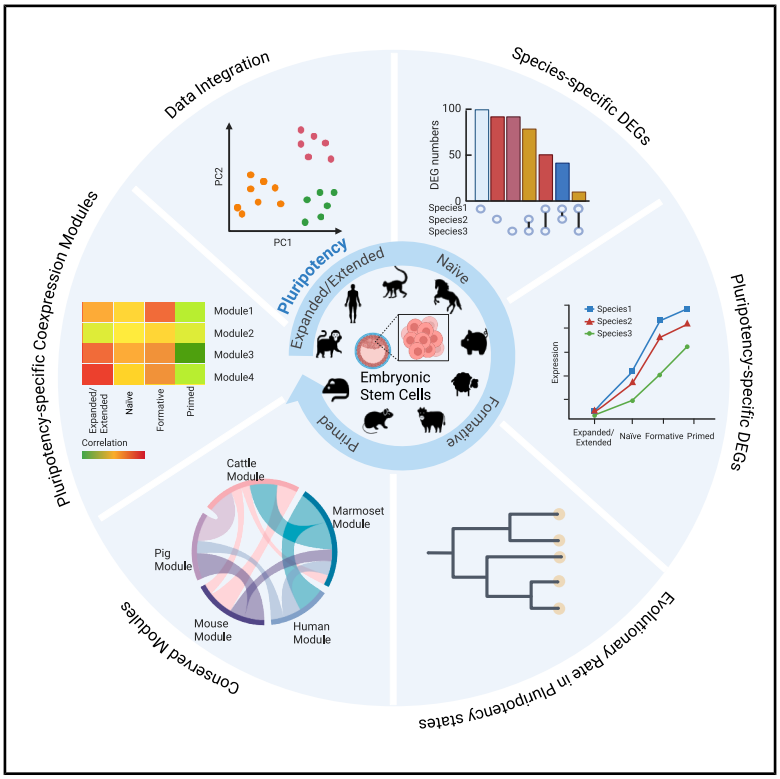
iScience

Article

Comparative transcriptomic analysis of embryonic stem cells across mammalian species

Graphical abstract



Highlights

- 1:1 ortholog-based analysis of four ESC pluripotency states across nine mammals
- Transcriptomic variation across species exceeds variation across ESC states
- *ZBTB10* in naive and *OTX2*, *ETV4,* and *EPHA1* in primed states are conserved markers
- Primed-state networks were conserved and show strong purifying selection

Authors

Yifei Fang, Yue Su, Richard Meisel, ..., Xiuchun (Cindy) Tian, Young Tang, Jingyue (Ellie) Duan

Correspondence

xiuchun.tian@uconn.edu (X.T.), youngtang@nwafu.edu.cn (Y.T.), jd774@cornell.edu (J.D.)

In brief

Evolutionary Developmental Biology; Molecular network; Stem cells research; Transcriptomics

Fang et al., 2026, iScience *29*, 114542
January 16, 2026
https://doi.org/10.1016/j.isci.2025.114542

CellPress

iScience

CellPress
OPEN ACCESS

Article

Comparative transcriptomic analysis of embryonic stem cells across mammalian species

Yifei Fang,[1] Yue Su,[2,7] Richard Meisel,[3] James R. Walters,[4] Tony Gamble,[5] Eric C. Randolph,[6] Xingtan Yu,[1] Meihong Shi,[1] Guangsheng Li,[1] Jingzhi Zhang,[1] IISAGE Consortium, Xiuchun (Cindy) Tian,[7,*] Young Tang,[2,*] and Jingyue (Ellie) Duan[1,8,*]

[1]Department of Animal Science, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14853, USA
[2]Shaanxi Centre of Stem Cells Engineering & Technology, Key Laboratory of Livestock Biology, College of Veterinary Medicine, Northwest A&F University, Yangling, Shaanxi, China
[3]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA
[4]Department of Ecology and Evolutionary Biology, The University of Kansas, Lawrence, KS, USA
[5]Department of Biological Sciences, Marquette University, Milwaukee, WI 53201, USA
[6]Department of Biology, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[7]Department of Animal Science, Institute for Systems Genomics, University of Connecticut, Storrs, CT, USA
[8]Lead contact
*Correspondence: xiuchun.tian@uconn.edu (X.T.), youngtang@nwafu.edu.cn (Y.T.), jd774@cornell.edu (J.D.)
https://doi.org/10.1016/j.isci.2025.114542

SUMMARY

Pluripotency, the ability of cells to self-renew and differentiate into all body lineages, is vital for mammalian early development. This study presented a comprehensive comparative transcriptomic analysis of embryonic stem cells across multiple mammalian species, defining their progression through expanded/extended, naive, formative, and primed pluripotency states. Our findings revealed both conserved and species-specific mechanisms underlying pluripotency regulation. We also emphasized the limitations of existing state-specific markers and their limited cross-species applicability, while identifying *de novo* pluripotency markers. Despite variability in gene expression dynamics, gene co-expression networks showed remarkable conservation across species. Among pluripotency states, the primed state demonstrated the highest conservation, evidenced by shared markers, preserved gene networks, and stronger selective pressures acting on its genes. These findings provide critical insights into the evolution and regulation of pluripotency, laying a foundation for refining stem cell models to enhance their translational potential in regenerative medicine, agriculture, and conservation biology.

INTRODUCTION

Embryonic stem cells (ESCs) are unique in their ability to self-renew and plasticity to differentiate into diverse cell types, offering a powerful *in vitro* model for studying developmental biology and advancing regenerative medicine.[1] These ESCs exhibit pluripotent states with distinct differentiation potentials and molecular profiles corresponding to cells of the specific stages of early embryonic development.[2] For example, mouse ESCs were initially established in the naive pluripotency state (nESCs), representing the inner cell mass of the pre-implantation stage embryos.[3] Subsequent research has derived additional pluripotency states, including the primed state (epiblast stem cells, EpiSCs), which corresponds to post-implantation stage cells,[4] and the formative state (formative stem cells, FSCs), defined as the intermediate transition between naive and primed states.[5] More recently, the expanded or extended state (extended pluripotency stem cells, EPSCs) has been established, representing an earlier embryonic stage with broader developmental lineage potential, such as the ability to generate both embryonic and extraembryonic tissues, thereby positioning them as pluripotent cells with certain totipotent-like characteristics.[6] Each pluripotency state is characterized by distinct differentiation potential and functional genomic profiles, serving as *in vitro* models to map ESC properties to specific embryonic stages and uncover the molecular mechanisms underlying the early embryonic programming.

Although these pluripotency states are well characterized in the mouse model, generating ESCs at equivalent states in other mammals, such as domestic animals like pigs, cattle, and sheep, remains challenging. Specifically, the naive state exhibits significant cross-species variability in culture conditions.[7] In contrast, the maintenance of the primed state is more consistent across species. For instance, WNT inhibition has been shown to effectively sustain EpiSCs in a range of species, including humans,[8] rats,[9] cattle,[10] pigs,[11] sheep,[11] and crab-eating macaques.[12] Moreover, comparative studies have identified conserved signaling pathways that stabilize the maintenance of a primed pluripotency state across species, such as elevating fibroblast growth factor (FGF) and transforming growth factor (TGF) pathways while inhibiting WNT signaling.[13] In contrast, the maintenance of naive pluripotency relies on species-specific

signaling mechanisms. For instance, TGF-β signaling is dispensable for maintaining mouse naive ESCs, but it is necessary for sustaining naive pluripotency in human cells.[14] The activation of distinct pathways highlights fundamental differences in the molecular regulation of pluripotency across species, which likely arise as evolutionary divergence associated to early developmental programming.[15] These differences are shaped by species-specific requirements for embryogenesis, such as differences in implantation strategies, developmental rates, and embryo-maternal interactions.[16] Such species-specific differences complicate the application of ESC derivation protocols across species and highlight the need to develop methodologies for identifying both universal and species-specific molecular regulatory mechanisms underlying the developmental and functional features of ESCs.

Historically, ESC lines have often been derived independently by different research groups, each focusing on a single species or a specific pluripotency state. This has limited opportunities for direct cross-species and cross-pluripotency state comparisons. Recent advances in comparative transcriptomics, supported by the availability of RNA sequencing (RNA-seq) data from various species and pluripotency states, now enable systematic investigation into conserved and species-specific gene expression patterns and regulatory mechanisms in ESCs. Comparative transcriptomics can facilitate the identification of core regulatory networks, explore evolutionary differences in gene expression, and discover new state-specific pluripotency markers that may function universally across species or exhibit distinct species-specific roles in stem cell biology.

In this study, we conducted a comprehensive comparative transcriptomics analysis of four ESC pluripotency states across nine mammalian species (three primates, four ungulates, and two rodents), using transcriptomic data from 28 research projects. We identified both common and species-specific differentially expressed genes (DEGs), enriched GO pathways between pluripotent states, assessed known state-specific markers, and discovered *de novo* pluripotency markers for each species. To further explore conserved and divergent gene regulation, we employed weighted gene co-expression network analysis (WGCNA) to construct species-specific gene networks and performed preservation analysis to identify networks conserved across species within specific pluripotency states. Additionally, our transcriptome-based evolutionary analysis revealed varying levels of selective constraints across different pluripotency states. Overall, these findings provide critical insights into the evolutionary conservation and divergence of gene regulatory mechanisms underlying pluripotency, offering a foundation for future studies to refine species-specific stem cell models and enhance cross-species applications in regenerative medicine and developmental biology.

RESULTS

Transcriptome integration and online RNA-seq datasets normalization

To achieve a comprehensive analysis, we curated 120 publicly available RNA-seq samples from 28 research projects, including four ESC types, EPSCs, nESCs, FSCs, and EpiSCs, across nine mammalian species: human, mouse, rat, marmoset, crab-eating macaque, cattle, pig, sheep, and horse (Figures 1A and 1B, Table S1). The pluripotency states of each sample were characterized in its original research based on criteria such as ESC colony morphology, self-renewal capacity, differentiation potential, *in vitro*/*in vivo* developmental capability, and surface marker expression. With the exception of mouse and human, most species currently have only a single dataset for each pluripotent state. To ensure balanced representation and minimize potential batch effects from combining data generated by different laboratories, we mostly analyzed one dataset per state per species. Moreover, for consistency in cross-species comparisons, we use "naive" as an inclusive term encompassing both the "naive" and "ground" states, as most datasets outside mouse and rat do not explicitly distinguish this pluripotency state.

All raw transcriptome data were processed using a uniform bioinformatics pipeline (Figure 1C and Table S1). Specifically, RNA-seq reads were first aligned to species-specific reference genomes (Table S2), and expression read counts were quantified. To enable cross-species comparative analysis, these samples were subsequently combined using the 9,942 1-to-1 orthologous genes (i.e., genes that have only one ortholog copy among studied species) across the nine mammals. These RNA-seq data originate from diverse research projects and may be influenced by differences in sample conditions, library construction, sequencing procedures, and other unknown factors. To detect and correct for such factors, including unrecognized "hidden" biases, we applied $\log_2$-transformed transcripts per million ($\log_2$ TPM) normalization, followed by surrogate variable analysis (SVA-log-TPM) (Figure 1C). This SVA-log-TPM normalization greatly improved the Pearson correlation within samples of the same pluripotency state in each species, regardless of their research project origins (Figure 1D), confirming the reliability of our bioinformatic pipeline in handling online curated datasets for analyzing ESC transcriptomic patterns.

Principal Component Analysis (PCA) revealed that RNA-seq samples primarily clustered by species rather than pluripotency states (Figure 1E), reflecting a stronger species effect than the pluripotency states. These findings align with previous *in vivo* studies showing that cross-species embryo samples from pre- and post-implantation stages cluster more strongly by species than by developmental stages.[17] Variance partitioning analysis confirmed that species divergence contributes more to transcriptional variances in PCA than pluripotency states among these samples (Figure S1A). We further identified highly variable pluripotency-specific genes, such as *DUSP6* and *SPRY1* (Figure S1B), which are key components of the ERK[18] and FGF[19] signaling pathways and are known to play critical roles in maintaining stem cell pluripotency. Notably, *DUSP6* has also been recognized as a novel pluripotency-related gene associated with the transition from naive to primed states through comparative studies of mammalian embryos.[17] Moreover, our PCA results showed a consistent pattern of naïve-to-primed distribution along PC1 was observed within each species cluster (Figure 1E), highlighting the conservation of transcription dynamics across pluripotency states among mammalian species.

