs (1,875 up-regulated and 2,024 down-regulated genes) and L-EPSCs (2,128 up-regulated and 1,619 down-regulated genes) than those between L-EPSCs and D-EPSCs (836 up-regulated in L-EPSCs and 1,573 up-regulated in D-EPSCs) (Fig 1E and Table S1), consistent with the correlation heat map (Fig 1C). The expression levels of Oct4 and Sox2 in both EPSCs resembled those in ESCs at mRNA (Figs 1E and F and S1A) and protein levels (Fig 1G), although Nanog showed a slightly lower mRNA level in EPSCs relative to ESCs and yet similar protein levels in EPSCs and ESCs (Fig 1E–G). EPSCs also showed reduced expression of a few pluripotency genes, including *Nr5a2* and *Esrrb*, while overexpressing other pluripotency-

associated genes such as *Utf1*, *Lin28a*, *Dnmt3l*, *Zic3*, and *Myc* (Figs 1F and S1A). Interestingly, whereas most early totipotent two-cell specific genes (i.e., 2C markers) do not express in ESCs or EPSCs, some of them, including *Zscan4c/d/f* and *Usp17le*, albeit lowly expressed, express at slightly higher levels, especially in L-EPSCs (Figs 1F and S1A), with an enrichment of H3K27ac active histone modification mark near their promoter regions (Fig S1B) than ESCs.

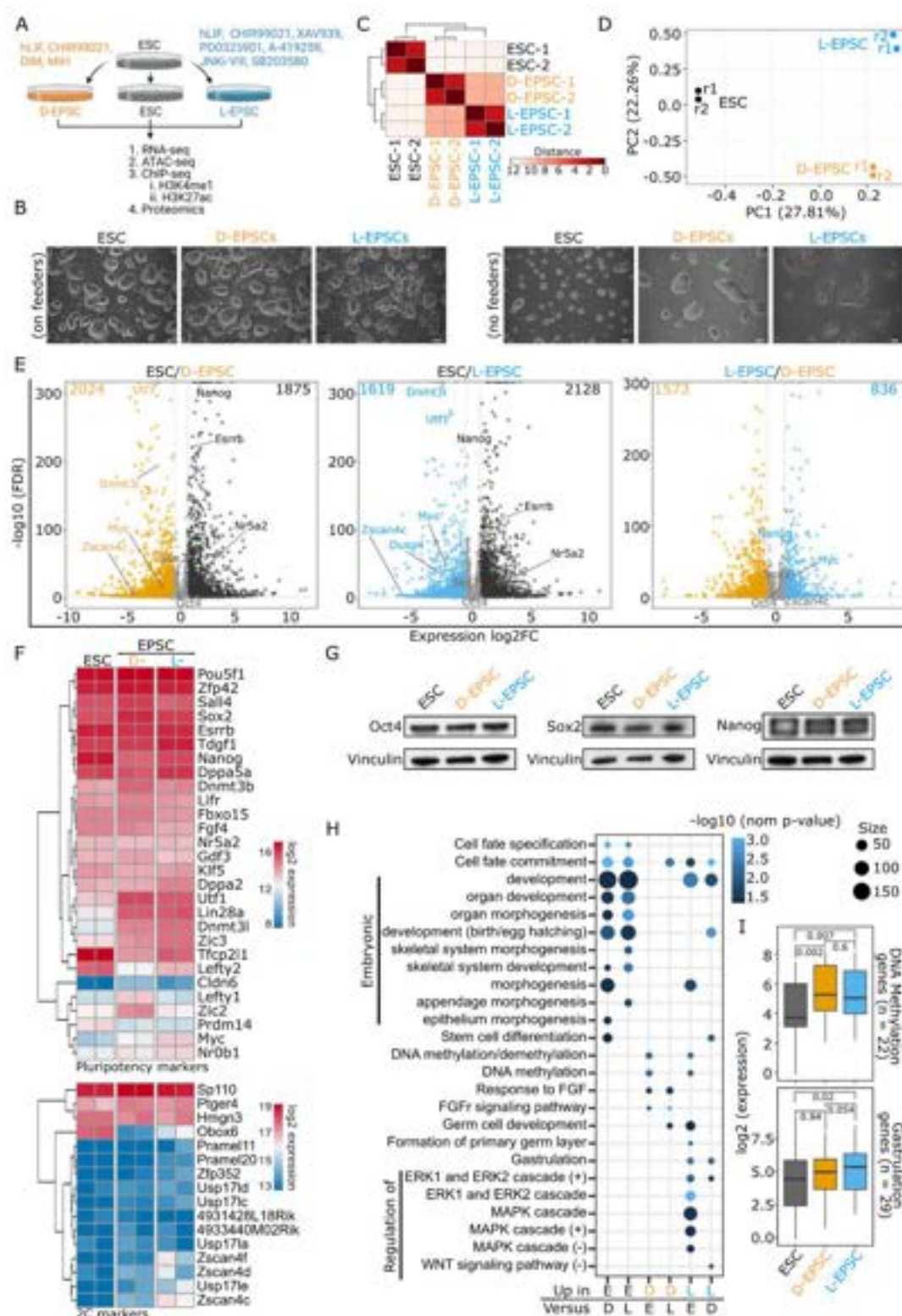
The intersection of differentially expressed genes (DEGs) in ESCs versus EPSCs from this study and previously published bulk RNA-seq of ESCs versus day 15 EPSCs (Posfai et al, 2021b) showed that over half (>55%) DEGs (734/1,334) in the latter study were recapitulated in our studies (Fig S1C). However, we captured a larger number of DEGs that distinguish D/L-EPSCs from ESCs, possibly due to different sequencing platform and depth as well as inherent transcriptomic heterogeneity among different ESC lines from different mouse strains. GSEA analysis using DEGs from all the three comparisons (Fig 1E) showed enrichment for cell fate commitment and embryonic development in ESCs and both EPSC lines (Fig 1H). Interestingly, D-EPSCs showed enrichment of FGF signaling pathway, whereas L-EPSCs are enriched for gastrulation-related terms. Consistently, both EPSC lines showed a strong enrichment of DNA methylation signature and a significant increase in expression levels of DNA methylation-associated genes ( $n = 22$ ), including *Dnmt3a/b/l* and *Mettl4* (Figs 1I and S1A). In contrast, only L-EPSCs showed a significantly higher expression of gastrulation-related genes ( $n = 29$ ) (Figs 1I and S1A).

In sum, these data indicate that, whereas the expression of Oct4, Sox2, and Nanog genes is similar in ESCs and EPSCs, some pluripotency and totipotency-related genes are differentially overexpressed in EPSCs relative to ESCs. Thus, EPSCs do reflect a departure from ESCs at the transcriptome level in both pluripotency and totipotency-related gene expression.

### Chromatin accessibility comparison identifies a subset of putative transcriptional regulators for the unique developmental potential of EPSCs

To examine how the gene expression differences observed above would be correlated with the chromatin status between ESCs and EPSCs, we probed chromatin accessibility by ATAC-seq (assay for transposase-accessible chromatin using sequencing) (Buenrostro et al, 2013, 2015). Our data are highly concordant between replicates (Fig S2A), and PCA revealed the occurrence of accessibility differences in the three cell populations (Fig S2B). Differential accessibility analysis of ATAC-seq peaks showed a large proportion of shared open chromatin among ESCs and both EPSC lines (Fig S2C), yet there also occur subsets of genomic regions that showed significant accessibility changes (Fig 2A). Differential accessibility analysis yielded six groups from three comparisons, that is, (i) ESCs (E) versus D-EPSCs (D): sites open (O) in ESCs and closed (C) in D-EPSCs (EO-DC  $n = 1,753$ ) and vice versa (DO-EC  $n = 9,113$ ); (ii) ESCs versus L-EPSCs (L) (EO-LC  $n = 525$ ; LO-EC  $n = 2,670$ ); and (iii) L-EPSCs versus D-EPSCs (LO-DC  $n = 798$ ; DO-LC  $n = 1,193$ ) (Fig 2A and Table S2). These comparisons reveal that ESCs show more differences in their chromatin opening with D-EPSCs than L-EPSCs (Figs 2A and S2A). The ATAC-seq signals corresponded to the chromatin accessibility group classification according to the identified six groups (Fig 2A)





**Figure 1.** Expanded potential stem cells (EPSCs) share transcriptomic profiles with embryonic stem cells (ESCs) but with discernible differences in pluripotency and 2C genes.

**(A)** A summary of the experimental design. **(B)** Representative phase-contrast images of indicated cellular states with (left panels) or without (right panels) feeders at passage 7. The scale bar is 100  $\mu$ m. **(C)** Hierarchical clustering correlation heat map of the RNA-seq signals from ESC, D-, and L-EPSCs replicates. Distance, calculated with DESeq2 (Love et al, 2014), represents sample-to-sample distances. **(D)** Principal component analysis for the RNA-seq gene expression data. **(E)** Volcano plots of differential gene expression between ESC and D-EPSC (left panel), ESC and L-EPSC (middle), and L- and D-EPSC (right) with selected candidates highlighted. The horizontal



with representative loci shown (Fig 2B). These analyses suggest that, despite major genome-wide chromatin accessibility overlap among ESCs and both EPSC lines (Fig S2C), there exist subsets of differentially accessible regions between EPSCs and ESCs and between the two EPSC types.

The genomic distribution of differentially open peaks in D/L-EPSC lines compared with ESCs showed a stronger enrichment around the TSS region, including CpG island, 5' UTR, and promoter regions (Fig 2C), indicating that access to gene-proximal regulatory elements may play a significant role during the ESC-to-EPSC conversion. Furthermore, in comparison to ESCs, the differentially open ATAC-seq peaks containing LTRs were depleted in both EPSC lines (Fig 2C), which corresponds with significantly lower expression levels of genes associated with these LTR-containing regions in EPSCs than ESCs (Fig 2D). Similar results were reported by a previous study showing reduced expression of porcine endogenous retroviruses (PERVs) in porcine-EPSCs (Kruger et al, 2021), suggesting a cross-species conserved phenomenon.

