

# Single-cell RNA-Seq reveals the earliest lineage specification and X chromosome dosage compensation in bovine preimplantation embryos

Bingjie Hu  | Hao Jin | Yan Shi | Haotian Yu | Xiaotong Wu | Shaohua Wang |  
 Kun Zhang 

Key Laboratory of Dairy Cow Genetic Improvement and Milk Quality Research of Zhejiang Province, College of Animal Sciences, Zhejiang University, Hangzhou, Zhejiang, China

## Correspondence

Kun Zhang, Key Laboratory of Dairy Cow Genetic Improvement and Milk Quality Research of Zhejiang Province, College of Animal Sciences, Zhejiang University, Room 301 E Building, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, China.

Email: [kzhang@zju.edu.cn](mailto:kzhang@zju.edu.cn)

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 32072731 and 32161143032; National Key R&D Program of China, Grant/Award Number: 2022YFE0100200; Qinghai Science and Technology Major Program, Grant/Award Number: 2021-NK-A5; Foundation of Key Laboratory of Veterinary Biotechnology, Grant/Award Number: BD1500010

## Abstract

Lineage specification and X chromosome dosage compensation are two crucial biological processes that occur during preimplantation embryonic development. Although extensively studied in mice, the timing and regulation of these processes remain elusive in other species, including humans. Previous studies have suggested conserved principles of human and bovine early development. This study aims to provide fundamental insights into these programs and the regulation using a bovine embryo model by employing single-cell transcriptomics and genome editing approaches. The study analyzes the transcriptomes of 286 individual cells and reveals that bovine trophectoderm/inner cell mass transcriptomes diverge at the early blastocyst stage, after cavitation but before blastocyst expansion. The study also identifies transcriptomic markers and provides the timing of lineage specification events in the bovine embryo. Importantly, we find that SOX2 is required for the first cell decision program in bovine embryos. Moreover, the study shows the occurrence of X chromosome dosage compensation from morula to late blastocyst and reveals that this compensation results from downregulation of X-linked genes in female embryonic cells. The transcriptional atlas generated by this study is expected to be widely useful in improving our understanding of mammalian early embryo development.

## KEY WORDS

cattle, embryo, pluripotency, single cell, SOX2, X dosage compensation

**Abbreviations:** BL\_ICM, wildtype ICM cells of blastocyst stage; BL\_KO, knockout cells of blastocyst stage; BL\_TE, wildtype TE cells of blastocyst stage; COCs, cumulus oocyte complexes; EB, early blastocyst; EPI, epiblast; ICM, inner cell mass; IVC, in vitro culture; IVF, in vitro fertilization; IVM, in vitro maturation; KO, knockout; LB, late blastocyst; MB, middle blastocyst; MO\_KO, knockout cells of morula stage; MO\_NC, wildtype cells of morula stage; OxPhos, oxidative phosphorylation; PE, primitive endoderm; TE, trophectoderm; UMAP, uniform manifold approximation and projection approach; WT, wildtype; XCI, X chromosome inactivation.

Bingjie Hu and Hao Jin contributed equally to this work.

© 2024 Federation of American Societies for Experimental Biology.

## 1 | INTRODUCTION

From zygote to blastocyst stage in mammals, embryos experience a series of cleavage and morphological transformations, acquire pluripotency, and proceed with the first cell fate decision. This developmental process results in two unique cell types: the trophectoderm (TE), contributing to the placenta, and the inner cell mass (ICM) cells. The ICM cells further differentiate into the epiblast (EPI), responsible for all fetal cells, and the primitive endoderm (PE), contributing predominantly to the extra-embryonic yolk sac.