To determine the impact of ESC derivation protocols on ESC transcriptome, we compared porcine, murine, and crab-eating

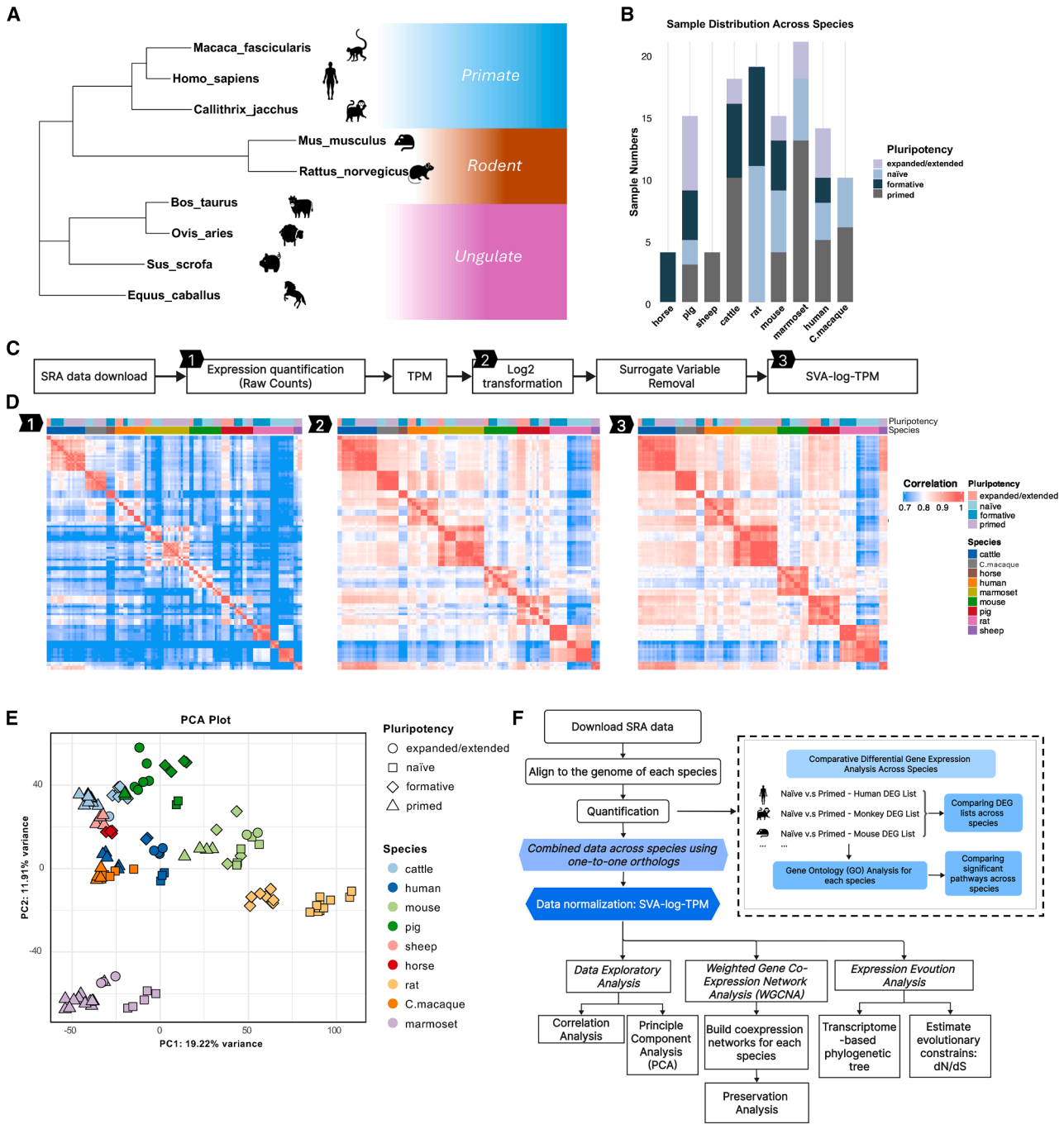


Figure 1. Overview of dataset integration and analytical pipeline

(A) Phylogenetic tree representing the nine mammalian species included in the analysis.

(B) Distribution of embryonic stem cell samples with varying pluripotency levels across species.

(C) Data normalization workflow: Raw RNA-seq data were downloaded from the GEO database and processed through a unified pipeline, including alignment and quantification to obtain raw counts. TPM values were then calculated and log-transformed. To account for hidden batch effects and technical biases, Surrogate Variable Analysis (SVA) was applied, generating SVA-log-TPM values for downstream analysis.

(D) Pearson correlation coefficients comparing samples at different stages of normalization: (1) raw counts, (2) log2-TPM values, and (3) SVA-log-TPM values post-surrogate variable adjustment.

(E) Principal Component Analysis (PCA) of all samples based on SVA-log-TPM values; colors denote species, and shapes indicate pluripotency states.

(F) Detailed bioinformatics pipeline used in this study.

macaque ESCs with their corresponding *in vivo* pre- or post-implantation stage embryos (Figures S1C–S1E). Mouse ESCs aligned well with their corresponding developmental stages: EPSC and nESC clustered closer to pre-implantation embryos, EpiSCs aligned with post-implantation embryos, and FSC, representing an intermediate state, located between pre- and post-implantation stages (Figure S1C). However, the correspondence between *in vitro* and *in vivo* samples appears weaker in some species, which may reflect inconsistencies in classification standards across studies. In crab-eating macaque, nESCs and EpiSCs show only a modest trend toward their respective pre- or post-implantation embryonic stages, with two main clusters of ESCs and embryo samples in the PCA space (Figure S1E). Intriguingly, porcine EpiSCs and FSCs exhibited an inverse alignment with their expected stages: EpiSCs aligned with pre-implantation embryos, while FSCs clustered closer to the post-implantation stage embryos (Figure S1D). We further traced back to the original studies and found that this discrepancy arose because porcine EpiSCs were derived from pre-implantation embryos,[20] whereas FSCs were derived from post-implantation embryos.[21] These results suggested that ESC transcriptomes are strongly influenced by their derivation protocols, aligning more closely with the developmental stage of their embryonic origin. Moreover, the observed impacts of derivation protocols, coupled with species-specific differences, raise limitations of applying mouse model-based practices and standards for establishing new stem cell lines across diverse species, highlighting the need to establish both universal and species-specific derivation standards and pluripotency markers.

Gene expression comparison among pluripotency states

To minimize the impact of the inconsistencies in pluripotency characterization and assignment across species, we first identified species-specific DEGs during ESC pluripotency state transition by performing pairwise DEG analysis within species across four ESC states: naive vs. primed, expanded/extended vs. naive, naive vs. formative, and formative vs. primed. We then compared identified DEGs and Gene Ontology (GO) biological process terms across species that had both ESC types presented (Figure 1F, Tables S3 and S4).

nESCs vs. EpiSCs comparison

We compared DEGs between the nESCs and EpiSCs across six species, rat, marmoset, macaque, pig, mouse, and humans (Figure 2), and further assessed the expression pattern of these common DEGs across different pluripotency states in species with multiple ESC states, excluding horse and sheep, which each had only one available pluripotency state (Figure S5). Our results showed that the species-specific DEGs dominated the upregulated DEGs in nESCs, with 2,026 in mice and 1,512 in humans (Figure 2A). Despite this species specificity, we identified two DEGs, *ZBTB10* and *IHO1*, that were consistently upregulated in nESCs across all six species (Figures 2A and S5A). *ZBTB10*, a transcription factor (TF), interacts with core pluripotency factors *OCT4* and *SOX2,* playing a role in maintaining stem cell pluripotency.[22] Interestingly, *ZBTB10* was consistently upregulated in nESCs across six species examined yet also elevated in the formative states in cattle (Figure S5A), suggesting a potentially expanded role in bovine pluripotency regulation. *IHO1* is known to promote the DNA double-strand break formation,[23] a crucial mechanism for maintaining ESC genome stability.[24] However, due to the absence of a clearly defined 1:1 ortholog in several species, we were unable to assess *IHO1* expression patterns across species (Figure S5A).

GO analysis revealed significant enrichment of pathways across all species, including cell differentiation and cell development, suggesting that these pathways are fundamental to the core processes that maintain naive pluripotency and are evolutionary conserved (Figures 2B and 2C, Table S4). Moreover, we identified several mouse-specific enriched pathways, including the purine nucleotide metabolic process, peptide metabolic process, ubiquitin-dependent protein catabolic process, and stem cell population maintenance (Figure 2B).

Similarly, species-specific DEGs also dominated the upregulated DEGs in EpiSCs compared with nESCs, with 2,386 identified in mice, 1,640 in humans, and 1,066 in pigs (Figure 2D). Sixteen commonly upregulated genes in EpiSCs were identified across six species, including *OTX2, EPHA1, SLC2A1, ETV4,* and *SPRY4* (Figure 2D). Most of these common DEGs also show an upregulated pattern in primed ESCs in other mammalian species (Figure S5A). For example, *OTX2,* a key factor that helps maintain the EpiSC state by preventing neural differentiation,[25] was similarly elevated in cattle primed ESCs (Figure S5A). In contrast, *EPHA1*, a member of the Eph receptor tyrosine kinase family, which promotes pluripotency maintenance in human ESCs and iPSCs,[26] showed a downregulation primed state but was highly expressed in formative cells in cattle, indicating a species-specific pluripotency regulation (Figure S5A). Moreover, *SLC2A1* (*GLUT1*) supports pluripotency in ESCs by facilitating glucose uptake to sustain the high glycolytic activity essential for stem cell survival and self-renewal.[27] *ETV4*, together with *ETV5*, facilitates the exit from naive pluripotency and promotes epiblast maturation during early embryonic development.[28] *SPRY4* functions as a positive regulator of stemness in mouse ESCs by supporting proliferation, stem cell marker expression, and proper embryoid body formation.[29] GO analysis of EpiSCs upregulated DEGs revealed significantly enriched pathways for primed state ESCs across species, such as non-canonical Wnt signaling, BMP signaling, ERK and MAPK cascade, transmembrane receptor protein tyrosine kinase signaling, embryo development, and regulation of cell adhesion (Figures 2E and 2F, Table S4). These pathways are essential for lineage specification, cellular interactions, signaling, and developmental processes in primed state ESC. Their conservation across species suggests their evolutionary importance in regulating early development and ensuring proper establishment of a robust primed pluripotency state.

nESCs vs. EPSCs comparison

In the comparison of nESCs and EPSCs across human, marmoset, mouse, and pig (Figure S2), two commonly upregulated genes in nESCs, *DDIT4* and *BCL3,* were identified (Figures S2A and S5B). Both genes also showed elevated expression in cattle expanded state cells and macaque naive state cells (Figure S5B). *DDIT4* mediates HIF1α and mTOR

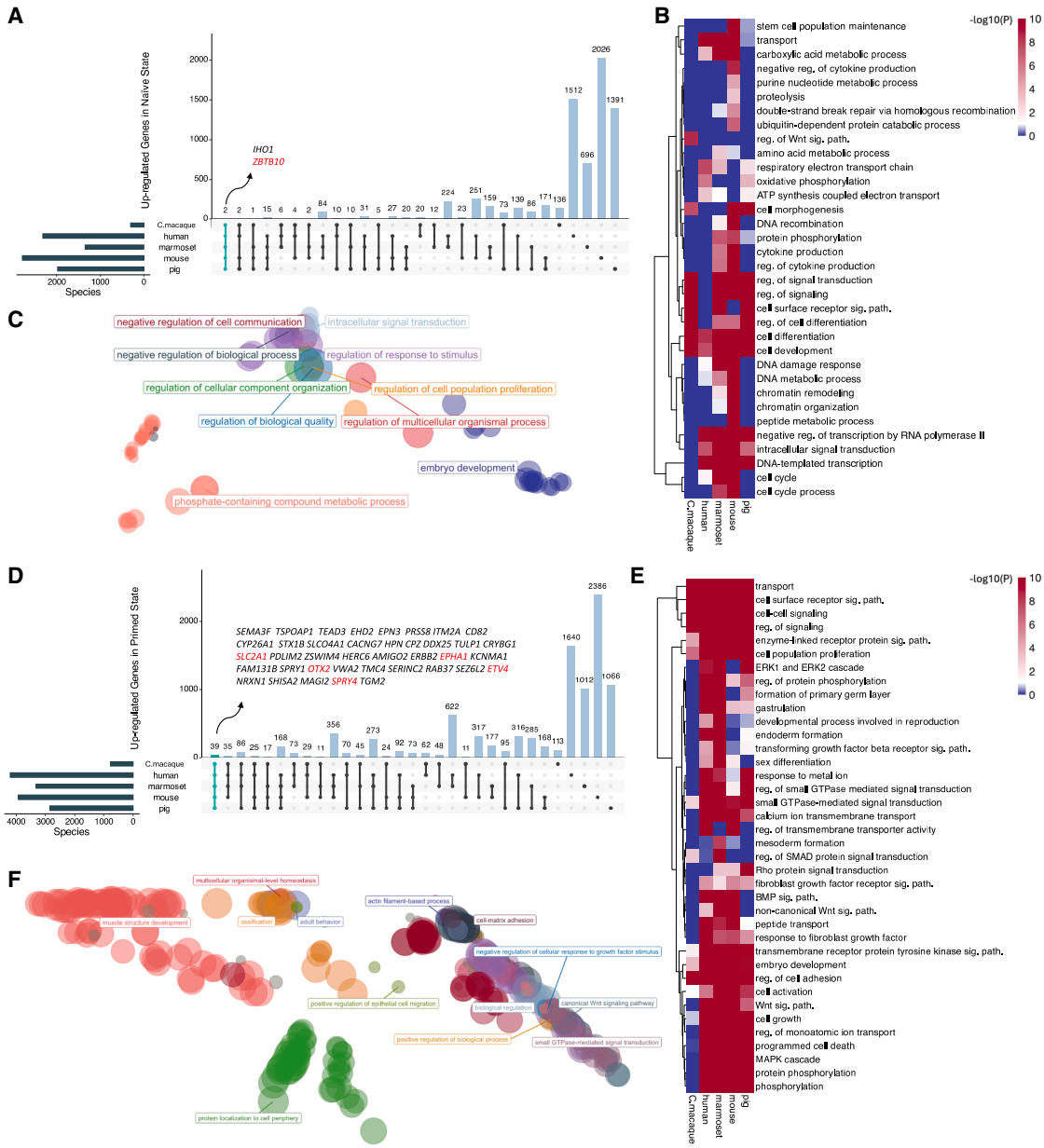


Figure 2. Comparative analysis of shared and species-specific genes and GO pathways in naive versus primed ESC states

(A) Up-regulated genes identified in the naive state across multiple species.