To gain mechanistic insight into the unique chromatin dynamics of EPSCs, we analyzed the DNA motifs associated with the overlapping and differentially accessible peaks in ESCs and both EPSC lines. The overlapping peaks among all three cell types mainly harbored the pluripotency-related TF motifs (Fig S2D), whereas the differentially accessible six groups (defined in Fig 2A) showed the presence of various classes of TF motifs in addition to the pluripotency related motifs (Fig 2E and Table S2). We further filtered the motifs and only kept the ones ( $n = 12$ ) present in more than 50% of open peaks in ESCs, D-EPSCs, and L-EPSCs (Figs 2F and S2E). We found that ESCs are enriched for Tcf/Left-like motifs, D-EPSCs for Ronin/Gly-Staf, REST-NRSF, Six4, NRF, and Zfp281 motifs, whereas L-EPSCs for retinoic acid receptor-retinoid X receptor (RAR-RXR) heterodimer binding motifs (Figs 2F and S2E). The enrichment of the Zfp281 motif in D-EPSCs and the RAR-RXR motif in L-EPSCs is contrasting and intriguing because Zfp281 inhibits (Wen et al, 2022), whereas the retinoic acid (RA)-signaling pathway promotes the transition of ESCs to totipotent two-cell-like cells (2CLCs) (Iturbide et al, 2021). Nonetheless, we noticed that the promoter regions of Zfp281 together with Rara, Rarg, Rxrb, and Rxrg genes among RAR and RXR gene families are accessible and enriched for active histone marks. Consistently, we found that their transcripts are overexpressed in EPSCs compared with ESCs (Fig S2F).

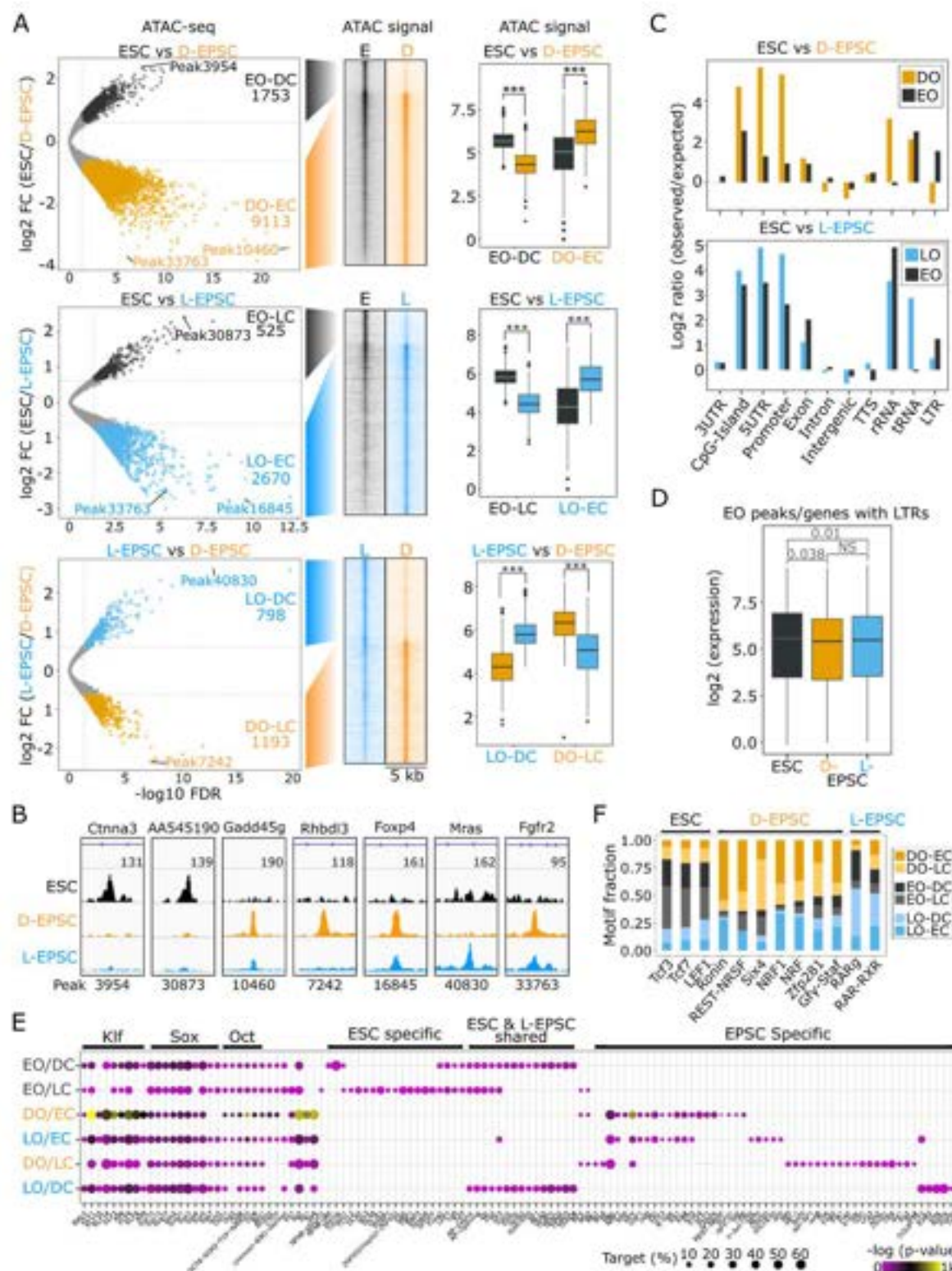
In sum, compared with ESCs, EPSCs show enrichment of a unique subset of TF motifs, including Zfp281 and RAR-RXR with corresponding gene expression changes and closing of LTR containing regions and their reduced gene expression, which may provide critical regulatory elements that drive the expanded potential of EPSCs, a hypothesis that warrants future investigations.

## Histone marks, transcription, and chromatin accessibility based molecular features of EPSCs

To investigate the potential impact of differential open chromatin (Fig 2A) on differential gene expression in EPSCs compared with ESCs, we performed ChIP-seq of H3K4me1 and H3K27ac to characterize active promoters/enhancers in ESCs and EPSCs and expand the resource for these cells in addition to the previously characterized bivalent marks (H3K4me3 and H3K27me3) in ESCs (Bernstein et al, 2006) and EPSCs (Yang et al, 2017a). Both ChIP-seq replicates correlate well (Fig S3A), enabling us to identify the chromatin environment and signature genes regulating EPSCs fate with the following approaches. First, we defined the peaks with differential enrichment of histone marks in ESCs versus D/L-EPSC lines (Fig S3B and C). Second, we performed a combinatorial analysis by intersecting the regions (i) with differential enrichment of histone marks (Fig S3B), (ii) that are accessible (Fig 2A), and (iii) with their nearby genes up-regulated in both EPSCs and ESCs (Fig 1E). As a result, we defined 131 and 85 signature genes for EPSCs and ESCs, respectively (Fig 3A and Table S3). Third, we performed GO analysis of these identified signature genes, revealing DNA methylation (e.g., *Dnmt3a*), pluripotency network (e.g., *Tfap2c*, *Zmym2*, and *Lefty1*), and MAPK/ERK signaling pathways (e.g., *Rora*, *Fgfr2*, and *Dusp6*) are prominent features associated with EPSC signature genes, whereas otic vesicle development (e.g., *Fgf10*, *Fgf3*, and *Eya1*) and LIF signaling pathways (e.g., *Nr5a2*, *Hk2*, *Trim2*, *Kat6b*, and *Arid5b*) are more pronounced in ESC signature genes (Fig 3B). The representative candidates from 131 EPSC signature genes have open promoter regions, higher H3K4me1 and H3K27ac signals, and higher mRNA expression levels in either D- or L-EPSCs compared with ESCs (Fig 3C). Of note, *Tfap2c* is a trophoblast (Kuckenberg et al, 2010) and naive pluripotency marker (Pastor et al, 2018). Among 272 mouse cell lines or tissue samples (Hutchins et al, 2017), *Tfap2c* transcripts are expressed at high levels in placenta, trophoblast stem cells, 4/8-cell embryos followed by ESCs (Fig S3D). We noticed that, unlike its mRNA expression (Fig 3C), *Tfap2c* is expressed at higher protein levels in both D/L-EPSCs than ESCs (Fig S3E), suggesting a potential posttranscriptional regulation. To understand its functional role in EPSCs, we used *Tfap2a/c* double knockout (dKO) mouse ESCs (Pastor et al, 2018). We found that under respective culture medium, *Tfap2a/c* dKO ESCs could be efficiently converted to D-EPSCs but less so to L-EPSCs compared with WT ESCs, evident with morphological changes (flat colonies with individualized cells) as signs of differentiation already within the first five passages which could not be further passaged and maintained (Fig S3F). These results indicate the differential requirement of *Tfap2a/c* for the EPSC state by different culture conditions, suggesting D- and L-EPSCs may represent two distinct expanded pluripotency states.

and vertical dashed lines represent the false discovery rate (FDR, 0.05) and log<sub>2</sub> fold change (x0.6) cut-offs, respectively. The numbers of differentially expressed genes are indicated on the top corners and details provided in Table S1. (F) Heat maps show the expression dynamics of selected pluripotency (top) and 2C (bottom) markers genes for ESC, D-EPSC and L-EPSC. Data are presented as log<sub>2</sub> normalized counts. (G) Western blots for Oct4, Sox2, and Nanog in three cell types. Vinculin is the loading control. (H) GSEA of the differentially expressed genes in E filtered by FDR < 0.05, log<sub>2</sub> FC > 0.6 & < -0.6. Select GO terms were extracted using keywords "Embryonic, FGF, Gastrulation, Stem cell, MAP, ERK, DNA methylation" out of 927 significant terms from six groups. Size represents the number of genes associated with each term and color -log<sub>10</sub> nominal P-value. (H, I) Box plots of RNA-seq signals of selected genes associated with GSEA terms DNA methylation ( $n = 22$ , Table S1) and gastrulation ( $n = 29$ , Table S1) from heat map in (H). Indicated P-values were calculated using the unpaired Mann-Whitney U test also known as Wilcoxon rank-sum test (R function Wilcoxon test).