While the preimplantation development shares similarities across mammals, distinct biological variations exist between species. In mouse embryos, the first lineage specification is believed to occur in the morula stage where cells differentiate into ICM cells and TE cells.<sup>1,2</sup> Subsequently, a second lineage specification takes place at the blastocyst stage, separating ICM cells into EPI and PE, driven by a Nanog/Fgf4/Gata6 axis.<sup>3</sup> However, lineage specification events in human embryos remain a topic of debate with two prevailing viewpoints. One perspective proposes that the establishment of TE, EPI, and PE lineages occurs simultaneously rather than following a two-step model like in mice.<sup>4</sup> Conversely, several single-cell transcriptome studies conducted on early human embryos suggest that ICM cells are transcriptionally distinct from EPI and PE, implying a two-step model akin to mice.<sup>5,6</sup>

However, studies on lineage species and pluripotency establishment in other species are limited. A couple of studies attempt to examine lineage differentiation at the level of individual cells, providing references for single-cell transcriptome study of bovine embryos but with a restricted number of cells and genes.<sup>7,8</sup> The mechanisms within bovine embryos remain elusive.

X chromosome dosage compensation is another important process during preimplantation development. Studies have shown that X chromosome dosage compensation dynamics vary between species and even between different stages of embryonic development.<sup>9</sup> In mouse embryos, two waves of X chromosome inactivation (XCI) occur with the first wave occurring at the four-cell stage and the second wave occurring in the epiblast.<sup>9–11</sup> In contrast to mice, chromosome-wide XCI is not established even by the blastocyst stage in humans, although the accumulation of XIST from the morula stage has been observed.<sup>4,9</sup> However, there is still much to be explored about its regulation and timing in different species.

Here, we use bovine embryo as a model and employ single-cell RNA sequencing technology to elucidate the transcriptional dynamics underlying the lineage specification, the emergence of pluripotency as well as X chromosome dosage compensation during preimplantation

development. This research provides valuable resources and fundamental knowledge for understanding lineage specification and X chromosome dosage compensation during bovine preimplantation development.

## 2 | MATERIALS AND METHODS

### 2.1 | Bovine embryo production

In vitro production of bovine embryos, including in vitro maturation (IVM), in vitro fertilization (IVF), and in vitro culture (IVC), was performed as published before.<sup>12</sup> Briefly, cumulus oocyte complexes (COCs) containing intact cumulus cells were collected from bovine ovaries obtained from a local abattoir and these COCs were matured in the IVM medium at 38.5°C under 5% CO<sub>2</sub> in humidified air for 22–24 h. The IVM medium consists of 10% fetal bovine serum (FBS; Gibco-BRL), 1 IU/mL follicle-stimulating hormone (Sansheng Biological Technology, Ningbo, China), 0.1 IU/mL luteinizing hormone (Solarbio, Beijing, China), 1 mM sodium pyruvate (Thermo Fisher Scientific), 2.5 mM GlutaMAX™ (Thermo Fisher Scientific), and 10 µg/mL gentamicin in Medium-199 (M4530). After maturation, COCs (60–100 COCs per well in 4-well plates) were then incubated with spermatozoa ( $1 \times 10^6$ – $5 \times 10^6$ ) purified from frozen-thawed semen using a percoll gradient in BO-IVF medium (IVF Bioscience, Falmouth, Cornwall, UK). The IVF condition was 38.5°C under 5% CO<sub>2</sub> for 9–12 h. Putative zygotes were then removed from cumulus cells by pipetting up and down in Medium-199 (M7528) supplemented with 2% FBS (Gibco-BRL). Embryos were incubated in BO-IVC medium (IVF Bioscience) at 38.5°C under 5% CO<sub>2</sub> in humidified air until use. Time post-fertilization and morphological states jointly determine the samples we collected. Based on morphological characteristic mentioned above, morula were collected on Day 6, while early blastocysts were harvested on Day 7. Middle blastocysts were collected on Day 8 or 9, and late blastocysts were obtained on Day 9. Twenty-seven embryos comprising six morula, six early blastocysts, eight middle blastocysts, and seven late blastocysts were collected (Figure S1A). Out of the consideration of embryo heterogeneity, we randomly selected 288 cells for single-cell library construction after isolating individual cells.