(B) Selected GO pathways significantly enriched by the up-regulated genes in the naive state across species.

(C) All conserved GO:BP pathways significantly enriched by up-regulated genes in the naive state across species, with terms clustered based on similarity. The color represents the group of similar GO terms based on their semantic similarity. The dot size reflects the number of genes associated with each GO term. The distance between points represents the similarity between terms. Only representative GO terms from each cluster are labeled on the plot.

(D) Up-regulated genes identified in the primed state across multiple species.

(E) Selected GO pathways significantly enriched by the up-regulated genes in the primed state across species.

(F) All conserved GO:BP pathways significantly enriched by up-regulated genes in the primed state across species, with terms clustered based on similarity. The color represents the group of similar GO terms based on their semantic similarity. The dot size reflects the number of genes associated with each GO term. The distance between points represents the similarity between terms. Only representative GO terms from each cluster are labeled on the plot.

signaling, which are critical for pluripotency,[30] while *BCL3* links LIF-STAT3 to core pluripotency genes *Oct4* and *Nanog* to promote the maintenance of naive pluripotency.[31] GO analysis revealed conserved pathways in nESCs, such as cellular developmental process, regulation of cellular and biological processes, cell differentiation, and intracellular signaling cassette (Figures S2B and S2C). In EPSCs, 18 commonly upregulated genes were identified, including well-known pluripotency

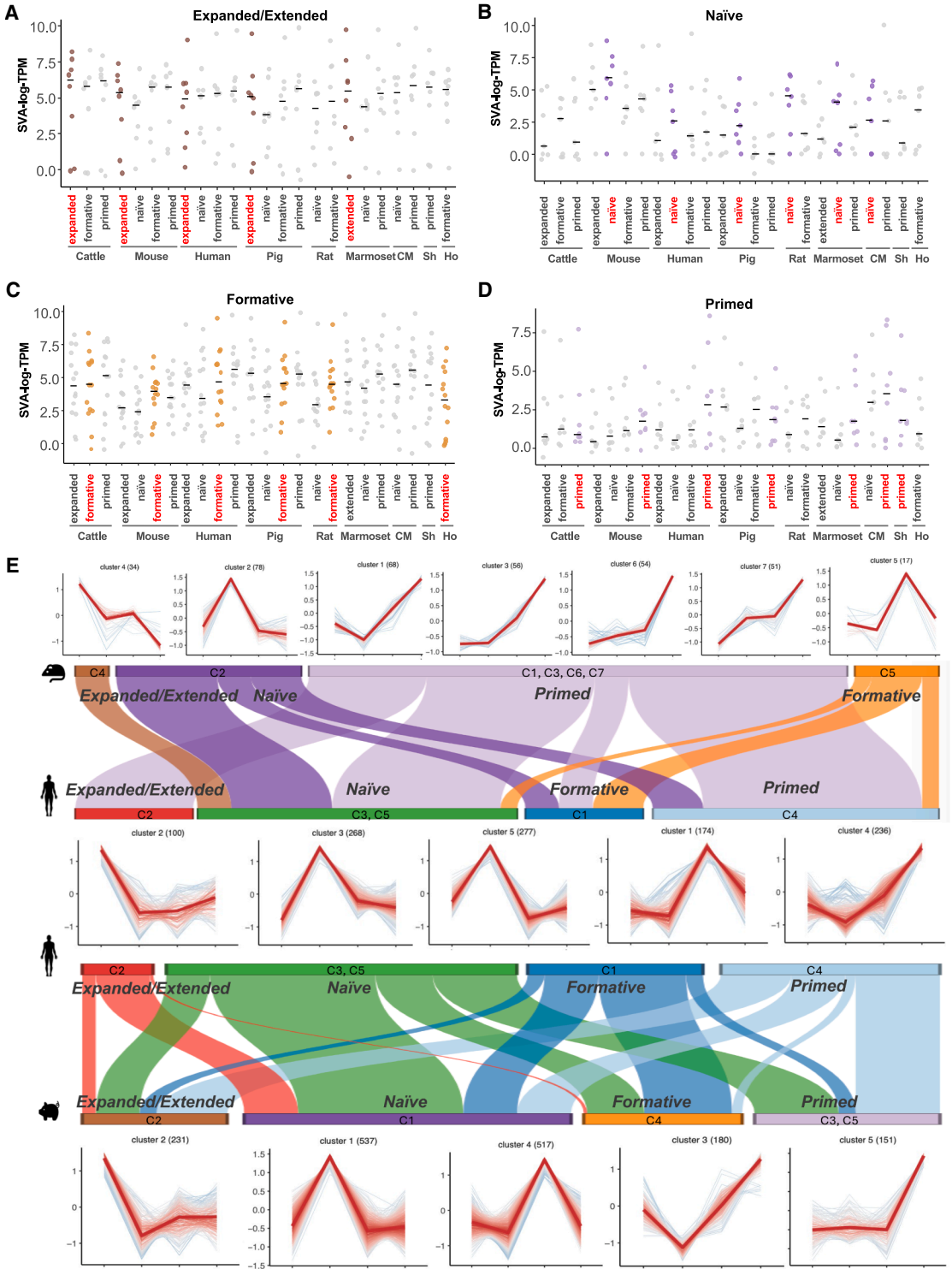


Figure 3. Assessment of known state-specific markers across species and identification of *de novo* markers

(A–D) Evaluation of known mouse state-specific markers for the expanded/extended state (A), naive state (B), formative state (C), and primed state (D). CM represents crab-eating macaque, Sh represents sheep, and Ho represents horse. The SVA-log-TPM values were used.

(legend continued on next page)

regulators *MYC* and *DNMT3B* (Figures S2D and S5B). *MYC* facilitates cell reprogramming and pluripotent state establishment,[32] while *DNMT3B* plays a critical role in enabling ESCs to differentiate into multiple lineages, and its depletion has been associated with sustained self-renewal and impaired differentiation capacity.[33] Unexpectedly, many of the DEGs identified as specific to the expanded/extended state also exhibited high expression levels in the formative and primed state cells across multiple species (Figures S5B and S5E), suggesting shared pluripotency gene regulatory mechanism between these states. Shared GO pathways in EPSCs included cell surface receptor signaling, regulation of cell adhesion, and embryonic morphogenesis, revealing core regulatory processes of expanded/extended pluripotency state maintenance (Figures S2E and S2F, Table S4).

FSCs vs. nESCs comparison

In the comparison between FSCs and nESCs across humans, mice, rats, and pigs (Figure S3), 63 commonly upregulated DEGs in formative states were identified. These DEGs were also enriched in primed states across species (Figure S5C), suggesting shared pluripotency regulation between these two states. Several key genes involved in transitioning from naive to primed pluripotency, ESC proliferation, and maintaining stem cell states, including *DUSP6*,[18] *USP44*,[34] and *DNMT3B*[35] (Figures S3A and S5C). Enriched GO pathways in FSCs among the four species included cell-cell signaling by WNT, cell differentiation, and MAPK cascade (Figures S3B and S3C). Conversely, 16 common upregulated genes were identified in the naive state, such as *MVP, WHAMM, WDR62,* and *ULK1* (Figures S3D and S5C). *ULK1* is essential for mouse ESC self-renewal and pluripotency.[36] Notably, most of these 16 naïve-specific DEGs also exhibited elevated expression in naive state cells of both macaque and marmoset, indicating conserved regulation mechanisms (Figure S5C). Shared GO pathways in nESCs included regulation of cell communication, signal transduction, cell cycle, and apoptotic process (Figures S3E and S3F, Table S4).

EpiSCs vs. FSCs comparison

When comparing EpiSCs with FSCs across humans, mice, pigs, and cattle (Figure S4), five commonly shared upregulated genes in the primed state were *MYL9, PNKD, ATP1B2*, *TSPAN7*, and *NRARP* (Figures S4A and S5D). *TSPAN7* plays an important role in neural differentiation and axonal development,[37] while *NRARP,* a part of the Notch signaling pathway, plays a key role in differentiation and pluripotency maintenance.[38,39] These genes also exhibited high expression in the primed states of macaque and marmoset, the formative state in rats, and were elevated in the expanded state but downregulated in the naive state across most species (Figure S5D). Shared GO terms in all species except mice included regulation of cell differentiation, cell surface receptor, and transmembrane receptor protein tyrosine kinase signaling pathways (Figures S4B and S4C, Table S4). In FSCs, two shared upregulated genes were *ZBTB11* and *GDF3* (Figures S4D and S5D). *ZBTB11* maintains pluripotency by repressing pro-differentiation genes in ESCs.[22] *ZBTB11* showed high expression in both naive and formative states across all species, except in mice, where its expression was elevated in the expanded state (Figure S5D). *GDF3*, a TGF-β superfamily member, regulates ESCs pluripotency through BMP inhibition in a species-specific manner.[40] We could not assess *GDF3* expression across pluripotency states among all species due to the absence of 1:1 conserved orthologs (Figure S5D). Enriched pathways in FSCs across four species involved regulation of gene expression and DNA-templated transcription (Figures S4E and S4F).

Evaluation of known and novel pluripotency state-specific markers

Next, we evaluated the performance of known pluripotency state-specific markers (Table S5) defined in mice to assess their expression levels and applicability across mammals (Figure 3). Our analysis revealed that, while these state-specific markers effectively differentiated pluripotency states in mouse ESCs, they failed to consistently distinguish their corresponding pluripotency states across all species (Figures 3A–3D). For example, expanded/extended markers (n = 8) did not show the highest expression of EPSCs compared with other ESC states in humans (Figure 3A). Similarly, naive markers (n = 7) generally performed well across species except for differentiating nESC from EpiSCs in macaque (Figure 3B). Moreover, formative (n = 14) and primed (n = 9) markers were able to differentiate FSCs and EpiSCs from EPSCs or nESCs but could not distinguish between each other in most species that contained these two states (Figures 3C and 3D).

Given the limitations of mouse-based ESC markers, we aim to identify *de novo* state-specific markers applicable across multiple mammalian species. Using human, mouse, and pig ESC datasets, each of which contains all four pluripotency states. We first conducted pairwise differential expression results to identify genes consistently upregulated in a single pluripotency state within each species. To further examine the expression pattern of these candidate markers, we applied soft clustering (*see STAR Methods*) to capture shared expression trajectories across states within each species (Figures 3E, S6A and S6B). Genes exhibiting similar expression dynamics were grouped into the same cluster, each representing a distinct expression pattern in which candidate markers exhibit peak expression in their associated pluripotency state. The Sankey plot (Figure 3E) showed how these state-specific markers align across species. For example, mouse primed state markers (clusters C1, C3, C6, C7) overlap with primed markers in human (C4) and pig (C3, C5). Additionally, about half of the human naïve-associated markers (C3, C5) derived from mouse primed state clusters align with pig naive

(E) *De novo* state-specific markers identified across species. Markers identified by differential expression analysis were grouped into clusters using the Mfuzz soft clustering algorithm. Each panel represents one cluster of genes (number of genes shown in parentheses), where the x axis corresponds to the four pluripotency states and the y axis indicates expression changes. Thin lines represent the expression profiles of individual genes, while the bold red line depicts the trend of the cluster, summarizing the dominant expression trajectory within each cluster. The top Sankey diagram compares shared marker gene flow between mouse and human, while the bottom one compares human and pig.

markers (C1). Overall, these analyses revealed both conserved and species-specific regulatory signatures of pluripotency.

Common genes assigned in the same state-specific clusters across species were further defined as shared state-specific markers. For example, *CHRNA4,* a potential regulator of pluripotency maintenance,[41] was identified as the only primed state-specific marker shared across all three species (Figure S6A). Other state-specific markers were significantly upregulated in their respective states in at least two species, such as *DBNDD1* in expanded/extended state, *BHMT*[42] in naive state, and *SNORD17*[43] in formative state. Some of these genes play known roles in pluripotency-related functions and show consistently elevated expression in their corresponding states across species. In contrast, pluripotency-specific genes that did not overlap across species were identified as species- and state-specific genes. Interestingly, genes such as *LEF1, SMC6,* and *LSR* were uniquely elevated in the expanded/extended state in humans, mice, or pigs, respectively, but not in the other two species, suggesting that each species possesses a unique set of markers tailored to their pluripotency regulation (Figure S6B and Table S6).