**Figure 2. Expanded potential stem cells (EPSCs) show differential chromatin accessibility near the promoter region than embryonic stem cells (ESCs).** (A) Volcano plots of differential genomic accessibility performed using DiffBind tool (Ross-Innes et al., 2012) between ESC and D-EPSC (top panel), ESC and L-EPSC (middle), and L- and D-EPSC (bottom); named using acronyms: Open (O) and Close (C) in ESC (E), D-EPSC (D), and L-EPSC (L) conditions. Peak numbers and select peaks with differential accessibility in each category are indicated. A heat map of the ATAC-seq signal (RPM normalized) within a 2.5-kb window centered at six accessibility peaks was drawn using EaSeq (Lerdrup et al., 2016) and corresponding boxplots. P-values ( $^*P < 0.05$ ;  $^{**}P < 0.01$ ;  $^{***}P < 0.001$ ) were calculated using the unpaired Mann-Whitney Wilcoxon Test (R function Wilcox test). Genomic locations are provided in Table S2. (A, B) Representative genome browser tracks of selected ATAC-seq



The overexpression of *Zmym2* and *Rora* in EPSCs is intriguing because we and others have shown that *Zmym2* and *RAR* family proteins play negative (Yang et al, 2020) and positive (Iturbide et al, 2021) roles, respectively, in 2CLC totipotency induction. Nonetheless, the enrichment of *RAR-RXR* motif in more than 50% differential open region in L-EPSCs (Fig 2F) and the overexpression of *RAR* and *RXR* family members in both EPSCs (Fig S2F) suggest the potential roles of *RAR/RXR* factors in EPSC fate regulation. Collectively, we identified a subset of potential EPSC signature genes that could serve as prime candidates in constructing the regulatory network governing the unique developmental potential of EPSCs.

### Like ESCs, EPSCs also rely on Oct4 and Sox2 but not Nanog for their maintenance

Our data indicate that although we identified EPSC-specific gene signatures, EPSCs also share with ESCs some of the gene expression, chromatin accessibility, and histone marks. In particular, we wondered to what extent EPSCs are dependent on the pluripotency network for their maintenance. In this regard, we focused on the core pluripotency factors Oct4 (Niwa et al, 2000), Sox2 (Masui et al, 2007), and Nanog (Chambers et al, 2003; Das et al, 2011). We used previously established ZHBTc4, 2TS22C, and NgCKO<sup>tet-off</sup>ESC lines for tetracycline/doxycycline-induced conditional knockout (cKO) of Oct4 (Niwa et al, 2000), Sox2 (Masui et al, 2007), and Nanog (Das et al, 2011), respectively. In these ESC lines, the respective tet-off transgene (Fig 4A) sustains the self-renewal of tet-off ESCs that are genetically null for each endogenous gene (Fig 4A) before the doxycycline (Dox) treatment. We were thus able to convert them to stable tet-off L-EPSCs without Dox (Fig 4A and B) using the published protocol (Yang et al, 2017a). By adding Dox to turn off the Oct4 and Sox2 transgenes, we observed the collapse of EPSCs concomitant with the protein loss (Fig 4B and C). In contrast, despite the Nanog protein loss (Fig 4C), Nanog-cKO<sup>tet-off</sup>EPSCs (NgCKO), and their converted tet-off L-EPSCs could still be maintained, although with reduced size and number of colonies (Fig 4B), indicating a similar role of Nanog in L-EPSC maintenance as that of ESC maintenance. Of note, it is well known that ESCs without Nanog can be maintained although they proliferate slower and are more prone to differentiation (Chambers et al, 2007).

To understand the downstream effect of the loss of Oct4, Sox2, and Nanog proteins on EPSCs relative to ESCs, we performed RNA-seq in all three tet-off ESC lines and corresponding tet-off L-EPSC lines with and without Dox. All replicates correlated well with each other (Fig S4A). All three tet-off ESC lines are separated from their individually derived tet-off L-EPSC lines at the transcriptome level (Fig 4D). However, whereas the Dox-treated Oct4-cKO and Sox2-cKO<sup>tet-off</sup>L-EPSCs clustered separately from their untreated counterparts, Dox-treated and untreated Nanog-cKO<sup>tet-off</sup>L-EPSCs

clustered together (Figs 4D and S4A). These data indicate two major gene expression changes happen, first during the establishment phase, that is, tet-off ESC-to-tet-off L-EPSC conversion in all three cKO cell types, and second during the maintenance phase, that is, Oct4-cKO and Sox2-cKO<sup>tet-off</sup>L-EPSCs after Dox addition. Of note, there were only minimal or modest gene expression changes in Nanog-cKO<sup>tet-off</sup>L-EPSCs upon Dox addition, consistent with the minimal effect of the Nanog loss on EPSC morphology/maintenance (Fig 4B). Consistently, global DEG analysis in Dox-treated and untreated tet-off L-EPSCs revealed that both Oct4-cKO and Sox2-cKO<sup>tet-off</sup>L-EPSCs showed more up-regulated ( $n = 1,517$  and  $n = 1,407$ ) and down-regulated ( $n = 1,895$  and  $n = 1,261$ ) genes compared with only a few hundred (up-regulated = 202 and down-regulated = 304) in Nanog-cKO<sup>tet-off</sup>L-EPSCs (Fig 4E and Table S4). GO analysis of these differentially regulated, especially down-regulated, genes in all three conditions revealed pluripotency network as one of the most affected terms (Fig S4B). The expression of major pluripotency-related markers was significantly down-regulated in Oct4-cKO and Sox2-cKO but not in Nanog-cKO<sup>tet-off</sup>L-EPSCs (Figs 4F and S4C). In contrast, the expression of totipotency-associated 2C marker genes did not show a significant change in all three cell lines (Fig S4D).

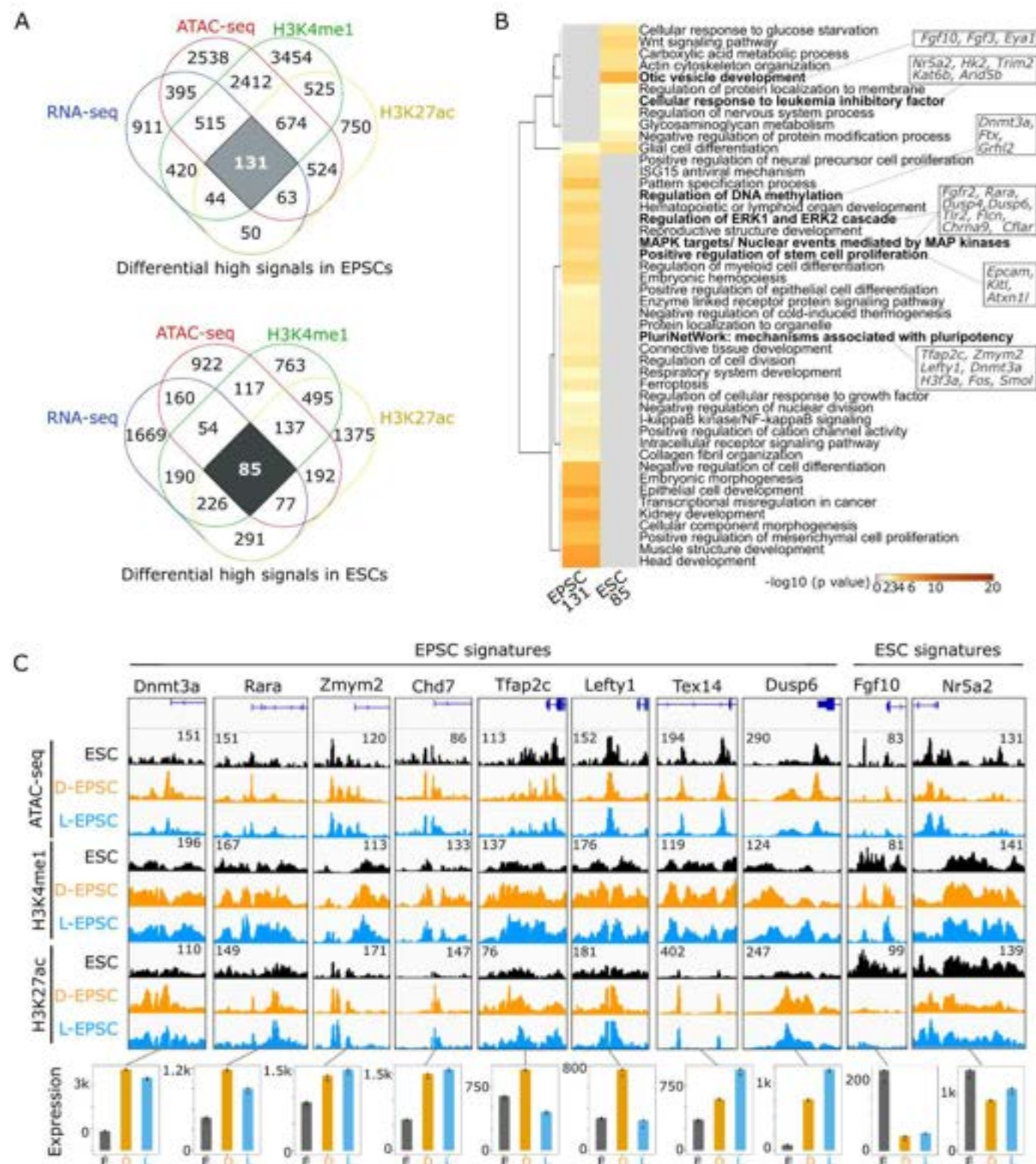
Puzzling enough, we could not generate tet-off D-EPSCs from these cKO cell lines despite repeated trials following strictly the published protocol (Yang et al, 2017b) (indicated by a red "X" in Fig 4A). We suspected that was due to the similarity of structure and mechanism of action between Dox and minocycline hydrochloride (Mih), a chemical compound required for the induction of D-EPSCs (Yang et al, 2017b). To ascertain this is the case, we resorted to an Oct4-degron ESC system allowing for rapid Oct4 protein degradation with dTAG treatment (Boija et al, 2018). Indeed, we could successfully convert Oct4-degron ESCs to both tet-off D- and tet-off L-EPSCs without dTAG treatment (Fig S4E), and dTAG addition induced Oct4 protein degradation concomitant with differentiation of tet-off ESCs and both tet-off EPSC lines and the eventual collapse of all these cells. Collectively, our data indicate that like ESCs, EPSC maintenance is also critically dependent on pluripotency factors Oct4 and Sox2 and, to a much lesser extent, Nanog.