### 2.2 | Single-cell isolation and library construction

The zona pellucidae were removed using 0.5% pronase E (Sigma). Morula and early blastocysts were dissociated by

mouth pipette under the microscope (Nikon). Single-cell dissociation of middle and late blastocysts was performed by incubation in TrypLE Express (Gibco) for 10–15 min at 37°C and repeated pipetting using thin capillaries under embryo microsystems (Eppendorf). Individual cells were washed in 0.05% BSA (Sigma) and manually collected into PCR tubes that contained 2.5 μL lysis buffer with 2 Units RNase Inhibitor (Promega), 0.2% Triton X 100 (Sigma), 2 mM deoxynucleotide triphosphate (dNTP) mixture (Promega), and 0.6 μM barcode reverse transcriptase (RT) primers. Single-cell library construction were performed according to the protocol described previously.<sup>13,14</sup> In brief, single cells were incubated at 72°C for 3 min with lysis mixture. Reverse transcription and PCR pre-amplification were carried out with SuperScript II (Invitrogen) and KAPA HiFi HotStart ReadyMix (KAPA Biosystems). After PCR amplification, the samples with different barcodes were mixed together and then purified with DNA Clean&Concentrator-5 (Zymo Research). The cDNAs of single cells were added with biotinylated index and purified using beads (Vazyme). The cDNAs were sheared into approximately 300 bp fragments using sonication (Covaries). The fragmented cDNAs were enriched by Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen). Adaptor ligation and final library were conducted by NEBNext Library Quant Kit (New England Biolabs). The single-cell libraries were then sent to Novogene for sequencing with PE150 strategy on NovaSeq (Illumina) platform.

## 2.3 | sgRNA synthesis and in vitro transcription

BE-Designer online software (<http://www.rgenome.net>) was used to design sgRNAs. SgRNA sequences with appropriate GC content and low probability for off-target were selected to target the coding region of the gene of interest. The sticky end of BpiI: 5'-3' CACC and 5'-3' AAAC was added to the 5' ends of the sense and antisense strand, respectively (Data S4). The DNA sequences were synthesized by SangonCo., LTD (Shanghai, China). Then, sgRNA DNA oligos were annealed and cloned into a PX458 vector containing BpiI restriction sites with a T7 promoter.

BE3 plasmids were purchased from Addgene. After linearization with NotI, the plasmid underwent in vitro transcription using mMESSAGE mMACHINE T7 kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and was purified by LiCl precipitation. sgRNAs were amplified and transcribed in vitro using MEGAshortscript T7 High Yield Transcription Kit (Invitrogen, Thermo Fisher Scientific) according to manufacturer's instructions. Primers are listed in Data S4. After transcription, sgRNAs were purified by ethanol precipitation.

## 2.4 | Genotyping

After single-cell collection, the rest of cells of each embryo were, respectively, collected into a tube containing lysis buffer (40 nM Tris-HCl, 1% NP-40, 1% Triton X-100 and 0.4 ng/mL Proteinase K). Then, samples were incubated at 55°C for 1 hour and 95°C for 10 min as PCR template. Nested PCR was performed with rTaq (Takara), and PCR product fragments were sent to sequenced with Sanger sequencing. PCR conditions were as follows: 98°C for 2 min followed by 35 cycles of 98°C for 10 s, 60°C for 5 s, 72°C for 1.5 min, and a final 5-min step at 72°C. All primers are listed in Data S4.

## 2.5 | Immunofluorescence (IF)

Embryos were first washed three times in 0.1% PBS/PVP (polyvinylpyrrolidone) and fixed with 4% paraformaldehyde in PBS for 30 min, permeabilized with 0.5% Triton X-100/PBS for 30 min. Then, fixed samples were blocked for 1 h using the buffer containing 10% FBS (fetal bovine serum) and 0.1% Triton X-100/PBS, followed by incubation with primary antibody overnight at 4°C. Next, embryos were washed three times in 0.1% PBS/PVP and transferred into secondary antibody for 2 h at room temperature. Embryos were stained by DAPI for 15 min and imaged in glass slides under Zeiss LSM880 confocal microscope system (Zeiss, Oberkochen, Germany). All antibodies are listed in Data S4.