Gene co-expression network construction and preservation analysis

To systematically define pluripotency state-specific gene co-expression networks and evaluate their evolutionary conservation, we performed WGCNA across the four pluripotency states in all species except sheep, horses, and macaque due to insufficient sample sizes (Figures 4 and S7–S10). This analysis identified several modules significantly correlated with each specific pluripotency state. For example, in the human pluripotency network, the modules labeled salmon (R = 0.87), turquoise (R = 0.96), tan (R = 0.95), black (R = 0.94), and blue (R = 0.85) were significantly associated with the expanded/extended, naive, formative, and primed states, respectively (Figure 4A). Similarly, in mice, the modules labeled light green (R = 0.85) and blue (R = 0.8) were significantly correlated with the primed state in mice (Figure 4B), while in pigs, modules labeled green (R = 0.89) and pink (R = 0.91) were enriched in the primed state (Figure S9).

To assess the conservation of pluripotency state-specific gene modules across species, we conducted a preservation analysis using the human network as a reference (Figures 4C and 4D, Figure S10). This approach evaluates whether the structure and connectivity of gene co-expression modules identified in one species are retained in another, providing a quantitative measure of cross-species network conservation. This analysis identified several conserved modules between humans and other species, where a module with a Z summary score >2 indicated weak preservation and a Z summary score >10 reflects strong preservation. Notably, the blue, black, yellow, brown, purple, and dark red modules from the human network were preserved across all species analyzed (Figure 4C).

Given that the blue and black modules had a strong correlation with the human primed state, we further investigated whether the preserved modules in other species also correlated with this state. We found that the human blue module was largely preserved within the blue, light green, and pink modules in mice (Figure 4D) and turquoise modules of marmosets and cattle. All of these modules showed significant correlations to the primed state in their respective species (Figures S9B and S9C). However, in pigs, the human blue module was preserved in the yellow and salmon modules (Figure 4C), which were significantly enriched in the extended/expanded state (Figure S9A). Similarly, the human black module, also associated with the primed state, was substantially preserved in modules such as pink in mice, brown and green in pigs, and turquoise and yellow in cattle (Figures 4D and S8). These preserved modules exhibited significant primed state associations in their respective species, except for the magenta and greenyellow modules in marmoset and black modules in mice, which lacked strong primed correlations (Figures 4D and S8C). These findings indicate that the human blue and black modules represented conserved gene co-expression networks associated with the primed state across most mammalian species examined, with some species-specific exceptions.

To highlight the internal structure of conserved co-expression modules, we visualized the human gene networks using module-specific subgraphs constructed from the topological overlap matrix, which quantifies the strength of connection between gene pairs based on their shared network neighbors and co-expression similarity (Figures 4E and S11A). We identified several key hub genes: *NAB1*, a transcriptional repressor,[44] along with *SLC27A2*, *MGAT4C*, and *AMPH*, involved in lipid metabolism,[45] glycosylation,[46] and endocytosis,[47] respectively, were identified as key hub genes in the blue module. Similarly, *ELAPOR2*, associated with the regulation of autophagy and apoptosis[48]; *LTA4H*, a mediator of leukotriene-associated inflammation[49]; and *TSPAN7*, contributing to cell adhesion, migration, and neural development[50]; were identified as key hub genes in the black module. GO analysis further revealed pathways related to neuronal development, tissue morphogenesis, and system differentiation (Figures S11B and S11C), indicating that primed ESCs are poised for lineage commitment, consistent with their role in later embryonic development stages.

Transcriptome-based phylogenies

To gain insights into the evolution of gene expression, we reconstructed transcriptome-based phylogenetic trees for four distinct pluripotency states (Figure 5A). These trees represent evolutionary relationships among species inferred from global gene expression patterns, offering a complementary perspective to sequence-based phylogenies by capturing functional divergence at the transcriptomic level. Each state exhibited a unique tree structure, reflecting dynamic transcriptional changes and interspecies variations across different pluripotency states. Rodents and primates consistently clustered closely across all states, whereas ungulates displayed a more varied clustering pattern. For example, porcine FSCs formed a distinct branch rather than clustering with cattle and horses in the formative-specific phylogenetic tree (Figure 5A). Given the limited availability of FSC datasets in these species, it remains unclear whether this divergence reflects a genuine biological difference or study-specific technical variation.

Among all states, the primed state showed the closest alignment with known mammalian phylogeny based on DNA or

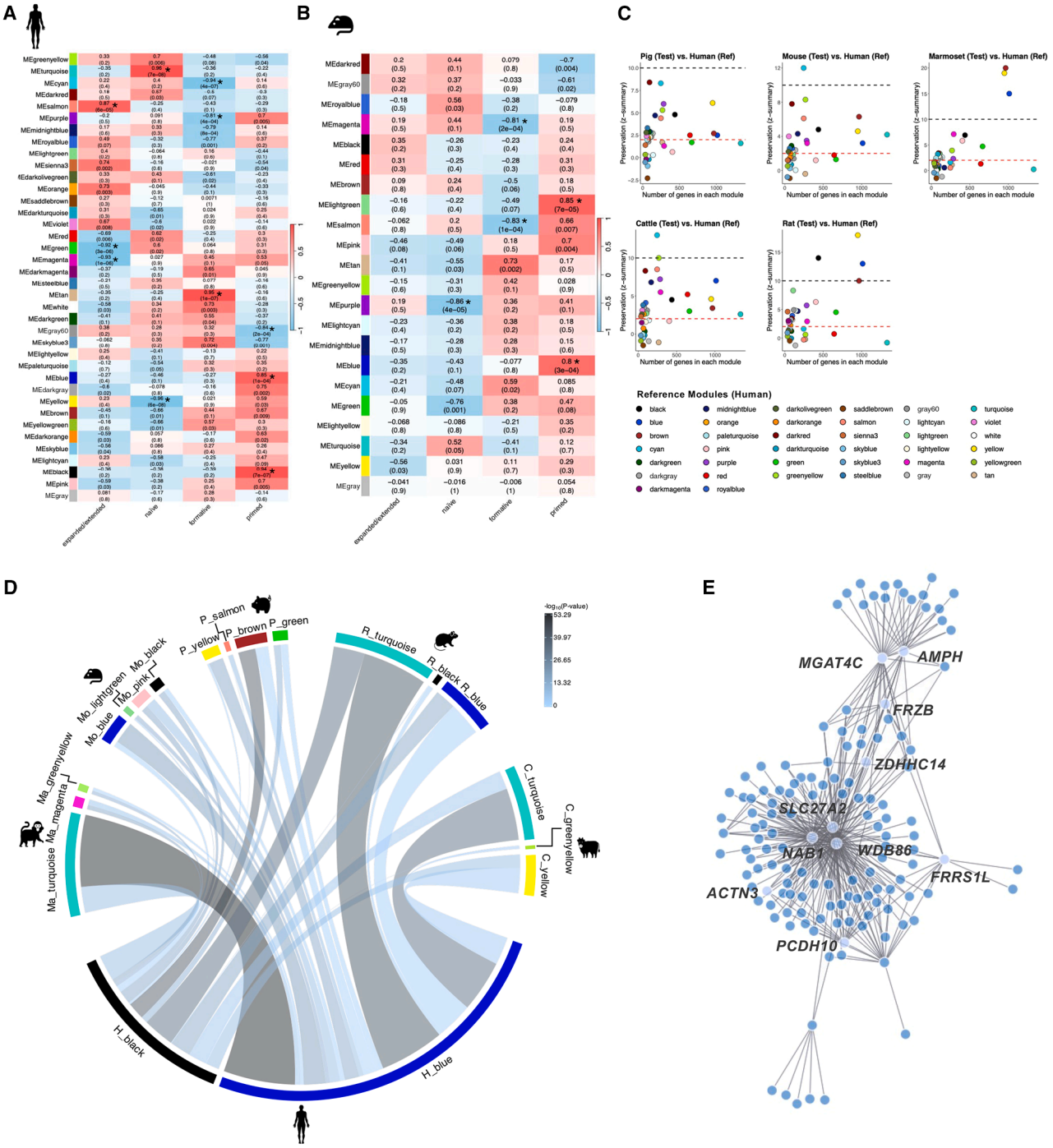


Figure 4. Cross-species comparison of WGCNA analysis

(A) Module-trait relationships within the human co-expression network, with modules showing a correlation coefficient ≥ 0.8 and $p < 0.05$ marked by an asterisk.

(B) Module-trait relationships within the mouse co-expression network, highlighting modules with a correlation coefficient >0.8 and $p < 0.05$ by an asterisk.

(C) Z-statistics depicting module preservation of the human network (reference network) in networks of other species. Thresholds for weak preservation (Z = 2) and strong preservation (Z = 10) are indicated in the plots.

(D) The top two modules from other species' networks where the human blue and black modules are mainly preserved. The thickness of the connecting bands represents the number of overlapping genes between the two modules, while the color indicates the $-\log_{10}(P)$ value, where P represents the statistical significance of the overlap.

(E) Co-expression network of the human blue module, with the top 10 hub genes (ranked by centrality scores using Kleinberg metric) labeled.

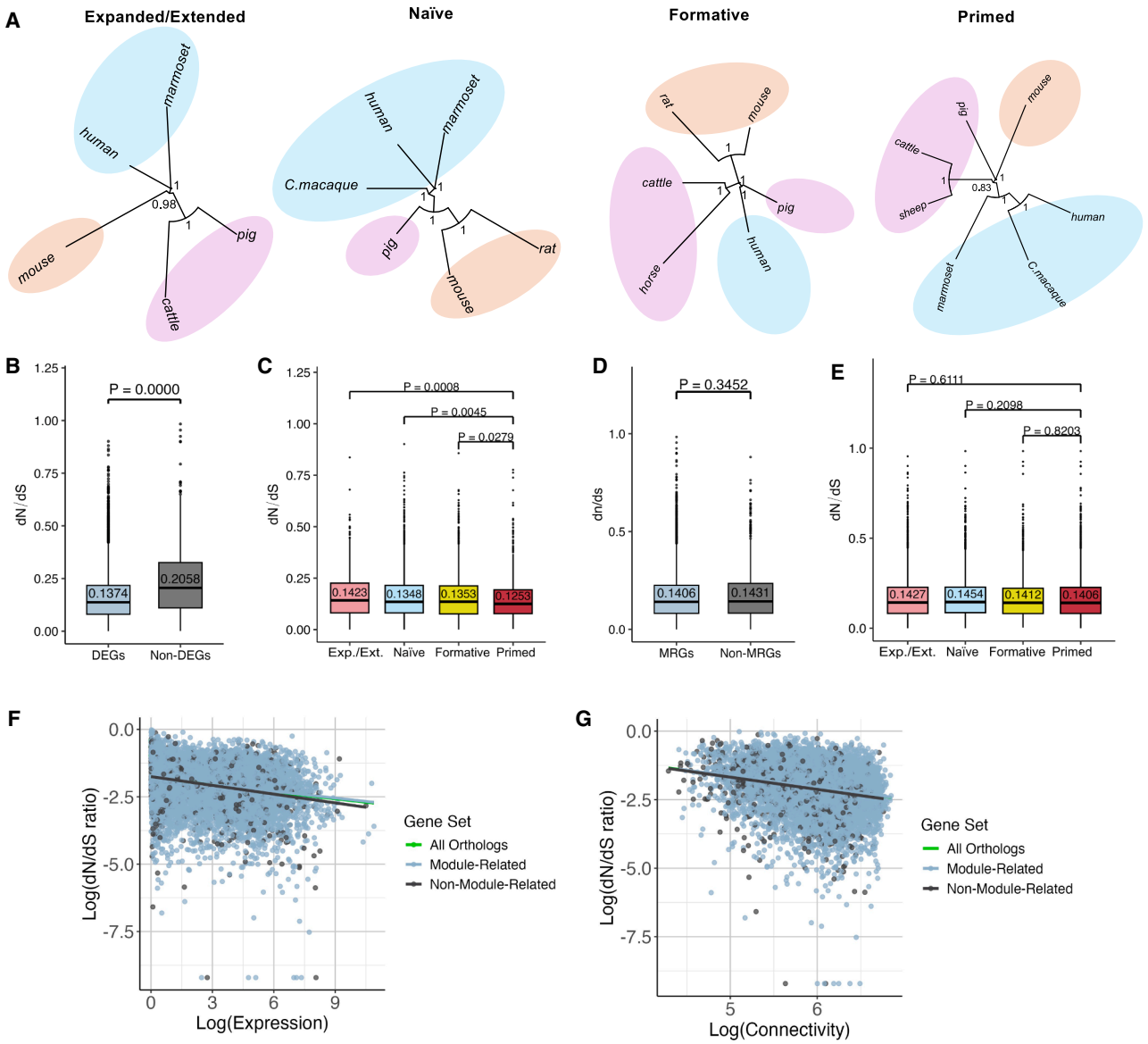


Figure 5. Analysis of expression evolution

(A) Transcriptome-based phylogenetic trees for the four pluripotency states. Neighbor-joining trees were constructed using 1:1 orthologous genes, with branch lengths representing the proportion of expression variation corresponding to evolutionary expression changes. Bootstrap values (calculated from 1,000 random resamplings of 1:1 orthologous genes) are indicated.

(B–E) Boxplots illustrating dN/dS ratios for various gene sets: DEGs versus non-DEGs (B), state-specific genes identified through DEG analysis (C), module-associated genes (MRGs) identified by WGCNA as related to any pluripotency state versus Non-MRGs (D), and genes in WGCNA modules associated with specific states (E). *p*-values above each comparison indicate the significance of dN/dS differences between gene sets (Mann-Whitney U test).