### The proteomic comparison reveals differential translational and metabolic control between ESCs and EPSCs

To understand the functional outcomes of the global genomic and chromatin differences in EPSCs versus ESCs, we interrogated the differential proteome in ESCs and EPSCs. We performed quantitative proteomics using SILAC-based MS with biological replicates to achieve a high-accuracy analysis of the proteome for each cellular state (Fig 5A). We identified a total of 1,103 proteins among which 129 and 79 are up-regulated ( $\log_2$  SILAC ratios > 0.6 & < -0.6) in

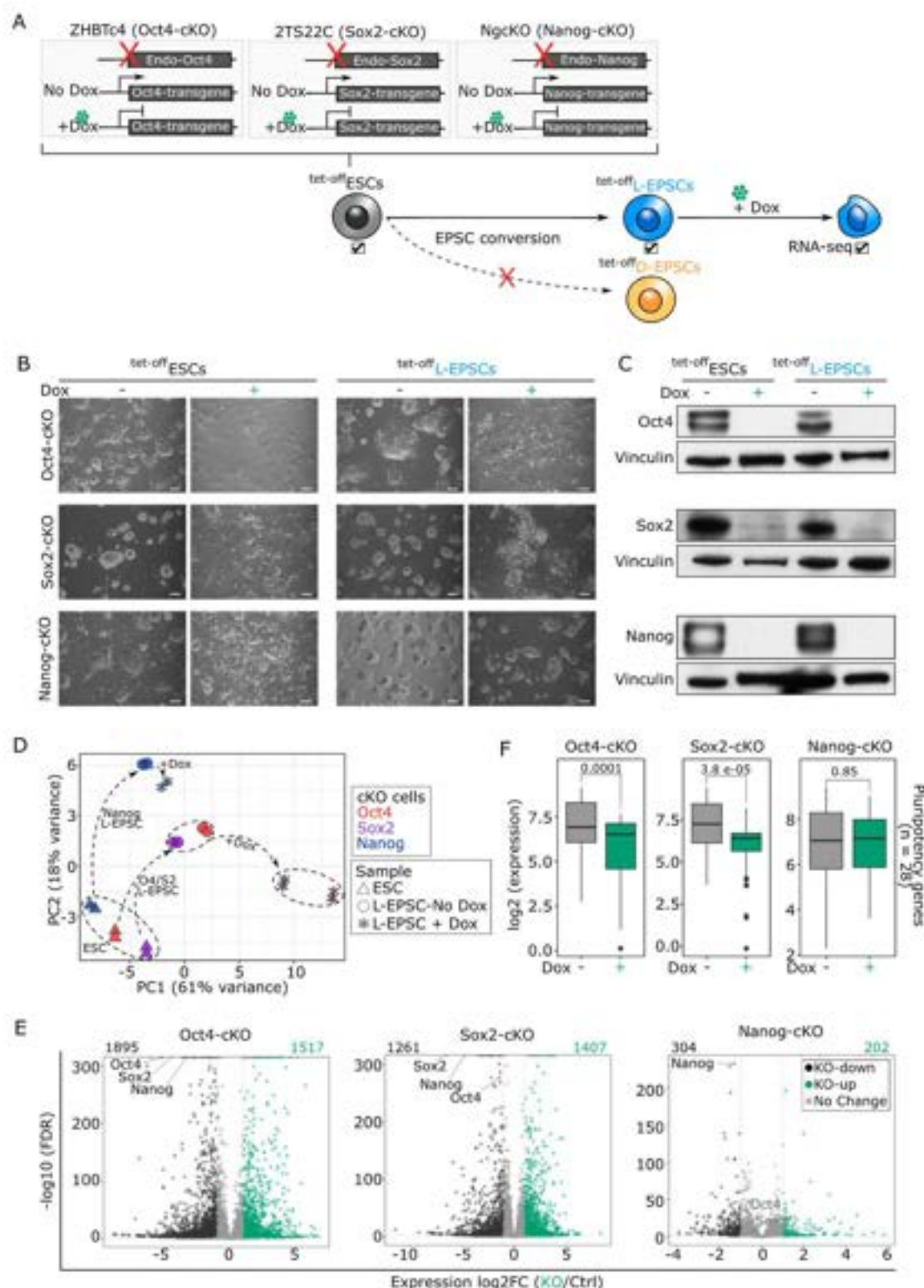
peaks (indicated in A). The nearest genes are displayed on the top, peak number identifiers at the bottom, and signal strength is shown with respective numbers. (C) Genomic distribution of ATAC-seq peaks from indicated groups are shown as bar plots. (D) Boxplots for RNA-seq signals of nearest genes associated with ATAC-seq peaks harboring LTRs from combined-ESC-open peaks compared with D- & L-EPSC categories. Indicated P-values were calculated using the unpaired Mann-Whitney Wilcoxon Test (R function Wilcox test); NS, not significant. (E) Enrichment of selected motifs in six differential-ATAC-seq categories obtained using HOMER (Heinz et al, 2010). Point size represents the proportion of sequences with the motif, and color gradient the P-value score. Major motif classes are indicated on top; list provided in Table S2. (F, G) Proportion of ATAC-seq peaks featuring select top motifs (from E) in each accessibility category. Motifs from (E) were selected if they occurred in more than 50% of ATAC-seq peaks in EO, DO, and LO categories.





**Figure 3. Defining the molecular signature of expanded potential stem cells (EPSCs).**

(A) Venn diagram showing the intersection of overexpressed genes from RNA-seq, open regions from ATAC-seq with the presence of H3K4me1 and H3K27ac signal in EPSCs (top) and embryonic stem cells (bottom). (A, B) GO analysis of 131 and 85 genes from (A) using metasplice (Zhou et al., 2019). Genes associated with selected terms are indicated. (A, C) Genome browser tracks represent ATAC-seq and indicated histone marks ChIP-seq signals around promoter regions of selected genes from 131 or 81 candidates in (A) and their expression levels in three cell states. ATAC-seq and ChIP-seq signal strengths, set same for embryonic stem cell, D-EPSC, L-EPSC samples, are indicated once with respective numbers for each gene loci.



**Figure 4. Requirement of Oct4, Sox2, and, to a lesser extent, Nanog for expanded potential stem cell (EPSC) maintenance.**

**(A)** Illustration of RNA-seq experiment (tick marks) on *tet-off* L-EPSCs converted from Oct4 (Niwa et al., 2000), Sox2 (Masui et al., 2007), and Nanog (Das et al., 2011) conditional knockout (Oct4-cKO, Sox2-cKO, and Nanog-cKO) *tet-off* ESCs, wherein each of these endogenous (endo) genes is deleted, and cells are maintained by their corresponding doxycycline (Dox)-suppressible transgene. **(B, C)** Cellular morphology (B) and protein expression (C) on day 5 after Dox treatment (1  $\mu$ g/ml) is shown in *tet-off* ESCs and *tet-off* L-EPSCs. The scale bar is 100  $\mu$ m. Vinculin is the housekeeping control. **(D)** Principal component analysis for the RNA-seq gene expression from indicated samples. **(E)** Volcano plots of differential gene expression of Dox-treated versus untreated *tet-off* L-EPSCs from O/S/N-cKO backgrounds. Differentially expressed



ESCs compared with D-EPSCs and L-EPSCs, respectively (Fig 5B). As expected, we observed higher levels of Parp1 in ESCs than both L- and D-EPSCs (Fig 5B) because both chemical Parp1 inhibition and genetic Parp1-KO are reported to be beneficial for EPSC maintenance and developmental potency (Yang et al, 2017b). D-EPSCs overexpress pluripotency-related proteins like Lin28a and Dnmt3L, whereas Utr1 is overexpressed in both D- and L-EPSCs compared with ESCs (Figs 5B and 55A). The GO analysis of proteins overexpressed in D-EPSCs compared with ESCs showed enrichment of terms related to posttranscriptional gene regulation and translation (Fig 5C). Consistently, D-EPSCs showed enrichment of eukaryotic translation initiation factors, namely, Eif3b/c/e/m, Eif4g1, Eif6 and translation initiation-associated ribosomal protein Rps15a compared with ESCs (Figs 5B and 56). During mouse early embryonic development, the transcript levels of the above-mentioned translation initiation factors are higher at E5.5 (an embryonic stage that correlates with D-EPSCs at the single-cell level [Posfai et al, 2021b]) than ESCs (Fig 55B). On the other hand, compared with D-EPSCs, L-EPSCs specifically showed enrichment of Eif4a2 protein (Fig 5B). Similarly, the transcript levels of Eif4a2 are abundant in the E4.0 state and subsequently decrease in both ESCs and further in D-EPSCs (Fig 55B). Notably, L-EPSCs are correlated at the mRNA level with the E4.5 stage (Posfai et al, 2021b). We recently showed that Eif4a2 mediates specialized translational control of mRNA targets governing stem cell and developmental potency (Li et al, 2022). Thus, overexpression of Eif4a2 in L-EPSCs compared with both ESCs and D-EPSCs and other translation-related factors in D-EPSCs suggest that the two EPSC types might use different translational mechanisms to control expanded pluripotency.

The GO analysis for proteins up-regulated in L-EPSCs compared with both ESCs and D-EPSCs showed enrichment of terms like TCA cycle and fatty acid  $\beta$ -oxidation (Fig 5C). We observed overexpression of multiple TCA cycle-related proteins, namely, Aco2, Sdhb/a, Ndufs1, Ndufs3, Ndufa9, Vdac1, and Acad9 in L-EPSCs compared with both ESCs and D-EPSCs (Figs 5B and 56). In contrast, ESCs mainly overexpress glycolytic proteins (e.g., Hk2, Tigar, Atp1a1, Vdac3, and Pfkfb1) and G1/S transition related proteins (e.g., Cdk2, Cdk4, Mcm3, and Mcm4) compared with D-EPSCs and L-EPSCs, respectively. Among these proteins, the majority of TCA (e.g., Aco2, Ndufs1, Ndufs3, Ndufa9, Sdhb, and Sdhb) and glycolysis (e.g., Atp1a1, Hk2, and Tigar) related factors showed concordant expression at both mRNA and protein levels in ESCs versus L-EPSCs (Fig S7). For the rest of the proteins (e.g., Acad9, Pfkfb1, and Vdac3) the mRNA and protein expression were not concordant suggesting possible post-transcriptional regulation. These results suggest that ESCs and EPSCs, especially L-EPSCs, likely depend on different metabolic needs, supported by previous studies demonstrating that mouse early embryos (with totipotency and/or extended pluripotency) are dependent on pyruvate (TCA cycle), whereas ESCs on glucose (glycolytic pathway) as their main energy source (Nagaraj et al, 2017; Zhang et al, 2018a).

In sum, our proteomics data on three cellular states highlight that EPSCs may use differential translational and metabolic

controls to acquire their expanded potential over ESCs, pointing another direction for future mechanistic inquiries into the EPSC biology.