## 2.6 | Base editing mRNA microinjection

In total, 10–20 pL mixture of 100 ng/μL sgRNA and 200 ng/μL BE3 mRNA were microinjected into bovine zygotes at 12 h post insemination (hpi) by using a micromanipulator (TransferMan, Eppendorf, Germany). Control embryos were injected with the same amount of mRNA without sgRNA. To maximize the editing efficiency of the gene of interest, a cocktail of two or three sgRNAs was microinjected together with BE3 mRNA. Each sgRNA was kept at the same concentration (100 ng/μL).

## 2.7 | Single-cell RNA-seq data pre-processing

Raw reads were trimmed by Trimgalore (version 0.6.7), a wrapper around Cutadapt (version 0.4.0) to remove adapter sequences and low quality ends. Trimmed reads were mapped to ARS-UCD1.2 (cattle) using Hisat2

(version 2.1.0).<sup>15</sup> FeatureCounts (version 1.6.3) was used to count the number of reads aligned to each gene.<sup>16</sup> Gene expression levels were estimated in terms of reads per kilobase exon model and per million mapped reads (RPKM) using Cufflinks (version 2.2.1).<sup>16</sup>

## 2.8 | Cell clustering and differential expression analysis

The R package Seurat (version 4.1.1) was applied to normalize read counts of genes for each good quality cell and do cell clustering and differential expression analysis.<sup>17</sup> Genes expressed in less than three cells were filtered out, and cells with expressed genes less than 500 were excluded. Uniform manifold approximation and projection (UMAP) was used to identify the relations between cells using normalized read counts with the Seurat. Each cluster was defined by specific marker genes highlighted on UMAP graph using “FeaturePlot” function. “FindAllMarkers” and “FindMarkers” functions (test.use = “wilcox”) and DEsingle with default parameters were used for differential gene expression analysis between clusters. Gene ontology and Kyoto Encyclopaedia of Genes and Genomes analysis were performed with the Database for Annotation, Visualization, and Integrated Discovery (DAVID).<sup>18,19</sup>

## 2.9 | Pseudotime analysis

For scRNA-seq data, the pseudotemporal analysis was performed using the R package Monocle2.<sup>20-22</sup> Briefly, differentially expressed genes were detected at corresponding state of each cell using Monocle2. After cell ordering, trajectories were constructed by “plot\_cell\_trajectory” function. Then, the expression levels of marker genes were plotted according to pseudotemporal ordering.

## 2.10 | RNA velocity

RNA velocity was performed from BAM of samples that have passed all quality control in the final counts table. First, we used *velocyto.py* using the command *velocyto run*, with the parameter-logic as “SmartSeq2,” and the parameter-m (RepeatMasker annotations) as a GTF downloaded from the UCSC genome browser. The global GTF was the same that were used for the computation of raw counts table. *velocyto.R* was used for computing Velocity matrix.

## 2.11 | Motif enrichment

Genomic regions were evaluated for binding motif enrichment using the *findMotifsGenome.pl* script from HOMER (v4.8).<sup>23</sup> The most significant known or de novo motifs were reported based on p value.

## 2.12 | Module scores

The “AddModuleScore” function from the Seurat was used to calculate the expression score for the glycolysis and OxPhos gene sets. The module score function facilitates the computation of the average expression of genes within a specific module at the single-cell level. This involves subtracting a control cell score. Control cell scores are determined by binning all genes according to their average expression and then randomly selecting a specified number of genes from each bin. Consequently, a positive module score indicates an elevated expression of a particular set of genes within a cell, compared to a randomly assembled gene set. Default settings were applied.

## 2.13 | Comparison of identity fraction

The identity fraction is computed through R package MuSiC.<sup>24</sup> The package use deconvolution analysis to estimate the similarity of distinctive cell type. The raw expression date with overlapping gene (FPKM > 0) of BL ICM and SOX2 KO was as reference data.