(F–G) dN/dS ratios as functions of (F) gene expression and (G) connectivity identified by WGCNA. Smoothing functions represent different gene sets, with the green line for all orthologous genes, the blue line for module-associated genes (MRGs), and the black line for non-module-associated genes (Non-MRGs). Negative correlations are observed between gene expression and dN/dS ratio, as well as between connectivity and dN/dS ratio.

protein sequences, suggesting it represents a conserved post-implantation development stage, whereas other states exhibit more developmental diversity or species-specific variations (Figure 5A). This is further supported by the consistent use of FGF2 and Activin-A (F/A) in the culture media for EpiSCs across species, in contrast to more variable supplements used for naive conditions between mice and humans.[13]

Evolutionary rates are correlated with pluripotency

To investigate whether evolutionary rates of protein genes are associated with pluripotency states, we calculated the dN/dS ratio, which indicates the rate of nonsynonymous (amino acid-changing) substitutions relative to synonymous (neutral) substitutions in a gene.[51] This metric serves as a proxy for selective pressure, where lower values suggest functional constraint and

higher values may indicate relaxed or positive selection. DEGs identified from pairwise pluripotency state comparisons exhibited a significantly lower dN/dS ratio (0.14) compared with non-DEGs (0.21, $p < 0.0001$, Mann-Whitney U test, Figures 5B and S12A). While both median dN/dS values were below 1, indicating purifying selection, DEGs had a significantly lower dN/dS ratio, suggesting stronger evolutionary constraints. However, since DEGs also showed higher expression levels than non-DEGs, we cannot fully exclude gene expression as a confounding factor in the observed evolutionary rate differences.

To ensure these results were not biased by a large number of total DEGs, we further divided the DEG and non-DEG by species (Figure S12E) or separated DEGs in specific pairwise comparisons (expanded extended/naive, naive/formative, formative/primed) within each species (Figure S12F). Across comparisons, DEGs consistently showed significantly lower dN/dS values than non-DEGs, except for the mouse formative/primed comparison (Figure S12F). These findings reinforce that genes involving pluripotency state transitions are generally under stronger purifying selection with a slower evolutionary rate.[52]

We further analyzed dN/dS ratios for pluripotency-specific genes uniquely expressed in different pluripotency states among all species. Our results revealed that the primed state had the significantly ($p < 0.05$, Mann-Whitney U test) lowest dN/dS ratios compared with other states (Figures 5C and S12B), suggesting higher conservation of the primed states related genes with slower evolutionary rates. Similarly, genes from WGCNA modules (MRGs) correlated with specific states exhibited lower dN/dS ratios than those from non-correlated modules (Non-MRGs, Figures 5D and S12C). Among these modules, the primed state showed the lowest dN/dS values (Figures 5E and S12D). Finally, we observed a statistically significant negative correlation between gene expression levels and dN/dS ratios (linear model coefficient = −0.0864, $P < 1e\text{-}10$, linear model, Figure 5F), as well as between WGCNA module connectivity and dN/dS ratios (linear model coefficient = −0.4158, $P < 1e\text{-}10$, linear model, Figure 5G). To further investigate the relationships among these variables, we performed a correlation and partial correlation analysis to assess how dN/dS, gene expression levels, and connectivity interact (Table S7). The results confirmed that genes with higher expression levels or greater connectivity in gene co-expression networks tend to evolve more slowly, consistent with the hypothesis that highly expressed and functionally central genes are subject to stronger purifying selection. These findings indicate that genes that were highly expressed or had more gene network connections evolve more slowly. Overall, these results highlight the close relationship between evolutionary rates and pluripotency across states, with genes associated with pluripotency transitions, particularly in the primed state, showing stronger evolutionary constraints and slower rates of evolution.

DISCUSSION

Defining and aligning pluripotency states is essential for meaningful cross-species comparisons. In this study, we used the state assignments reported by the original publications, which were generally supported by multiple lines of evidence, including colony morphology, transcriptomic profiling, *in vivo* stage mapping, and functional validation such as embryoid body formation, teratoma formation, or directed differentiation. However, perfect alignment across species remains challenging due to differences in developmental timelines and experimental conditions, particularly for less-characterized states like the formative and expanded or extended states. Given the limited number of currently available datasets, we cannot fully exclude the influence of technical noise when interpreting biological variation from these states. Nevertheless, our analysis revealed consistent preservation of state-associated transcriptomic trajectories and co-expression patterns, suggesting core molecular features defining each pluripotency state are broadly conserved and comparable across species. Future studies with standardized experimental protocols and expanding datasets, especially for under-characterized states, will help refine these cross-species comparisons.

Dynamics of gene expression in pluripotency states transition

Cross-species DEG analysis revealed conserved genes such as *ZBTB10* and *IHO1* in the naive state and *OTX2* in the primed state. *ZBTB10*, a C2H2 zinc finger BTB domain TF expressed in human and mouse early embryos, is downregulated upon differentiation.[53] Although *ZBTB10* knockdown has minimal effects on ESC morphology or OCT4-positive cell numbers, it may have accessory or self-renewal functions.[22] Conversely, the functions of *IHO1* in pluripotency remain underexplored, underscoring the need for further investigation. *OTX2,* a key transcription factor facilitating the transition from naive to primed state, has been widely recognized as essential for maintaining primed pluripotency.[25,54,55]

To capture intermediate dynamics of the pluripotency transition, we incorporated the formative state between naive and primed states. This analysis identified conserved upregulated genes including *DNMT3B* and *DUSP6* during the naïve-to-formative transition, while *ZBTB11* was downregulated during the formative-to-primed transition. These findings align with previous studies. For example, *DNMT3B*[56] is a well-documented formative marker in mice, while *DUSP6* is known to associate with the naïve-to-primed transition[17] and up-regulated in the intermediate state of human PSCs.[57] Further GO analysis highlighted consistently conserved pathways across several states, such as the MAPK cascade, which has been shown to play a pivotal role in balancing proliferation and differentiation in stem cells.[58] Additionally, genes upregulated in the formative state were enriched in pathways related to DNA recombination and chromosome organization. These pathways may be linked to X chromosome inactivation, a hallmark event during the transition between naive and primed states, which involves extensive chromatin reorganization.[13,59]

Identification of pluripotency state-specific markers and co-expression hub genes

Our analysis revealed that mouse pluripotency state-specific markers were insufficient to distinguish pluripotency states in other species. We further identified conserved and species-specific *de novo* markers for each pluripotent state. We observed that the primed state had the highest number of shared markers,

with *CHRNA4* conserved across humans, mice, and pigs. *CHRNA4* encodes the neuronal acetylcholine receptor subunit alpha-4 (nAChRα4) and is involved in neural differentiation.[60] Our findings revealed substantial variability in species-specific gene expression patterns and uncovered an intriguing phenomenon that a state-specific cluster in one species aligned with a different state cluster in another species. This indicates the need for caution when selecting or applying markers from one species to another in cross-species analyses.

While state-specific genes are poorly conserved across species, the structure of WGCNA co-expression networks appears to be well preserved among mammals, consistent with previous comparative transcriptomic analyses involving data from eight tissues, such as skin, mammary gland, and marrow.[61] This conservation may reflect the integrated nature of pluripotency regulation, which relies on interactions between various signaling pathways rather than isolated gene activity.[62] In our analysis, several human co-expression primed-specific modules were consistently preserved in cattle, pigs, mice, rats, and marmosets. Notably, the human blue module and its preserved counterparts in other species showed a significant correlation with the primed state across species. GO analysis revealed enrichment for pathways associated with neuronal development and differentiation, supporting the notion that cells in the primed state are poised for lineage commitment while retaining elements of the pluripotency network.[63]

Primed state exhibits the highest evolutionary conservation

We further calculated the dN/dS ratios for DEGs, non-DEGs, and state-specific upregulated genes. Our results revealed that DEGs generally have lower dN/dS values than non-DEGs, suggesting that genes involved in pluripotency regulation are subject to stronger purifying selection. Such evolutionary constraints align with our observation that the core pluripotency regulatory network demonstrates significant conservation, indicating the critical role of pluripotency-related genes in maintaining essential cellular processes.[64] Interestingly, primed state-specific genes exhibited significantly lower dN/dS values compared with genes specific to other states, indicating intensified selective pressure on this state. This aligns with the primed state showing the highest evolutionary conservation, as demonstrated by transcriptome-based phylogenetic clustering, a greater number of shared DEGs, and a higher WGCNA module preservation across species. This conservation may be due to the consistent derivation of EpiSCs from the embryonic epiblast cells of the post-implantation embryonic stage across species, in contrast to nESCs, which are derived from the more variable pre-implantation stage. Moreover, the uniformity of culture conditions for EpiSCs across species may contribute to the transcriptomic similarities, in contrast to the species-specific conditions required for naive states.[13] Additionally, a previous study[65] comparing several pre- and post-implantation stage embryos of monkeys, pigs, and humans revealed that post-implantation samples clustered closer across species than pre-implantation samples in PCA. This finding suggests that the observed conservation of the primed state primarily reflects the intrinsic evolutionary stability of post-implantation epiblast cells during early development, supporting the idea of the hourglass model, which proposes that embryonic divergence is highest at the earliest and latest stages, while the mid-embryonic (phylotypic) stage exhibits the greatest evolutionary conservation.[66]

Our study provides a comparative transcriptomic analysis across mammalian species, revealing both conserved and species-specific mechanisms of pluripotency regulation. The primed state emerged as the most evolutionarily conserved state, with shared pathways and gene networks highlighting its critical role in preparing cells for lineage commitment. While species-specific gene expression patterns highlight evolutionary diversity, the preservation of gene co-expression network structures suggests a broader regulatory network conserved across species. The knowledge gained here provides a foundation for standardizing pluripotency state characterization within species and optimizing ESC culture systems across diverse species. These advancements hold the potential to enhance applications in regenerative medicine, sustainable agriculture, and conservation while deepening our understanding of the evolutionary mechanisms shaping early mammalian development.

Limitations of the study

This study presents a comprehensive cross-species transcriptomic comparison of pluripotency states in ESCs. However, several limitations should be acknowledged. For example, all datasets were sourced from previously published studies, which introduces variability in sample collection, culture conditions, maintenance protocols, and sequencing methods. While we applied a standardized processing pipeline with surrogate variable analysis to reduce technical noise, residual batch effects cannot be fully ruled out. Moreover, the assignment of pluripotency states across species was based on original publications using criteria such as morphology and molecular markers, which may not be fully aligned across species. Our comparative approach also focused on 1:1 orthologous genes, enabled cross-species analysis, but excluded non-orthologous or species-specific genes that may be functionally relevant. Furthermore, formative and expanded/extended states remain underrepresented in public datasets, limiting the resolution of our comparative analyses. While our work provides valuable transcriptomic insights, the absence of matching epigenomic and proteomic data constrains the interpretation of regulatory mechanisms. Additionally, these findings were based on nine species and may not fully capture the diversity of pluripotency regulation across all mammals. Broader species inclusion and integration of multi-omics data will be crucial in future studies to validate and expand upon these findings.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Jingyue (Ellie) Duan (jd774@cornell.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: All datasets analyzed in this work are publicly available and listed in Table S1.
- Code: The source code used for data processing and analysis is available at: https://github.com/coderFaye/CS-ESC.
- Other: No additional resources were generated or analyzed in this study.

ACKNOWLEDGMENTS

The author thanks all members of the IISAGE Consortium for the discussion of the results. This study was supported by grants from the NSF BII (#2213824) to J.E.D. and the IISAGE Consortium.

AUTHOR CONTRIBUTIONS

Y.F., J.E.D., X.T., and Y.T. conceptualized the research. Y.F. conducted the sample curation and data analysis. Y.S. supported dataset curation. R.M., J.R.W., T.G., E.C.R., X.T., Y.T., and IISAGE Consortium contributed to the analysis methods and interpretation of the data. X.Y., M.S., G.L., and J.Z. helped with manuscript preparation and data discussion. Writing-original draft, Y.F. and J.E.D. All co-authors participated in revising the manuscript. Funding acquisition J.E.D. and the IISAGE Consortium.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Curation and quantification of transcriptome data
  - Ortholog retrieval and data normalization
  - Data exploratory analysis
  - Differential gene expression analysis
  - Identification of *de novo* state-specific markers
  - Gene co-expression network construction and preservation analysis
  - Gene expression phylogenies
  - dN/dS ratio calculation
  - Gene ontology enrichment analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.114542.