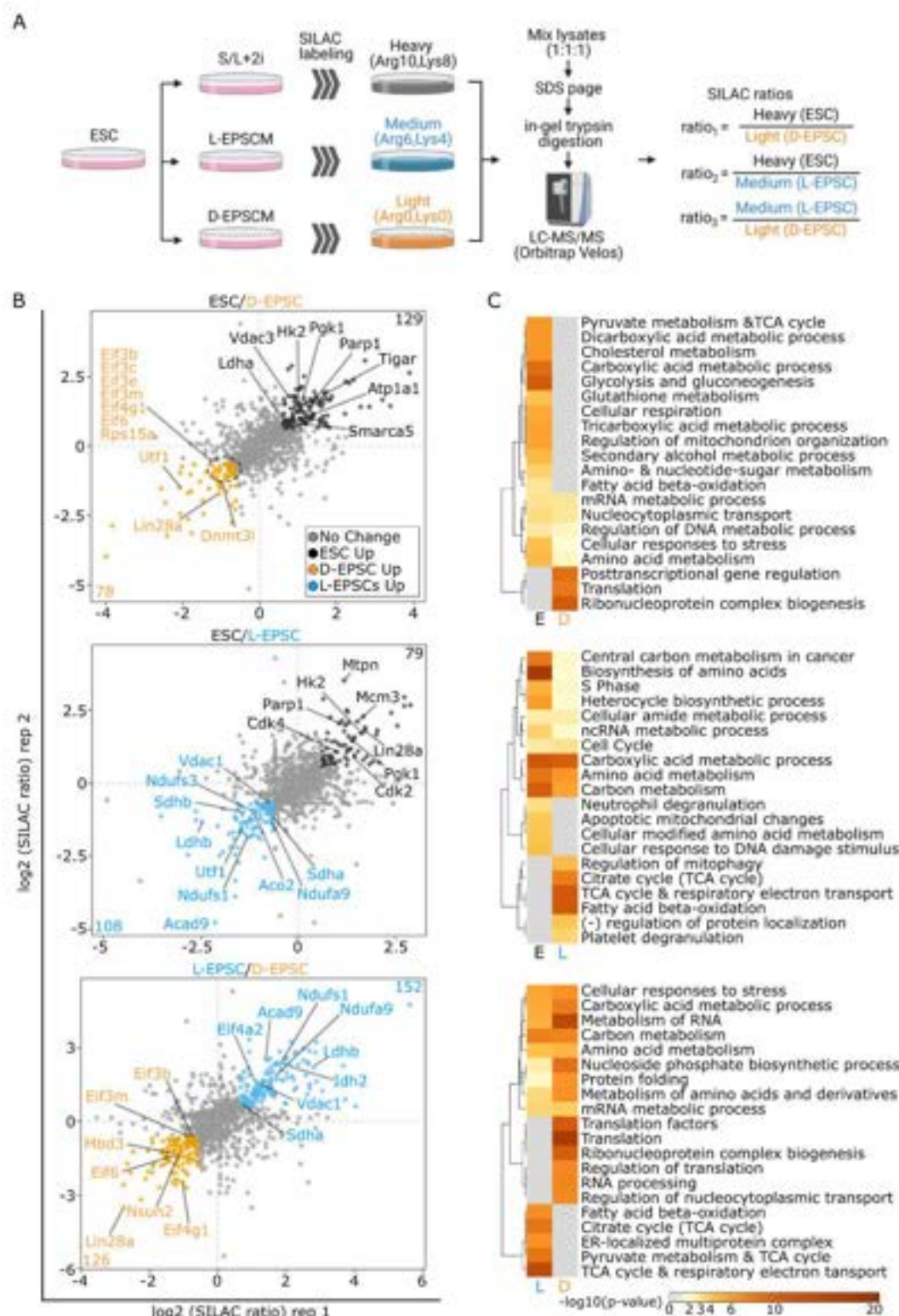
## Discussion

The Deng (Yang et al, 2017b) and Liu (Yang et al, 2017a) laboratories simultaneously established EPSCs with unique developmental potential over ESCs, which ushered in valuable applications of these cells in enhanced directed differentiation (Wang et al, 2020), blastoid and interspecies chimera generation (Sozen et al, 2019; Li et al, 2019b; Tan et al, 2021), and faster mouse model generation (Li et al, 2019a). Despite the challenge raised on their expanded potential (Posfai et al, 2021b), EPSCs were subsequently further proved to be advantageous over ESCs in interspecies monkey-human chimera generation (Tan et al, 2021) and the derivation of totipotent potential stem cells (TPSCs) (Xu et al, 2022). These studies definitely highlight the untapped potential of EPSCs for further exploration to drive studies towards in vitro establishment of authentic totipotent stem cells, although the molecular foundation of these EPSCs are still poorly characterized. Here we used comprehensive genomics, epigenetic and proteomics approaches to compare the molecular features of the two EPSC lines and their similarities and differences compared with ESCs that remained unexplored by previous studies. Our data demonstrate that EPSCs express the core pluripotency factors Oct4, Sox2, and Nanog but differentially overexpress other pluripotency-associated factors such as *Utr1*, *Dnmt3a/b/l*, *Zfp281*, and *Zmym2* compared with ESCs (Fig 6). In addition, EPSCs, especially L-EPSCs, also show slight up-regulation of *Zscan4c*, a totipotency-associated factor, compared with ESCs (Figs 1F and 6). Apart from highlighting the subtle differences in pluripotency and totipotency related genes which Posfai et al (2021a, 2021b) had overlooked, we also showed how subsets of genes related to DNA methylation and gastrulation (Figures 1H and I) could be explored further to guide studies that are focused on establishing the authentic totipotent stem cells in vitro.

The actual developmental potency of totipotent-like cells generated across various laboratories remains murky because these cell types show variable molecular features (Yang et al, 2017a, 2017b; Sozen et al, 2019; Li et al, 2019b; Tan et al, 2021; Posfai et al, 2021b; Xu et al, 2022). Thus, it is critical to know where precisely these cells map/position on the early development trajectory (Posfai et al, 2021a; Malik & Wang, 2022). Single-cell transcriptomic profile comparison with mouse early embryonic development had classified L-EPSCs slightly earlier (E4.5) than D-EPSCs (E5.5) (Posfai et al, 2021b). We also noticed some remarkable differences in a few molecular features that suggest the placement of L-EPSCs at a slightly earlier developmental stage than D-EPSCs. For example, higher expression of Lin28a protein, representative of primed pluripotency state (Zhang et al, 2016), was observed in D-EPSCs, whereas overexpression of TCA cycle-related proteins, a feature of

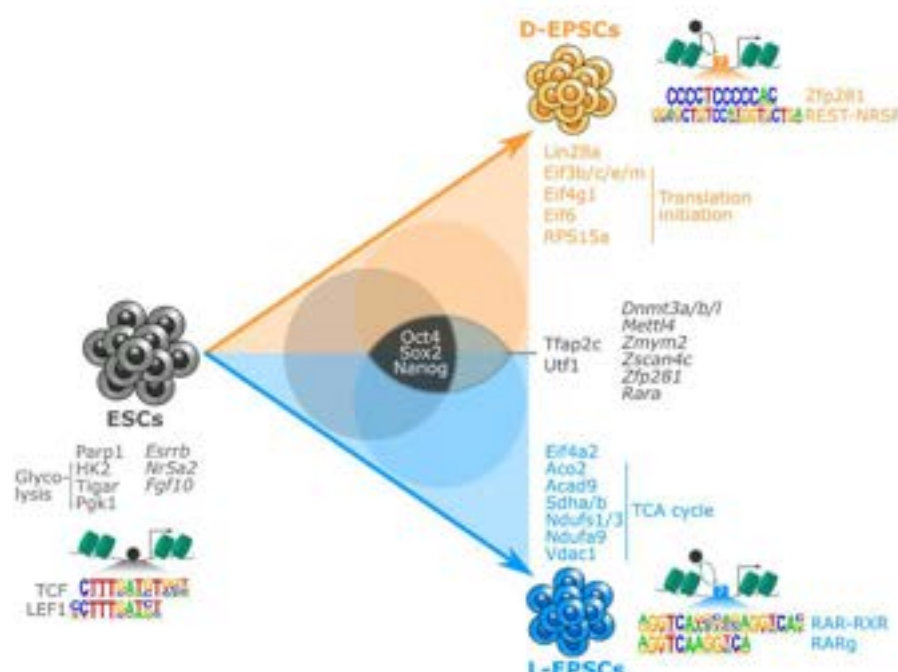
gene numbers are indicated, and a list is provided in Table S4. (F) Boxplots showing expression of pluripotency related genes ( $n = 29$  from Fig 1F and Table S4) in Dox-treated and untreated  $10^6$ -d L-EPSCs.





**Figure 5. D- and L-expanded potential stem cells (EPSCs) show overexpression of translation initiation and TCA cycle-related proteins.** (A) Illustration showing proteomics experiment using SILAC-based mass spectrometry (MS). (B) Protein ratios of two independent mass spec whole proteome measurements of embryonic stem cells, D-EPSCs, and L-EPSCs SILAC labeling are shown as dot plots (Table S5). Differentially expressed proteins ( $\log_2$  ratio  $> 0.6$  and  $< -0.6$ ) are highlighted with darker colors, and their numbers are indicated in respective quadrants. Select proteins are labeled. (B, C) GO analysis using the differentially expressed proteins ( $\log_2$  ratio  $> 0.6$  &  $< -0.6$  i.e., colored dots except grey) from (B).





**Figure 6. Summary of the molecular features associated with embryonic stem cells, D-, and L-expanded potential stem cells (EPSCs).** Schematic representation showing conversion of embryonic stem cells to D- and L-EPSCs with shared proteins at the center (Oct4, Sox2, and Nanog) and uniquely enriched proteins and mRNA (italicized) at the edges of the triangle. D- and L-EPSCs shared proteins and mRNAs are in the middle of the two cell types. Open chromatin and enriched motifs in each cell type are indicated. The pluripotency factors (filled grey circle) and EPSCs fate regulators (squares with the question mark) are shown.

much earlier developmental stage (Nagaraj et al, 2017; Zhang et al, 2018a), was observed in L-EPSCs (Figs 5B and 6). The relatively higher proliferation rate of D-EPSCs compared with L-EPSCs could be further explored to highlight if there is any connection with their dependence on different metabolic pathways (Yang et al, 2017a, 2017b). Similarly, L-EPSCs with the overexpression of Eif4a2 are better associated with an earlier embryonic development stage, whereas D-EPSCs show overexpression of multiple translation initiation factor proteins representing a later developmental stage than ESCs. These datasets suggest an intriguing possibility that these factors may contribute to the differential developmental potential of D/L-EPSCs and/or the existence of alternative states of expanded pluripotency of EPSCs. It also raises an interesting question of whether one EPSC line might have higher plasticity than the other.

The conversion of ESCs to EPSCs led to differential chromatin opening concomitant with enrichment of various transcription factor motifs in differentially accessible sites (Figs 2E and 6). It will be interesting to know which factors are directly (pioneers) or indirectly (settlers or migrants) (Sherwood et al, 2014) involved in bringing about these chromatin accessibility changes. Immediate genome-wide occupancy studies of crucial transcription factor(s) combined with our existing ATAC-seq data analysis would help understand the dynamic processes that possibly involve passive displacement of pluripotency factors leading to chromatin closing and/or active involvement of the pioneering role of expanded pluripotency related factors leading to chromatin opening (Fig 6). In addition, it would be important to know whether major pluripotency factors, especially Oct4 and Sox2, bind to different locations in EPSCs than ESCs. Sox2 acts as a major pioneer factor in pluripotency induction (Zhu et al, 2018; Malik et al, 2019) and partners with different proteins in a cell context-dependent manner (Adachi

et al, 2013; Lodato et al, 2013), so it would be interesting to know if there are changes in Sox2 binding locations and partners in EPSCs as well. Similarly, the determination of binding locations of identified EPSC-enriched factors (Fig 6) would help us understand their regulation more precisely. Zfp281 and Rara transcripts are expressed at nearly identical levels in both EPSCs (Figs S2F and 6). Yet, their DNA binding motifs are enriched differentially in the accessible regions in D- and L-EPSCs, respectively (Figs 2F and 6). To understand whether these differentially open locations containing the abovementioned DNA motifs are actually bound by their respective transcription factors, it would be of prime interest to catalog all the genome-wide locations bound by expanded potential related factors. We noticed the presence of the RAR binding motif upstream of Eif4a2 in L-EPSCs (Fig S5C). Using our previously published ChIP-seq data (Fidalgo et al, 2016), we found that Zfp281 binds the Dnmt3a and Lin28a genomic regions in ESCs (Fig S5D). A comparative genome-wide binding study in ESCs and EPSCs will help identify more regulatory signatures like this. Such studies will thus delineate if and how Zfp281 and RAR-RXR family proteins may contribute to the expanded potential of EPSCs, particularly considering activation of RAR signaling are important for inducing and maintaining totipotent features of TPSCs (Xu et al, 2022) and both Zfp281 and RAR-RXR family proteins may play a role in cell division and DNA replication-free reprogramming of somatic nuclei for embryonic transcription (Tomikawa et al, 2021). Recently Hu et al succeeded in capturing 2CLCs in vitro using three chemicals including TTNPB (RAR-agonist) (Hu et al, 2022), thus corroborating our findings and strengthening the demand to explore the possible role of these factors in expanded potential of EPSCs.

Open chromatin regions in the mouse genome (about 45%) are associated with repeat elements. The accessibility around these repeat elements is highly dynamic during early embryonic



development and ESCs (Lu et al, 2020). Similarly, global DNA methylation is also dynamic during mouse early development stages and controls the expression of repeat elements (Smith et al, 2012). Although it is well known that mouse early embryos and ESCs show higher expression of ERVs (regulated by LTRs), we found that, compared with ESCs, differentially open regions in EPSCs were depleted around LTRs and matched with significant down-regulation of the corresponding RNA expression (Fig 2C and D) in alignment with a previous report (Kruger et al, 2021). We also noticed that EPSCs overexpress Dnmt3 enzymes (Figs S1A, S8, and 6). Because DNA methylation levels control the repression of repeat elements (Smith et al, 2012; He et al, 2019), it raises more questions as to whether differential overexpression of DNA methylation enzymes is at interplay to control the selective repression of LTR-containing open regions in EPSCs.

In summary, we systematically compared ESCs and EPSCs with combined genomic, transcriptomic, and proteomic approaches. We identified critical genomic loci and proteins that provide a rich resource for further investigations as the potential molecular targets endowing EPSCs with superior developmental potential over ESCs (Fig 6). Similar studies are warranted for other currently available totipotent-like cells, including 2CLCs (Macfarlan et al, 2012), totipotent blastomere-like cells (TBLCs) (Shen et al, 2021), totipotent-like stem cells (TLSCs) (Yang et al, 2022), TPSCs (Xu et al, 2022), chemically induced totipotent stem cells (ciTotiSCs) (Hu et al, 2022), and human BCLCs (Mazid et al, 2022; Taubenschmid-Stowers et al, 2022). Ultimately, we will be able to capture authentic totipotent stem cells in vitro (Malik & Wang, 2022) and likely also reveal alternative paths to totipotency and/or alternative totipotent states.

## Materials and Methods

### ESC and EPSC culture

#### Mouse ES cells (mESCs) J1

Mouse ES cells (mESCs) J1 were cultured on inactive MEF feeder cells (~30,000 cells per cm<sup>2</sup>) or on 0.1% gelatin-coated plates in a serum-based medium with 2i (3  $\mu$ M CHIR99021; 1  $\mu$ M PD0325901) addition, which was prepared as follows: DMEM supplemented with FBS (15%), recombinant LIF (1,000 U/ml),  $\beta$ -mercaptoethanol (0.1 mM), L-glutamine (2 mM), MEM non-essential amino acids (NEAA, 0.1 mM), nucleoside mix (1%), and penicillin/streptomycin (50 U/ml). The medium was changed daily or every other day, and cells were passaged every 3 d using trypsin (0.05%).

#### D-EPSCs

D-EPSCs were cultured in a base medium of N2B27 prepared as follows: DMEM/F12 and Neurobasal (1:1 ratio), N2 supplement (1 $\times$ ), B27 supplement (1 $\times$ ), GlutaMAX (1 $\times$ ), and  $\beta$ -mercaptoethanol (0.1 mM). The base medium was supplemented with KSR (5%), NEAA (0.1 mM), LIF (1,000 U/ml), CHIR99021 (3  $\mu$ M); dimethindene maleate (DiM, 2  $\mu$ M); and minocycline hydrochloride (MiH, 2  $\mu$ M). L-EPSCs were cultured on feeder cells or gelatin-coated plates in a base medium prepared as follows: DMEM/F12, KnockOut Serum

Replacement (KSR, 20%), L-glutamine (2 mM), NEAA (0.1 mM), and  $\beta$ -mercaptoethanol (0.1 mM). The base medium was supplemented with LIF (1,000 U/ml), CHIR99021 (3  $\mu$ M), PD0325901 (1  $\mu$ M), A-49259 (0.3  $\mu$ M), XAV939 (5  $\mu$ M), JNK inhibitor VIII (4  $\mu$ M), and SB203580 (10  $\mu$ M). For both EPSCs, the medium was changed daily, and cells were passaged every 3 d with accutase. Both EPSCs were cultured in feeder-free conditions for about 15 d, which equated to five passages before they were used for ATAC-seq, ChIP-seq, and RNA-seq experiments.

Oct4-FKBP mESCs, used for degron (dTAG) treatment based Oct4 protein depletion experiment, were cultured on 0.1% gelatin-coated plates in N2B27/2i/LIF medium formulated as: DMEM/F12 and Neurobasal (1:1 mix) medium supplemented with N2 (1 $\times$ ) and B27 (1 $\times$ ) supplements, NEAA (0.1 mM), LIF (1,000 U/ml), CHIR99021 (3  $\mu$ M), and PD0325901 (1  $\mu$ M).

All the cells were cultured in an incubator with 5% CO<sub>2</sub> at 37°C. Detailed reagents and tools list is provided in Table S6.

### RNA-seq and data analysis

Total RNA from ESCs and EPSCs replicates were extracted using Trizol according to the manufacturer's protocol. RNA quality was evaluated by Agilent 2100 BioAnalyzer. Total RNA from each sample was isolated and used to prepare RNA-seq libraries. RNA-seq libraries were prepared manually using Universal Plus mRNA-Seq with NuQuant kit, according to the manufacturer's protocol. For each sample, 500 ng total RNA was used to isolate mRNA via poly(A) selection. Captured mRNA was washed, fragmented, and primed with a mix of random oligo(dT) primers. After cDNA synthesis, ends were repaired and ligated with Unique Dual Index (UDI) adaptor pairs. Libraries were amplified by 14 PCR cycles and purified with AMPure XP beads, sequenced on the NovaSeq 6000 platform with 150 bp paired-end read length with Novogene.

RNA-seq reads quality assessment and adaptor trimming of fastq files were performed using TrimGalore v.0.6.4, retaining reads with a minimum length of 60 and a minimum Phred score of 20. Processed reads were mapped against the mouse genome (mm9/NCBIM37.67) and sorted by coordinate using STAR v2.7.9a. The number of reads per gene was counted using htseq v.0.11.2, providing the genome annotation (GTF format) from the NCBIM37.67 mouse genome. Differential gene expression was analyzed using DESeq2 v.4.1.1 R package. Genes not expressed in all samples (rowSums  $\leq$  1) were filtered out from the analysis. Correlation plots and PCA were performed on vst-transformed values (variance stabilized transformation; implemented in DESeq2 package). Genes were considered differentially expressed if they had an FDR value of < 0.05 and a log<sub>2</sub> fold-change >  $\pm$ 0.6 unless otherwise indicated.

Gene Set Enrichment Analysis (GSEA v.4.2.2) was used to assess the ontology terms enriched in each sample using the C5 Gene Ontology Gene Set Database (v.7.5.1). Only significant genes (FDR < 0.05 and Log<sub>2</sub>FC >  $\pm$ 0.6) were selected for this analysis. Briefly, P-values were calculated based on 1,000 permutations, with permutation type set to gene\_set. Enrichment analysis was set to *weighted* for the enrichment score calculation, and log<sub>2</sub> Ratio\_of\_Classes was used for gene ranking. Only GO terms with NOM P-val < 0.05 were retained for further analysis. Gene



Ontology terms heat map, volcano plots, PCA, and correlation plots were plotted using R software.