Received: March 6, 2025
Revised: August 15, 2025
Accepted: December 19, 2025
Published: December 24, 2025

REFERENCES

1. Liang, G., and Zhang, Y. (2013). Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. Cell Res. *23*, 49–69. https://doi.org/10.1038/cr.2012.175.
2. Varzideh, F., Gambardella, J., Kansakar, U., Jankauskas, S.S., and Santulli, G. (2023). Molecular Mechanisms Underlying Pluripotency and Self-Renewal of Embryonic Stem Cells. Int. J. Mol. Sci. *24*, 8386. https://doi.org/10.3390/ijms24098386.
3. Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 516–528. https://doi.org/10.1038/ncb2965.
4. Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A.F., Jones, V., Hor, A., de Alencastro, G., Logan, G.J., et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell *14*, 107–120. https://doi.org/10.1016/j.stem.2013.09.014.
5. Yu, L., Wei, Y., Sun, H.-X., Mahdi, A.K., Pinzon Arteaga, C.A., Sakurai, M., Schmitz, D.A., Zheng, C., Ballard, E.D., Li, J., et al. (2021). Derivation of Intermediate Pluripotent Stem Cells Amenable to Primordial Germ Cell Specification. Cell Stem Cell *28*, 550–567.e12. https://doi.org/10.1016/j.stem.2020.11.003.
6. Shen, H., Yang, M., Li, S., Zhang, J., Peng, B., Wang, C., Chang, Z., Ong, J., and Du, P. (2021). Mouse totipotent stem cells captured and maintained through spliceosomal repression. Cell *184*, 2843–2859.e20. https://doi.org/10.1016/j.cell.2021.04.020.
7. Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282–286. https://doi.org/10.1038/nature12745.
8. Kurek, D., Neagu, A., Tastemel, M., Tüysüz, N., Lehmann, J., van de Werken, H.J.G., Philipsen, S., van der Linden, R., Maas, A., van IJcken, W.F.J., et al. (2015). Endogenous WNT Signals Mediate BMP-Induced and Spontaneous Differentiation of Epiblast Stem Cells and Human Embryonic Stem Cells. Stem Cell Rep. *4*, 114–128. https://doi.org/10.1016/j.stemcr.2014.11.007.
9. Iwatsuki, K., Oikawa, M., Kobayashi, H., Penfold, C.A., Sanbo, M., Yamamoto, T., Hochi, S., Kurimoto, K., Hirabayashi, M., and Kobayashi, T. (2023). Rat post-implantation epiblast-derived pluripotent stem cells produce functional germ cells. Cell Rep. Methods *3*, 100542. https://doi.org/10.1016/j.crmeth.2023.100542.
10. Bogliotti, Y.S., Wu, J., Vilarino, M., Okamura, D., Soto, D.A., Zhong, C., Sakurai, M., Sampaio, R.V., Suzuki, K., Izpisua Belmonte, J.C., and Ross, P.J. (2018). Efficient derivation of stable primed pluripotent embryonic stem cells from bovine blastocysts. Proc. Natl. Acad. Sci. USA *115*, 2090–2095. https://doi.org/10.1073/pnas.1716161115.
11. Kinoshita, M., Kobayashi, T., Planells, B., Klisch, D., Spindlow, D., Masaki, H., Bornelöv, S., Stirparo, G.G., Matsunari, H., Uchikura, A., et al. (2021). Pluripotent stem cells related to embryonic disc exhibit common self-renewal requirements in diverse livestock species. Development *148*, dev199901. https://doi.org/10.1242/dev.199901.
12. Sakai, Y., Nakamura, T., Okamoto, I., Gyobu-Motani, S., Ohta, H., Yabuta, Y., Tsukiyama, T., Iwatani, C., Tsuchiya, H., Ema, M., et al. (2020). Induction of the germ cell fate from pluripotent stem cells in cynomolgus monkeys. Biol. Reprod. *102*, 620–638. https://doi.org/10.1093/biolre/ioz205.
13. Du, P., and Wu, J. (2024). Hallmarks of totipotent and pluripotent stem cell states. Cell Stem Cell *31*, 312–333. https://doi.org/10.1016/j.stem.2024.01.009.
14. Osnato, A., Brown, S., Krueger, C., Andrews, S., Collier, A.J., Nakanoh, S., Quiroga Londoño, M., Wesley, B.T., Muraro, D., Brumm, A.S., et al. (2021). TGFβ signalling is required to maintain pluripotency of human naïve pluripotent stem cells. eLife *10*, e67259. https://doi.org/10.7554/eLife.67259.
15. McGowen, M.R., Erez, O., Romero, R., and Wildman, D.E. (2014). The evolution of embryo implantation. Int. J. Dev. Biol. *58*, 155–161. https://doi.org/10.1387/ijdb.140020dw.
16. Uddin, M., Goodman, M., Erez, O., Romero, R., Liu, G., Islam, M., Opazo, J.C., Sherwood, C.C., Grossman, L.I., and Wildman, D.E. (2008). Distinct genomic signatures of adaptation in pre- and postnatal environments

during human evolution. Proc. Natl. Acad. Sci. USA *105*, 3215–3220. https://doi.org/10.1073/pnas.0712400105.

17. Bernardo, A.S., Jouneau, A., Marks, H., Kensche, P., Kobolak, J., Freude, K., Hall, V., Feher, A., Polgar, Z., Sartori, C., et al. (2018). Mammalian embryo comparison identifies novel pluripotency genes associated with the naïve or primed state. Biol. Open *7*, bio033282. https://doi.org/10.1242/bio.033282.
18. Yoo, D.H., Im, Y.S., Oh, J.Y., Gil, D., and Kim, Y.O. (2023). DUSP6 is a memory retention feedback regulator of ERK signaling for cellular resilience of human pluripotent stem cells in response to dissociation. Sci. Rep. *13*, 5683. https://doi.org/10.1038/s41598-023-32567-8.
19. Jung, J.E., Moon, S.H., Kim, D.K., Choi, C., Song, J., and Park, K.S. (2012). Sprouty1 regulates neural and endothelial differentiation of mouse embryonic stem cells. Stem Cells Dev. *21*, 554–561. https://doi.org/10.1089/scd.2010.0568.
20. Choi, K.-H., Lee, D.-K., Kim, S.W., Woo, S.-H., Kim, D.-Y., and Lee, C.-K. (2019). Chemically Defined Media Can Maintain Pig Pluripotency Network In Vitro. Stem Cell Rep. *13*, 221–234. https://doi.org/10.1016/j.stemcr.2019.05.028.
21. Zhi, M., Zhang, J., Tang, Q., Yu, D., Gao, S., Gao, D., Liu, P., Guo, J., Hai, T., Gao, J., et al. (2022). Generation and characterization of stable pig pregastrulation epiblast stem cell lines. Cell Res. *32*, 383–400. https://doi.org/10.1038/s41422-021-00592-9.
22. Garipler, G., Lu, C., Morrissey, A., Lopez-Zepeda, L.S., Pei, Y., Vidal, S.E., Zen Petisco Fiore, A.P., Aydin, B., Stadtfeld, M., Ohler, U., et al. (2022). The BTB transcription factors ZBTB11 and ZFP131 maintain pluripotency by repressing pro-differentiation genes. Cell Rep. *38*, 110524. https://doi.org/10.1016/j.celrep.2022.110524.
23. Stanzione, M., Baumann, M., Papanikos, F., Dereli, I., Lange, J., Ramlal, A., Tränkner, D., Shibuya, H., de Massy, B., Watanabe, Y., et al. (2016). Meiotic DNA break formation requires the unsynapsed chromosome axis-binding protein IHO1 (CCDC36) in mice. Nat. Cell Biol. *18*, 1208–1220. https://doi.org/10.1038/ncb3417.
24. Nagaria, P., Robert, C., and Rassool, F.V. (2013). DNA double-strand break response in stem cells: Mechanisms to maintain genomic integrity. Biochim. Biophys. Acta *1830*, 2345–2353. https://doi.org/10.1016/j.bbagen.2012.09.001.
25. Acampora, D., Di Giovannantonio, L.G., and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. Development *140*, 43–55. https://doi.org/10.1242/dev.085290.
26. Son, M.-Y., Seol, B., Han, Y.-M., and Cho, Y.S. (2014). Comparative receptor tyrosine kinase profiling identifies a novel role for AXL in human stem cell pluripotency. Hum. Mol. Genet. *23*, 1802–1816. https://doi.org/10.1093/hmg/ddt571.
27. Ohtsuki, S., Kikkawa, T., Hori, S., and Terasaki, T. (2006). Modulation and Compensation of the mRNA Expression of Energy Related Transporters in the Brain of Glucose Transporter 1-Deficient Mice. Biol. Pharm. Bull. *29*, 1587–1591. https://doi.org/10.1248/bpb.29.1587.
28. Simon, C.S., Hur, W., Garg, V., Kuo, Y.-Y., Niakan, K.K., and Hadjantonakis, A.-K. (2025). ETV4 and ETV5 orchestrate FGF-mediated lineage specification and epiblast maturation during early mouse development. Development *152*, dev204278. https://doi.org/10.1242/dev.204278.
29. Lee, J.-Y., Park, S., Kim, K.-S., Ko, J.-J., Lee, S., Kim, K.P., and Park, K.-S. (2016). Novel Function of Sprouty4 as a Regulator of Stemness and Differentiation of Embryonic Stem Cells. Dev. Reprod. *20*, 171–177. https://doi.org/10.12717/DR.2016.20.2.171.
30. Gharibi, B., Ghuman, M., and Hughes, F.J. (2016). DDIT4 regulates mesenchymal stem cell fate by mediating between HIF1α and mTOR signalling. Sci. Rep. *6*, 36889. https://doi.org/10.1038/srep36889.
31. Chen, C.-Y., Lee, D.S., Yan, Y.-T., Shen, C.-N., Hwang, S.-M., Lee, S.T., and Hsieh, P.C.H. (2015). Bcl3 Bridges LIF-STAT3 to Oct4 Signaling in the Maintenance of Naïve Pluripotency. Stem Cell. *33*, 3468–3480. https://doi.org/10.1002/stem.2201.
32. Chappell, J., and Dalton, S. (2013). Roles for MYC in the Establishment and Maintenance of Pluripotency. Cold Spring Harb. Perspect. Med. *3*, a014381. https://doi.org/10.1101/cshperspect.a014381.
33. Wu, B., Li, Y., Li, B., Zhang, B., Wang, Y., Li, L., Gao, J., Fu, Y., Li, S., Chen, C., et al. (2021). DNMTs Play an Important Role in Maintaining the Pluripotency of Leukemia Inhibitory Factor-Dependent Embryonic Stem Cells. Stem Cell Rep. *16*, 582–596. https://doi.org/10.1016/j.stemcr.2021.01.017.
34. Suresh, B., Lee, J., Kim, H., and Ramakrishna, S. (2016). Regulation of pluripotency and differentiation by deubiquitinating enzymes. Cell Death Differ. *23*, 1257–1264. https://doi.org/10.1038/cdd.2016.53.
35. Wongtrakoongate, P., Li, J., and Andrews, P.W. (2014). DNMT3B inhibits the re-expression of genes associated with induced pluripotency. Exp. Cell Res. *321*, 231–239. https://doi.org/10.1016/j.yexcr.2013.11.024.
36. Gong, J., Gu, H., Zhao, L., Wang, L., Liu, P., Wang, F., Xu, H., and Zhao, T. (2018). Phosphorylation of ULK1 by AMPK is essential for mouse embryonic stem cell self-renewal and pluripotency. Cell Death Dis. *9*, 38. https://doi.org/10.1038/s41419-017-0054-z.
37. Katolikova, N.V., Khudiakov, A.A., Shafranskaya, D.D., Prjibelski, A.D., Masharskiy, A.E., Mor, M.S., Golovkin, A.S., Zaytseva, A.K., Neganova, I.E., Efimova, E.V., et al. (2023). Modulation of Notch Signaling at Early Stages of Differentiation of Human Induced Pluripotent Stem Cells to Dopaminergic Neurons. Int. J. Mol. Sci. *24*, 1429. https://doi.org/10.3390/ijms24021429.
38. Ishitani, T., Matsumoto, K., Chitnis, A.B., and Itoh, M. (2005). Nrarp functions to modulate neural-crest-cell differentiation by regulating LEF1 protein stability. Nat. Cell Biol. *7*, 1106–1112. https://doi.org/10.1038/ncb1311.
39. Yu, X., Zou, J., Ye, Z., Hammond, H., Chen, G., Tokunaga, A., Mali, P., Li, Y.-M., Civin, C., Gaiano, N., and Cheng, L. (2008). Notch signaling activation in human embryonic stem cells is required for embryonic but not trophoblastic lineage commitment. Cell Stem Cell *2*, 461–471. https://doi.org/10.1016/j.stem.2008.03.001.
40. Levine, A.J., and Brivanlou, A.H. (2006). GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. Development *133*, 209–216. https://doi.org/10.1242/dev.02192.
41. Yazawa, T., Imamichi, Y., Kitano, T., Islam, M.S., Khan, M.R.I., Takahashi, S., Sekiguchi, T., Suzuki, N., Umezawa, A., and Uwada, J. (2023). Expression of Chrna9 is regulated by Tbx3 in undifferentiated pluripotent stem cells. Sci. Rep. *13*, 1611. https://doi.org/10.1038/s41598-023-28814-7.
42. Sim, E.Z., Enomoto, T., Shiraki, N., Furuta, N., Kashio, S., Kambe, T., Tsuyama, T., Arakawa, A., Ozawa, H., Yokoyama, M., et al. (2022). Methionine metabolism regulates pluripotent stem cell pluripotency and differentiation through zinc mobilization. Cell Rep. *40*, 111120. https://doi.org/10.1016/j.celrep.2022.111120.
43. Isakova, A., Neff, N., and Quake, S.R. (2021). Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. Proc. Natl. Acad. Sci. USA *118*, e2113568118. https://doi.org/10.1073/pnas.2113568118.
44. Russo, M.W., Sevetson, B.R., and Milbrandt, J. (1995). Identification of NAB1, a repressor of NGFI-A- and Krox20-mediated transcription. Proc. Natl. Acad. Sci. USA *92*, 6873–6877. https://doi.org/10.1073/pnas.92.15.6873.
45. Falcon, A., Doege, H., Fluitt, A., Tsang, B., Watson, N., Kay, M.A., and Stahl, A. (2010). FATP2 is a hepatic fatty acid transporter and peroxisomal very long-chain acyl-CoA synthetase. Am. J. Physiol. Endocrinol. Metab. *299*, E384–E393. https://doi.org/10.1152/ajpendo.00226.2010.
46. Dong, S., Wang, Z., Huang, B., Zhang, J., Ge, Y., Fan, Q., and Wang, Z. (2017). Bioinformatics insight into glycosyltransferase gene expression in gastric cancer: POFUT1 is a potential biomarker. Biochem. Biophys.

Res. Commun. *483*, 171–177. https://doi.org/10.1016/j.bbrc.2016.12.172.

47. Wigge, P., Köhler, K., Vallis, Y., Doyle, C.A., Owen, D., Hunt, S.P., and McMahon, H.T. (1997). Amphiphysin Heterodimers: Potential Role in Clathrin-mediated Endocytosis. Mol. Biol. Cell *8*, 2003–2015.
48. Araki, T., Kusakabe, M., and Nishida, E. (2011). A Transmembrane Protein EIG121L Is Required for Epidermal Differentiation during Early Embryonic Development. J. Biol. Chem. *286*, 6760–6768. https://doi.org/10.1074/jbc.M110.177907.
49. Adams, J.M., Rege, S.V., Liu, A.T., Vu, N.V., Raina, S., Kirsher, D.Y., Nguyen, A.L., Harish, R., Szoke, B., Leone, D.P., et al. (2023). Leukotriene A4 hydrolase inhibition improves age-related cognitive decline via modulation of synaptic function. Sci. Adv. *9*, eadf8764. https://doi.org/10.1126/sciadv.adf8764.
50. Becic, A., Leifeld, J., Shaukat, J., and Hollmann, M. (2021). Tetraspanins as Potential Modulators of Glutamatergic Synaptic Function. Front. Mol. Neurosci. *14*, 801882. https://doi.org/10.3389/fnmol.2021.801882.
51. Jeffares, D.C., Tomiczek, B., Sojo, V., and dos Reis, M. (2015). A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome. In Parasite Genomics Protocols, C. Peacock, ed. (Springer), pp. 65–90. https://doi.org/10.1007/978-1-4939-1438-8_4.
52. Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature *267*, 275–276. https://doi.org/10.1038/267275a0.
53. Boroviak, T., Stirparo, G.G., Dietmann, S., Hernando-Herraez, I., Mohammed, H., Reik, W., Smith, A., Sasaki, E., Nichols, J., and Bertone, P. (2018). Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. Development *145*, dev167833. https://doi.org/10.1242/dev.167833.
54. Yang, S.-H., Kalkan, T., Morrisroe, C., Smith, A., and Sharrocks, A.D. (2012). A Genome-Wide RNAi Screen Reveals MAP Kinase Phosphatases as Key ERK Pathway Regulators during Embryonic Stem Cell Differentiation. PLoS Genet. *8*, e1003112. https://doi.org/10.1371/journal.pgen.1003112.
55. Yang, S.-H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 Drive Early Enhancer Activation during Embryonic Stem Cell Transition from Naive Pluripotency. Cell Rep. *7*, 1968–1981. https://doi.org/10.1016/j.celrep.2014.05.037.
56. Wang, X., Xiang, Y., Yu, Y., Wang, R., Zhang, Y., Xu, Q., Sun, H., Zhao, Z.-A., Jiang, X., Wang, X., et al. (2021). Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. Cell Res. *31*, 526–541. https://doi.org/10.1038/s41422-021-00477-x.
57. Cornacchia, D., Zhang, C., Zimmer, B., Chung, S.Y., Fan, Y., Soliman, M.A., Tchieu, J., Chambers, S.M., Shah, H., Paull, D., et al. (2019). Lipid Deprivation Induces a Stable, Naive-to-Primed Intermediate State of Pluripotency in Human PSCs. Cell Stem Cell *25*, 120–136.e10. https://doi.org/10.1016/j.stem.2019.05.001.
58. Haston, S., Pozzi, S., Carreno, G., Manshaei, S., Panousopoulos, L., Gonzalez-Meljem, J.M., Apps, J.R., Virasami, A., Thavaraj, S., Gutteridge, A., et al. (2017). MAPK pathway control of stem cell proliferation and differentiation in the embryonic pituitary provides insights into the pathogenesis of papillary craniopharyngioma. Development *144*, 2141–2152. https://doi.org/10.1242/dev.150490.
59. Gendrel, A.-V., and Heard, E. (2014). Noncoding RNAs and Epigenetic Mechanisms During X-Chromosome Inactivation. Annu. Rev. Cell Dev. Biol. *30*, 561–580. https://doi.org/10.1146/annurev-cellbio-101512-122415.
60. Kato, T., Nishimura, K., Hirao, M., Shimohama, S., and Takata, K. (2023). Expression and role of nicotinic acetylcholine receptors during midbrain dopaminergic neuron differentiation from human induced pluripotent stem cells. Neuropsychopharmacol. Rep. *43*, 440–445. https://doi.org/10.1002/npr2.12361.
61. Crow, M., Suresh, H., Lee, J., and Gillis, J. (2022). Coexpression reveals conserved gene programs that co-vary with cell type across kingdoms. Nucleic Acids Res. *50*, 4302–4314. https://doi.org/10.1093/nar/gkac276.
62. Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160. https://doi.org/10.1126/science.1248882.
63. Mitschka, S., Ulas, T., Goller, T., Schneider, K., Egert, A., Mertens, J., Brüstle, O., Schorle, H., Beyer, M., Klee, K., et al. (2015). Co-existence of intact stemness and priming of neural differentiation programs in mES cells lacking Trim71. Sci. Rep. *5*, 11126. https://doi.org/10.1038/srep11126.
64. Endo, Y., Kamei, K.I., and Inoue-Murayama, M. (2020). Genetic Signatures of Evolution of the Pluripotency Gene Regulating Network across Mammals. Genome Biol. Evol. *12*, 1806–1818. https://doi.org/10.1093/gbe/evaa169.
65. Liu, T., Li, J., Yu, L., Sun, H.-X., Li, J., Dong, G., Hu, Y., Li, Y., Shen, Y., Wu, J., and Gu, Y. (2021). Cross-species single-cell transcriptomic analysis reveals pre-gastrulation developmental differences among pigs, monkeys, and humans. Cell Discov. *7*, 8–17. https://doi.org/10.1038/s41421-020-00238-x.
66. Irie, N., and Kuratani, S. (2014). The developmental hourglass model: a predictor of the basic body plan? Development *141*, 4649–4655. https://doi.org/10.1242/dev.107318.
67. Bi, Y., Tu, Z., Zhou, J., Zhu, X., Wang, H., Gao, S., and Wang, Y. (2022). Cell fate roadmap of human primed-to-naive transition reveals preimplantation cell lineage signatures. Nat. Commun. *13*, 3147. https://doi.org/10.1038/s41467-022-30924-1.
68. Yang, Y., Liu, B., Xu, J., Wang, J., Wu, J., Shi, C., Xu, Y., Dong, J., Wang, C., Lai, W., et al. (2017). Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. Cell *169*, 243–257. https://doi.org/10.1016/j.cell.2017.02.005.
69. Zheng, R., Geng, T., Wu, D.Y., Zhang, T., He, H.N., Du, H.N., Zhang, D., Miao, Y.L., and Jiang, W. (2021). Derivation of feeder-free human extended pluripotent stem cells. Stem Cell Rep. *16*, 1686–1696. https://doi.org/10.1016/j.stemcr.2021.06.001.
70. Liu, C., Wang, R., He, Z., Osteil, P., Wilkie, E., Yang, X., Chen, J., Cui, G., Guo, W., Chen, Y., et al. (2018). Suppressing Nodal Signaling Activity Predisposes Ectodermal Differentiation of Epiblast Stem Cells. Stem Cell Rep. *11*, 43–57. https://doi.org/10.1016/j.stemcr.2018.05.019.
71. Bao, S., Tang, W.W., Wu, B., Kim, S., Li, J., Li, L., Kobayashi, T., Lee, C., Chen, Y., Wei, M., et al. (2018). Derivation of hypermethylated pluripotent embryonic stem cells with high potency. Cell Res. *28*, 22–34. https://doi.org/10.1038/cr.2017.134.
72. Vilarino, M., Alba Soto, D., Soledad Bogliotti, Y., Yu, L., Zhang, Y., Wang, C., Paulson, E., Zhong, C., Jin, M., Carlos Izpisua Belmonte, J., et al. (2020). Derivation of sheep embryonic stem cells under optimized conditions. Reproduction *160*, 761–772. https://doi.org/10.1530/REP-19-0606.
73. Debowski, K., Drummer, C., Lentes, J., Cors, M., Dressel, R., Lingner, T., Salinas-Riester, G., Fuchs, S., Sasaki, E., and Behr, R. (2016). The transcriptomes of novel marmoset monkey embryonic stem cell lines reflect distinct genomic features. Sci. Rep. *6*, 29122. https://doi.org/10.1038/srep29122.
74. Shiozawa, S., Nakajima, M., Okahara, J., Kuortaki, Y., Kisa, F., Yoshimatsu, S., Nakamura, M., Koya, I., Yoshimura, M., Sasagawa, Y., et al. (2020). Primed to Naive-Like Conversion of the Common Marmoset Embryonic Stem Cells. Stem Cells Dev. *29*, 761–773. https://doi.org/10.1089/scd.2019.0259.
75. Yoshimatsu, S., Nakajima, M., Sonn, I., Natsume, R., Sakimura, K., Nakatsukasa, E., Sasaoka, T., Nakamura, M., Serizawa, T., Sato, T., et al.

(2022). Attempts for deriving extended pluripotent stem cells from common marmoset embryonic stem cells. Genes Cells *28*, 156–169. https://doi.org/10.1111/gtc.13000.