## ATAC-seq and data analysis

The ATAC-seq libraries of ESCs and both EPSCs were prepared in technical replicates as previously described (Buenrostro et al, 2013, 2015) with minor modification. Briefly,  $5 \times 10^6$  cells were lysed by lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl<sub>2</sub>, and 0.15% NP-40) for 10 min on ice to prepare the nuclei. Immediately after lysis, nuclei were spun down at 500g for 5 min. Next, the pellet was incubated with the Tn5 transposase and tagmentation buffer at 37°C for 30 min (Vazyme Biotech). After the tagmentation, the stop buffer was added directly into the reaction to end the tagmentation. PCR was performed to amplify the library for 15 cycles using the following PCR conditions: 72°C for 3 min; 98°C for 30 s; and thermocycling at 98°C for 15 s, 60°C for 30 s, and 72°C for 3 min; following by 72°C for 5 min. After the PCR reaction, libraries from 200 to 700 bp were purified using gel extraction before sequencing. NextGen sequencing was performed by NovaSeq 6000 platform with 150 paired-end reads.

ATAC-seq reads were processed as previously described (Yang et al, 2020). Briefly, sequencing reads were aligned to mouse genome (mm9) using the bowtie2 (v2.3.5) program, with parameters -X 2000 -no-mixed. Aligned reads were filtered by samtools (v1.10) program with parameters -F 0x04 -f 0x02 -q 20. ATAC-seq peaks were determined by the MACS program (v2.2.7) with default settings. Differential peak accessibility was determined using the Diffbind tool v3.4.11 (Ross-Innes et al, 2012). Briefly, reads were counted over each peak, normalized, and differential analysis was performed using DESeq2. All peaks ( $t_h = 1$ ) were extracted for further analysis. Peaks were considered differentially accessible if they had a value of FDR < 0.05 and a log<sub>2</sub> fold-change > 0.6 and < -0.6. Peak filtering and downstream analysis were performed using bedtools software v2.27.1, including peak intersection and coverage depth computing (bedtools intersect and bedtools coverage functions, respectively). Motif analysis was performed using HOMER v4.11.1 (findMotifsGenome.pl script). Briefly, the top 15 most significant known motifs (ranked by FDR value) were selected in all comparisons and merged. Motifs were then separated into several groups based on their sample enrichment and shown in Figs 2E and S2D. Peak genomic distribution was assessed by homer annotatePeaks.pl script (-annStats option). Log<sub>2</sub> (obs/exp) ratio was plotted in Fig 2C. Motif fraction enrichment in each sample was computed using HOMER with annotatePeaks.pl function. Briefly, the identified 88 motif types (Fig 2E) that present differential accessibility between mESCs and EPSCs were downloaded from the homer motif database (<http://homer.ucsd.edu/homer/motif/HomerMotifDB/homerResults.html>) and merged into a single file. A merged motif file was provided as input to HOMER using the annotatePeaks function (-m option). Motif counts were calculated, and the motif fraction was obtained by normalizing with the peak numbers in each accessibility group. The most enriched motifs

in mESC and both EPSCs were plotted in Fig 2F. ATAC-seq peak annotation was performed using the homer annotatePeaks.pl function to assign the nearest gene name to the peaks. Enrichment heat maps in Figs 2A, S2C, and S3C were plotted using EaSeq v.1.111, ranked by read coverage. Correlation and volcano plots were plotted using R software.

## ChIP-seq and data analysis

H3K4me1 and H3K27ac ChIP-seq experiments were performed in replicates as described (Ding et al, 2015). One million cells were used for each sample. Massively parallel sequencing was performed with the Illumina NovaSeq 6000 according to the manufacturer's protocol, and pair-end 150 bp length reads were produced. FastQC was used to check the sequencing quality. ChIP-seq reads were aligned to the mouse genome mm9 using the bowtie2 (v2.3.5) program, with parameters -X 1000 -no-mixed -no-discordant. The mapped reads were sorted and converted to a binary bam file using SAMtools (v1.10). ChIP-seq peaks were determined by the MACS program (v2.2.7) with the -broad option, using input as the control data. Differential peak enrichment was determined using Diffbind software (similar as in the "ATAC-seq and data analysis" section). Peak annotation was performed using HOMER annotatePeaks.pl script to assign nearby genes to each peak.

## SILAC-MS

The SILAC-MS procedure is illustrated in Fig 4A. Briefly, D-EPSCs were cultured in SILAC Light (Lys0, Arg0), L-EPSCs were cultured in SILAC Medium (Lys4, Arg6), and ESCs were cultured in SILAC Heavy (Lys8, Arg10) media for at least five passages. Cell lysates of each population were equally mixed for the following steps. Protein lysates were dissolved in 8M Urea buffer, followed by in-gel tryptic digestion and liquid chromatography-tandem mass spectrometry (LC-MS/MS) using an Orbitrap-Velos mass spectrometer. Proteome Discoverer Software with mouse proteome was used for protein quantification and identification. The relative intensities of heavy, medium, and light fraction for each protein were exported for further analysis.

## Western blot

Whole-cell protein extracts were isolated from the cultured cells using RIPA lysis buffer (NC9193720; Boston BioProducts) supplemented with protease inhibitor cocktail (P8340; Sigma-Aldrich) and phosphatase inhibitor cocktail (78428; Thermo Fisher Scientific). Blots were incubated in 2% BSA/TBST at room temperature for 1 h, and then they were incubated with the corresponding antibodies in 5% skimmed milk powder/TBST at 4°C overnight. Secondary antibodies were anti-rabbit IgG, HRP-linked antibody (1:5,000; R&D System, HAF008) and anti-mouse IgG, HRP-linked antibody (1:5,000; Cell Signaling Technology, 70765), which were incubated for 1 h at room temperature while shaking. The blots were developed using XRAY FILM (Cat. no. XAR ALF 2025; LabScientific) in a dark room.



## Data Availability

Sequencing data (ATAC-seq, ChIP-seq, and RNA-seq) that supports the findings of this study are available in Gene Expression Omnibus. Accession number for the study is [GSE201305](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201305).

## Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lsa.202201608>.

## Acknowledgements

We thank Drs. Amander T Clark, Richard A Young, Austin Smith, and Hitoshi Niwa for providing Ttap2a/c dKO, Oct4-degrom ESCs, ZHBTc4 ESCs, and 2T522C ESCs, respectively. This work in the Wang laboratory is funded by grants from the National Institutes of Health (R01HD095938 and R01HD097268) and by contracts from New York State Stem Cell Science (NYSTEMC35583GG). A Fuentes-Iglesias is the recipient of a fellowship from the MINECO of Spain (BES-2017-082007).

## Author Contributions

V Malik: data curation, formal analysis, validation, investigation, methodology, and writing—original draft, review, and editing.  
R Zang: data curation, investigation, methodology, and writing—review and editing.  
A Fuentes-Iglesias: data curation, formal analysis, investigation, and writing—review and editing.  
X Huang: data curation, formal analysis, investigation, methodology, and writing—review and editing.  
D Li: data curation, investigation, methodology, and writing—review and editing.  
M Fidalgo: writing—review and editing.  
H Zhou: data curation, investigation, methodology, project administration, and writing—review and editing.  
J Wang: conceptualization, supervision, funding acquisition, investigation, visualization, methodology, project administration, and writing—original draft, review, and editing.

## Conflict of Interest Statement

The authors declare that they have no conflict of interest.

## References

Adachi K, Nikaio I, Ohta H, Ohtsuka S, Ura H, Kadota M, Wakayama T, Ueda HR, Niwa H (2013) Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. *Mol Cell* 52: 380–392. doi:10.1016/j.molcel.2013.09.002

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315–326. doi:10.1016/j.cell.2006.02.041

Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175: 1842–1855.e16. doi:10.1016/j.cell.2018.10.042

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10: 1213–1218. doi:10.1038/nmeth.2688

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 109: 21.29.1–21.29.9. doi:10.1002/0471142727.mb2129s109

Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113: 643–655. doi:10.1016/s0092-8674(03)00392-1

Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* 450: 1230–1234. doi:10.1038/nature06403

Das S, Jena S, Levasseur DN (2011) Alternative splicing produces Nanog protein variants with different capacities for self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* 286: 42690–42703. doi:10.1074/jbc.M111.290189

Ding J, Huang X, Shao N, Zhou H, Lee DF, Faiola F, Fidalgo M, Guallar D, Saunders A, Shihah PV, et al (2015) Tbx3 coordinates epigenetic control of super-enhancer activity in pluripotency and reprogramming. *Cell Stem Cell* 16: 653–668. doi:10.1016/j.stem.2015.04.001

Du Y, Wang T, Xu J, Zhao C, Li H, Fu Y, Xu Y, Xie L, Zhao J, Yang W, et al (2019) Efficient derivation of extended pluripotent stem cells from NOD-scid Il2rg<sup>-/-</sup> mice. *Protein Cell* 10: 31–42. doi:10.1007/s13238-018-0558-2

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292: 154–156. doi:10.1038/292154a0

Fidalgo M, Huang X, Guallar D, Sanchez-Priego C, Valdes VJ, Saunders A, Ding J, Wu WS, Clavel C, Wang J (2016) Zfp281 coordinates opposing functions of Tet1 and Tet2 in pluripotent states. *Cell Stem Cell* 19: 355–369. doi:10.1016/j.stem.2016.05.025

Gao X, Nowak-Imialek M, Chen X, Chen D, Herrmann D, Ruan D, Chen ACH, Eckersley-Maslin MA, Ahmad S, Lee YL, et al (2019) Establishment of porcine and human expanded potential stem cells. *Nat Cell Biol* 21: 687–699. doi:10.1038/s41556-019-0333-2

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589. doi:10.1016/j.molcel.2010.05.004

He J, Fu X, Zhang M, He F, Li W, Abdul MM, Zhou J, Sun L, Chang C, Li Y, et al (2019) Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat Commun* 10: 34. doi:10.1038/s41467-018-08006-y

Hu Y, Yang Y, Tan P, Zhang Y, Han M, Yu J, Zhang X, Jia Z, Wang D, Li Y, et al (2022) Induction of mouse totipotent stem cells by a defined chemical cocktail. *Nature* doi:10.1038/s41586-022-04967-9

Hutchins AP, Yang Z, Li Y, He F, Fu X, Wang X, Li D, Liu K, He J, Wang Y, et al (2017) Models of global gene expression define major domains of cell type and tissue identity. *Nucleic Acids Res* 45: 2354–2367. doi:10.1093/nar/gkx054

Iturbide A, Ruiz Tejada Segura ML, Noll C, Schorpp K, Rothenaigner I, Ruiz-Morales ER, Lubatti G, Agami A, Hadian K, Scialdone A, et al (2021) Retinoic acid signalling is critical during the totipotency window in early mammalian development. *Nat Struct Mol Biol* 28: 521–532. doi:10.1038/s41594-021-00590-w