76. Li, T.-B., Feng, G.-H., Li, Y., Wang, M., Mao, J.J., Wang, J.Q., Li, X., Wang, X.P., Qu, B., Wang, L.Y., et al. (2017). Rat embryonic stem cells produce fertile offspring through tetraploid complementation. Proc. Natl. Acad. Sci. USA *114*, 11974–11979. https://doi.org/10.1073/pnas.1708710114.
77. Oikawa, M., Kobayashi, H., Sanbo, M., Mizuno, N., Iwatsuki, K., Takashima, T., Yamauchi, K., Yoshida, F., Yamamoto, T., Shinohara, T., et al. (2022). Functional primordial germ cell–like cells from pluripotent stem cells in rats. Science *376*, 176–179. https://doi.org/10.1126/science.abl4412.
78. 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. USA *118*, e2018505118. https://doi.org/10.1073/pnas.2018505118.
79. Su, Y., Zhao, R., Fang, Y., Renxiu, M., Li, G., Jin, L., Liu, J., Yang, Z., Li, N., Zhu, J., et al. (2025). Bovine Formative Embryonic Stem Cell Plasticity in Embryonic and Extraembryonic Differentiation. Stem Cells, sxaf068. https://doi.org/10.1093/stmcls/sxaf068.
80. Zhang, M.L., Jin, Y., Zhao, L.H., Zhang, J., Zhou, M., Li, M.S., Yin, Z.B., Wang, Z.X., Zhao, L.X., Li, X.H., and Li, R.F. (2021). Derivation of Porcine Extra-Embryonic Endoderm Cell Lines Reveals Distinct Signaling Pathway and Multipotency States. Int. J. Mol. Sci. *22*, 12918. https://doi.org/10.3390/ijms222312918.
81. Meek, S., Watson, T., Eory, L., McFarlane, G., Wynne, F.J., McCleary, S., Dunn, L.E.M., Charlton, E.M., Craig, C., Shih, B., et al. (2022). Stem cell-derived porcine macrophages as a new platform for studying host-pathogen interactions. BMC Biol. *20*, 14. https://doi.org/10.1186/s12915-021-01217-8.
82. Gao, X., Nowak-Imialek, M., Chen, X., Chen, D., Herrmann, D., Ruan, D., Chen, A.C.H., Eckersley-Maslin, M.A., Ahmad, S., Lee, Y.L., et al. (2019). Establishment of porcine and human expanded potential stem cells. Nat. Cell Biol. *21*, 687–699. https://doi.org/10.1038/s41556-019-0333-2.
83. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
84. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
85. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.
86. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., et al. (2002). The Ensembl genome database project. Nucleic Acids Res. *30*, 38–41. https://doi.org/10.1093/nar/30.1.38.
87. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883. https://doi.org/10.1093/bioinformatics/bts034.
88. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinf. *17*, 483. https://doi.org/10.1186/s12859-016-1323-z.
89. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
90. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.
91. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
92. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Systems *1695*.
93. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7–W13. https://doi.org/10.1093/nar/gkq291.
94. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
95. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.
96. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. *9*, ELIXIR-709. https://doi.org/10.12688/f1000research.24956.2.
97. Sayols, S. (2023). rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. MicroPubl. Biol. *2023*, 10.17912/micropub.biology.000811. https://doi.org/10.17912/micropub.biology.000811.
98. Ignatiadis, N., Klaus, B., Zaugg, J.B., and Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat. Methods *13*, 577–580. https://doi.org/10.1038/nmeth.3885.
99. Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is My Network Module Preserved and Reproducible? PLoS Comput. Biol. *7*, e1001057. https://doi.org/10.1371/journal.pcbi.1001057.
100. Wang, Z.Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., et al. (2020). Transcriptome and translatome co-evolution in mammals. Nature *588*, 642–647. https://doi.org/10.1038/s41586-020-2899-z.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Human naive ESC RNA-seq datasets	Bi et al.[67]	GSE174771
Human primed ESC RNA-seq datasets	Yang et al.[68]	GSE80732
Human formative ESC RNA-seq datasets	Kinoshita et al.[11]	GSE131551
Human extended ESC RNA-seq datasets	Yang et al.[68]	GSE80732
Human extended ESC RNA-seq datasets	Zheng et al.[69]	GSE137208
Mouse naive/extended ESC RNA-seq datasets	Yang et al.[68]	GSE80732
Mouse formative ESC RNA-seq datasets	Kinoshita et al.[11]	GSE131553
Mouse formative ESC RNA-seq datasets	Yu et al.[5]	GSE135989
Mouse primed ESC RNA-seq datasets	Liu et al.[70]	GSE92634
Mouse naive ESC RNA-seq datasets	Bao et al.[71]	GSE99491
Crab-eating macaque naive/primedESC RNA-seq datasets	Chen et al.[31]	GSE69708
Sheep primed ESC RNA-seq datasets	Vilarino et al.[72]	PRJNA609175
Horse formative ESC RNA-seq datasets	Yu et al.[5]	GSE135991
Marmoset primed ESC RNA-seq datasets	Debowski et al.[73]	GSE70897
Marmoset naive/primed ESC RNA-seq datasets	Shiozawa et al.[74]	GSE138944
Marmoset extended ESC RNA-seq datasets	Yoshimatsu et al.[75]	GSE189633
Rat naive ESC RNA-seq datasets	Li et al.[76]	GSE97966
Rat naive/formative ESC RNA-seq datasets	Iwatsuki et al.[9]	GSE220805
Rat naive ESC RNA-seq datasets	Oikawa et al.[77]	GSE178701
Cattle primed ESC RNA-seq datasets	Bogliotti et al.[10]	GSE110036
Cattle expanded ESC RNA-seq datasets	Zhao et al.[78]	GSE129760
Cattle formative ESC RNA-seq datasets	Su et al.[79]	GSE240176
Pig naive/primed ESC RNA-seq datasets	Zhang et al.[80]	GSE183270
Pig expanded ESC RNA-seq datasets	Meek et al.[81]	PRJNA787759
Pig primed ESC RNA-seq datasets	Choi et al.[20]	GSE120031
Pig expanded ESC RNA-seq datasets	Gao et al.[82]	E-MTAB-7253
Pig formative ESC RNA-seq datasets	Zhi et al.[21]	PRJCA004466
Software and algorithms		
SRA Toolkit (v3.0.5)	NCBI	https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit
fastp (v0.23.4)	Chen et al.[83]	https://github.com/OpenGene/fastp
STAR (v2.7.10b)	Dobin et al.[84]	https://github.com/alexdobin/STAR
RSEM (v1.3.3)	Li and Dewey[85]	https://deweylab.github.io/RSEM/
Ensembl Genome Browser (Release 109)	Hubbard et al.[86]	https://www.ensembl.org
SVA (v3.50.0)	Leek et al.[87]	https://bioconductor.org/packages/sva
variancePartition (v1.32.3)	Hoffman and Schadt[88]	https://bioconductor.org/packages/variancePartition
DESeq2 (v1.42.0)	Love et al.[89]	https://bioconductor.org/packages/DESeq2
UpSetR (v1.4.0)	Conway et al.[90]	https://cran.r-project.org/package=UpSetR
ClusterGVis (v0.1.1)	Junjun Lab	https://github.com/junjunlab/ClusterGVis
WGCNA (v1.72-5)	Langfelder and Horvath[91]	https://cran.r-project.org/package=WGCNA

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
igraph (v2.0.3)	Csárdi et al.[92]	https://igraph.org/r/
networkD3 (v0.4)	Christopher Gandrud	https://github.com/christophergandrud/networkD3
TranslatorX	Abascal et al.[93]	http://translatorx.co.uk
RAxML (v8.2.12)	Stamatakis et al.[94]	https://cme.h-its.org/exelixis/web/software/raxml/
PAML (v4.10.6)	Yang et al.[95]	https://github.com/abacus-gene/paml
gprofiler2 (v0.2.3)	Kolberg et al.[96]	https://cran.r-project.org/package=gprofiler2
rrvgo (v1.14.2)	Sayols et al.[97]	https://bioconductor.org/packages/rrvgo
Custom analysis code	This paper	https://github.com/coderFaye/CS-ESC

METHOD DETAILS

In this study, we performed a comparative transcriptomic analysis of embryonic stem cells across multiple mammalian species. Publicly available RNA-seq datasets representing distinct pluripotency states were integrated and uniformly processed for downstream analyses. Detailed descriptions of each analytical step are provided in the following sections.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Curation and quantification of transcriptome data

A total of 120 RNA-seq samples from embryonic stem cells (ESCs) at various pluripotent and totipotent states (expanded or extended, naïve, formative, and primed) were collected across nine species: human, mouse, rat, marmoset, crab-eating macaque, cattle, pig, sheep, and horse. These samples, derived from 28 studies, were sourced from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/). The specific references for the RNA-seq data are provided, and detailed information for all samples is summarized in Table S1. For certain species, RNA-seq data for specific pluripotency stages were unavailable in public databases (last searched on January 1, 2024), and these cell types were excluded from further analysis.

All samples underwent a standardized upstream analysis pipeline. The RNA-seq data were downloaded using the SRA Toolkit v3.0.5 (https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit) with the fastq-dump function. Quality control was performed using fastp v0.23.4[83] with the parameters: -f 15 -F 15 -q 20 -u 40 -n 8 -L 30. Subsequently, alignment was carried out with STAR v2.7.10b,[84] mapping each sample to its corresponding genome (Table S2) using the parameters: –quantMode TranscriptomeSAM –outSAMtype BAM SortedByCoordinate –outSAMunmapped Within –readFilesCommand zcat. Finally, RSEM v1.3.3[85] was used for gene-level quantification with default settings.

Ortholog retrieval and data normalization

We retrieved the one-to-one orthologs for the species in our dataset from the Ensembl database release 109,[86] identifying 9,941 orthologous genes with 1:1 orthology relationships among all the species studied. The RNA-seq samples from different species were integrated using these orthologous genes, enabling cross-species comparisons based on the merged data. TPM normalization was performed, with values calculated using RSEM, and subsequently transformed to log2 (N + 1) values (log2-TPM). To detect and correct for hidden biases like those arising from variations in sampling conditions and sequencing protocols across different experiments, we employed Surrogate Variable Analysis (SVA) method using the SVA R package (v3.50.0),[87] which identifies latent sources of variation in high-dimensional genomic data. By estimating and adjusting for these surrogate variables, SVA effectively reduces the impact of unmodeled batch effects and other confounding factors. This approach generated SVA-log2-TPM values, which were subsequently used for further analysis.

Data exploratory analysis

Principal Component Analysis (PCA) was performed using the prcomp function in R, based on the SVA-log2-TPM values. Pearson correlation analysis was conducted on the raw counts, log2-TPM values, and SVA-log2-TPM values, both before and after normalization. To further dissect sources of gene expression variation, we applied variance partitioning analysis, a statistical approach that estimates the proportion of variance in gene expression attributable to different factors. This analysis was implemented using the variancePartition R package (v1.32.3),[88] with species and state as fixed effects in a linear mixed model.

Differential gene expression analysis

We employed the DESeq2 R package (v1.42.0)[89] to identify differentially expressed genes (DEGs) in each species by comparing nESCs to EpiSCs and across successive states: EPSCs to nESCs, nESCs to FSCs, and FSCs to EpiSCs. To control for multiple hypothesis testing and improve statistical power, we applied Independent Hypothesis Weighting.[98] Genes exhibiting an absolute log2

fold-change greater than 1 and an adjusted *p*-value less than 0.05 were classified as DEGs. We then performed an overlap analysis of DEGs across different species to identify those that are commonly shared using the UpSetR R package (v1.4.0).[90]

Identification of *de novo* state-specific markers

State-specific markers were defined as genes significantly upregulated in a specific state compared to all other states. Marker identification was first performed within each species to capture intra-species specificity. To examine expression dynamics across states, we applied Mfuzz, a soft clustering approach well-suited for capturing gradual changes in gene expression across conditions. Genes with similar expression patterns across the four states are assigned to the same Mfuzz cluster, each representing a distinct expression trajectory. Clustering was performed using the clusterGVis R package (v0.1.1) (https://github.com/junjunlab/ClusterGVis) to group genes into predominant expression trajectories. Subsequently, cross-species comparisons were conducted to identify both conserved markers shared across species and those unique to individual species, providing insights into universal and species-specific pluripotency regulation.

Gene co-expression network construction and preservation analysis

To identify modules of co-expressed genes and explore their association with biological traits, we applied Weighted Gene WGCNA. We utilized the WGCNA R package (v1.72-5)[91] to construct gene co-expression networks for each species individually. The blockwiseModules function from WGCNA in R was employed for the analysis, using SVA-log-TPM data as input. The soft-thresholding power was determined with the pickSoftThreshold function to optimize the scale-free topology fit of the network. A signed network was constructed to retain the sign of the correlation in the adjacency matrix. To assess the conservation of gene co-expression patterns across species, we performed module preservation analysis, which quantifies how well modules identified in one co-expression network are retained in another. Using network statistics, we conducted a preservation test with the modulePreservation function from WGCNA package, taking the human network as the reference. Z-scores were estimated to measure the preservation, with $Z > 10$ suggesting strong preservation, $2 < Z < 10$ representing weak preservation and $Z < 2$ meaning no preservation.[99] Hub genes within the networks were identified based on the hub centrality scores via Kleinberg's metric using the hub_score function from the igraph v2.0.3 package.[92] The relationships among genes within the modules were visualized using networkD3 R package (v0.4) (https://github.com/christophergandrud/networkD3).

Gene expression phylogenies

To infer evolutionary relationships among species from a transcriptomic perspective, we reconstructed gene expression phylogenies using orthologous gene expression profiles. Aligned with Brownian-motion-based models which assumes that gene-expression evolution occurs through a succession of independent changes in gene-expression levels, we employed a variance-based method to estimate expression divergence between each pair of species.[100] Distance matrices were calculated for each pluripotency state based on the SVA-log2-TPM values of all orthologous genes. Using these distance matrices, we constructed phylogenetic trees using the neighbor-joining (NJ) approach. The reliability of the branching patterns was assessed through bootstrap analyses, wherein 1:1 orthologous genes were randomly sampled with replacement 1000 times. The bootstrap values represent the proportions of replicate trees that support the branching pattern of the consensus tree displayed in the figures.

dN/dS ratio calculation

To assess evolutionary pressures acting on protein-coding genes, we calculated the dN/dS ratio, which compares the rate of nonsynonymous (dN) to synonymous (dS) substitutions. A dN/dS ratio significantly lower than 1 indicates purifying selection, whereas a ratio near or above 1 suggests relaxed or positive selection, respectively. A lower dN/dS ratio reflects stronger evolutionary constraint, suggesting that the protein-coding sequence is functionally important and has been conserved to maintain its biological role. The coding sequences (CDS) of all orthologous genes were retrieved from the Ensembl database. Multiple sequence alignments were performed using TranslatorX[93] with MAFFT as the alignment method. For phylogenetic reconstruction, RAxML v8.2.12[94] was used to infer maximum-likelihood gene trees. The ratio of nonsynonymous substitutions per nonsynonymous site (dN) to synonymous substitutions per synonymous site (dS) was calculated using PAML (codeml) v4.10.6.[95] A homogeneous model was applied (model = 0, NSsites = 0) to estimate the overall dN/dS ratio for each orthologous gene. Additionally, we examined the relationships between dN/dS ratios and two gene properties: gene expression levels and gene connectivity (i.e., the number of direct interactions a gene has with other genes). Gene connectivity was derived from WGCNA. To further investigate the interplay among these factors, we performed a correlation and partial correlation analysis to assess how dN/dS, gene expression, and connectivity interact. Specifically, we computed Spearman correlations for all pairwise comparisons and further estimated partial correlations to disentangle direct and indirect relationships between these factors. To ensure robustness, we applied a bootstrapping approach (1,000 replicates) to calculate confidence intervals for the partial correlations.

Gene ontology enrichment analyses

Gene ontology analysis is conducted using gprofiler2 R package (v 0.2.3).[96] To facilitate interpretation of enriched biological processes, we performed GO term reduction to summarize redundant and overlapping categories. GO lists were grouped by similar terms based on their semantic similarity to reduce redundancy by rrvgo R package (v1.14.2).[97]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters were reported either in individual figures or corresponding figure legends. Statistical details of experiments can be found in method details. All statistical analyses were performed in R.