- Kruger L, Nowak-Imialek M, Kristiansen Y, Herrmann D, Petersen B, Denner J (2021) Unexpected low expression of porcine endogenous retroviruses (PERVs) in porcine expanded potential stem cells (EPSCs). *Virus Res* 294: 198295. doi:10.1016/j.virusres.2021.198295
- Kuckenberger P, Buhl S, Woynecki T, van Furden B, Tolkunova E, Seiffe F, Moser M, Tomilin A, Winterhager E, Schorle H (2010) The transcription factor TCFAP2C/AP-2y cooperates with CDX2 to maintain trophoblast formation. *Mol Cell Biol* 30: 3310–3320. doi:10.1128/mcb.01215-09
- Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K (2016) An interactive environment for agile analysis and visualization of ChIP-seq data. *Not Struct Mol Biol* 23: 349–357. doi:10.1038/nsmb.3180
- Li D, Yang J, Huang X, Zhou H, Wang J (2022) eIF4A2 targets developmental potency and histone H3.3 transcripts for translational control of stem cell pluripotency. *Sci Adv* 8: eabm0478. doi:10.1126/sciadv.abm0478
- Li H, Zhao C, Xu J, Xu Y, Cheng C, Liu Y, Wang T, Du Y, Xie L, Zhao J, et al (2019a) Rapid generation of gene-targeted EPS-derived mouse models through tetraploid complementation. *Protein Cell* 10: 20–30. doi:10.1007/s13238-018-0556-1
- Li R, Zhong C, Yu Y, Liu H, Sakurai M, Yu L, Min Z, Shi L, Wei Y, Takahashi Y, et al (2019b) Generation of blastocyst-like structures from mouse embryonic and adult cell cultures. *Cell* 179: 687–702.e18. doi:10.1016/j.cell.2019.09.029
- Liu B, Chen S, Xu Y, Lyu Y, Wang J, Du Y, Sun Y, Liu H, Zhou H, Lai W, et al (2021) Chemically defined and xeno-free culture condition for human extended pluripotent stem cells. *Not Commun* 12: 3017. doi:10.1038/s41467-021-23320-8
- Lodato MA, Ng CW, Wamstad JA, Cheng AW, Thai KK, Fraenkel E, Jaenisch R, Boyer LA (2013) SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* 9: e1003288. doi:10.1371/journal.pgen.1003288
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. doi:10.1186/s13059-014-0550-8
- Lu JY, Shao W, Chang L, Yin Y, Li T, Zhang H, Hong Y, Percharde M, Guo L, Wu Z, et al (2020) Genomic repeats categorize genes with distinct functions for orchestrated regulation. *Cell Rep* 30: 3296–3311.e5. doi:10.1016/j.celrep.2020.02.048
- Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487: 57–63. doi:10.1038/nature11244
- Malik V, Glaser LV, Zimmer D, Velychko S, Weng M, Holzner M, Arend M, Chen Y, Srivastava Y, Veerapandian V, et al (2019) Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. *Not Commun* 10: 3477. doi:10.1038/s41467-019-11054-7
- Malik V, Wang J (2022) Pursuing totipotency: Authentic totipotent stem cells in culture. *Trends Genet* 38: 632–636. doi:10.1016/j.tig.2022.03.012
- Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78: 7634–7638. doi:10.1073/pnas.78.12.7634
- Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, et al (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Not Cell Biol* 9: 625–635. doi:10.1038/ncb1589
- Mazid MA, Ward C, Luo Z, Liu C, Li Y, Lai J, Wu L, Li J, Jia W, Jiang Y, et al (2022) Rolling back human pluripotent stem cells to an eight-cell embryonic stage. *Nature* 605: 315–324. doi:10.1038/s41586-022-04625-0
- Nagaraj R, Sharpley MS, Chi F, Braas D, Zhou Y, Kim R, Clark AT, Banerjee U (2017) Nuclear localization of mitochondrial TCA cycle enzymes as a critical step in mammalian zygotic genome activation. *Cell* 168: 210–223.e11. doi:10.1016/j.cell.2016.12.026
- Niwa H, Miyazaki JI, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Not Genet* 24: 372–376. doi:10.1038/74199
- Pastor WA, Liu W, Chen D, Ho J, Kim R, Hunt TJ, Lukianchikov A, Liu X, Polo JM, Jacobsen SE, et al (2018) TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Not Cell Biol* 20: 553–564. doi:10.1038/s41556-018-0089-0
- Posfai E, Lanner F, Mulas C, Leitch HG (2021a) All models are wrong, but some are useful: Establishing standards for stem cell-based embryo models. *Stem Cell Rep* 16: 1117–1141. doi:10.1016/j.stemcr.2021.03.019
- Posfai E, Schell JP, Janiszewski A, Rovic I, Murray A, Bradshaw B, Yamakawa T, Pardon T, El Bakkali M, Talon I, et al (2021b) Evaluating totipotency using criteria of increasing stringency. *Not Cell Biol* 23: 49–60. doi:10.1038/s41556-020-00609-2
- Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481: 389–393. doi:10.1038/nature10730
- Shen H, Yang M, Li S, Zhang J, Peng B, Wang C, Chang Z, Ong J, Du P (2021) Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* 184: 2843–2859.e20. doi:10.1016/j.cell.2021.04.020
- Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Not Biotechnol* 32: 171–178. doi:10.1038/nbt.2798
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484: 339–344. doi:10.1038/nature10960
- Sozen B, Cox AL, De Jonghe J, Bao M, Hollfelder F, Glover DM, Zernicka-Goetz M (2019) Self-organization of mouse stem cells into an extended potential blastoid. *Dev Cell* 51: 698–712.e8. doi:10.1016/j.devcel.2019.11.014
- Tan T, Wu J, Si C, Dai S, Zhang Y, Sun N, Zhang E, Shao H, Si W, Yang P, et al (2021) Chimeric contribution of human extended pluripotent stem cells to monkey embryos ex vivo. *Cell* 184: 2020–2032.e14. doi:10.1016/j.cell.2021.03.020
- Taubenschmid-Stowers J, Rostovskaya M, Santos F, Ljung S, Argelaguet R, Krueger F, Nichols J, Reik W (2022) 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell* 29: 449–459.e6. doi:10.1016/j.stem.2022.01.014
- Tomikawa J, Penfold CA, Kamiya T, Hibino R, Kosaka A, Anzai M, Matsumoto K, Miyamoto K (2021) Cell division- and DNA replication-free reprogramming of somatic nuclei for embryonic transcription. *iScience* 24: 103290. doi:10.1016/j.isci.2021.103290
- Wang Q, Sun D, Liang Z, Wang J, Zhong X, Lyu Y, Cao J, Lin Z, Du Y, Miao Z, et al (2020) Generation of human hepatocytes from extended pluripotent stem cells. *Cell Res* 30: 810–813. doi:10.1038/s41422-020-0293-x
- Wen X, Lin Z, Wu H, Cao L, Fu X (2022) Zfp281 inhibits the pluripotent-to-totipotent state transition in mouse embryonic stem cells. *Front Cell and Dev Biol* 10: 879428. doi:10.3389/fcell.2022.879428
- Xu Y, Zhao J, Ren Y, Wang X, Lyu Y, Xie B, Sun Y, Yuan X, Liu H, Yang W, et al (2022) Derivation of totipotent-like stem cells with blastocyst-like structure forming potential. *Cell Res* 32: 513–529. doi:10.1038/s41422-022-00668-0
- Yang F, Huang X, Zang R, Chen J, Fidalgo M, Sanchez-Priego C, Yang J, Caichen A, Ma F, Macfarlan T, et al (2020) DUX-miR-344-ZMYM2-mediated activation of MERVL LTRs induces a totipotent 2C-like state. *Cell Stem Cell* 26: 234–250.e7. doi:10.1016/j.stem.2020.01.004
- Yang J, Ryan DJ, Wang W, Tsang JCH, Lan G, Masaki H, Gao X, Antunes L, Yu Y, Zhu Z, et al (2017a) Establishment of mouse expanded potential stem cells. *Nature* 550: 393–397. doi:10.1038/nature24052

- Yang M, Yu H, Yu X, Liang S, Hu Y, Luo Y, Izsvak Z, Sun C, Wang J (2022) Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell* 29: 400–418.e13. doi:10.1016/j.stem.2022.01.010
- Yang Y, Liu B, Xu J, Wang J, Wu J, Shi C, Xu Y, Dong J, Wang C, Lai W, et al (2017b) Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell* 169: 243–257.e25. doi:10.1016/j.cell.2017.02.005
- Zhang J, Ratanasirintrawoot S, Chandrasekaran S, Wu Z, Ficarro SB, Yu C, Ross CA, Cacchiarelli D, Xia Q, Seligson M, et al (2016) LIN28 regulates stem cell metabolism and conversion to primed pluripotency. *Cell Stem Cell* 19: 66–80. doi:10.1016/j.stem.2016.05.009
- Zhang J, Zhao J, Dahan P, Lu V, Zhang C, Li H, Teitell MA (2018a) Metabolism in pluripotent stem cells and early mammalian development. *Cell Metab* 27: 332–338. doi:10.1016/j.cmet.2018.01.008
- Zhang Y, Xiang Y, Yin Q, Du Z, Peng X, Wang Q, Fidalgo M, Xia W, Li Y, Zhao ZA, et al (2018b) Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat Genet* 50: 96–105. doi:10.1038/s41588-017-0003-x
- Zhao L, Gao X, Zheng Y, Wang Z, Zhao G, Ren J, Zhang J, Wu J, Wu B, Chen Y, et al (2021) Establishment of bovine expanded potential stem cells. *Proc Natl Acad Sci U S A* 118: e2018505118. doi:10.1073/pnas.2018505118
- Zheng R, Geng T, Wu DY, Zhang T, He HN, Du HN, Zhang D, Miao YL, Jiang W (2021) Derivation of feeder-free human extended pluripotent stem cells. *Stem Cell Rep* 16: 1686–1696. doi:10.1016/j.stemcr.2021.06.001
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523. doi:10.1038/s41467-019-09234-6
- Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M, et al (2018) The interaction landscape between transcription factors and the nucleosome. *Nature* 562: 76–81. doi:10.1038/s41586-018-0549-5



**License:** This article is available under a Creative Commons License (Attribution 4.0 International), as described at <https://creativecommons.org/licenses/by/4.0/>.