## 2.14 | Diffusion component analysis

Diffusion component analysis is the method based on nonlinear dimensionality reduction to reconstruct the developmental bifurcation and pseudotime of the target cells. The map was constructed with the R package Destiny, using default parameters.<sup>25</sup>

## 3 | RESULTS

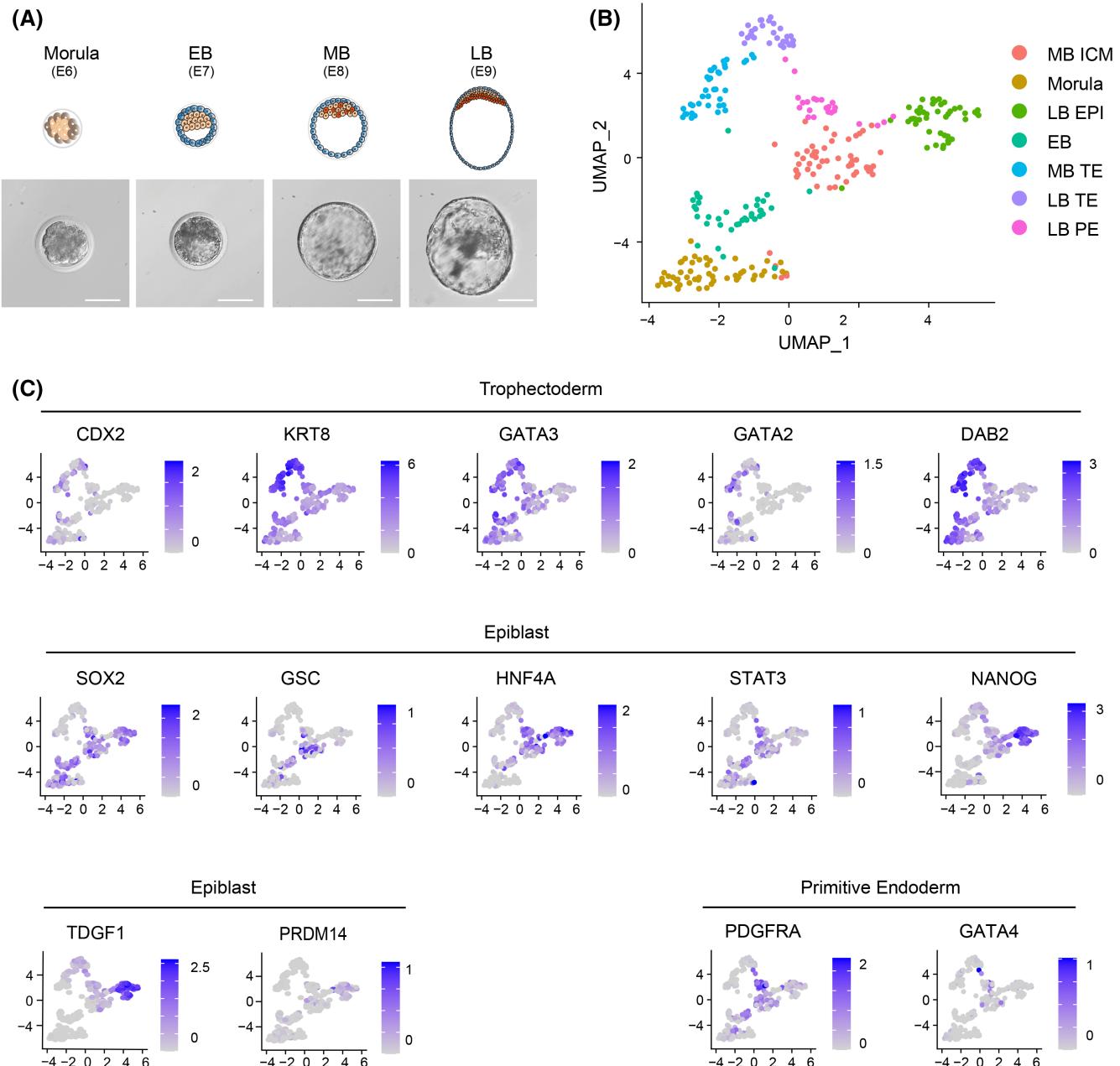
### 3.1 | Single-cell transcriptomic analysis of bovine peri-cavitating embryos

In our study, we harvested bovine embryos from four distinct developmental stages: morula, early blastocyst (EB), middle blastocyst (MB), and late blastocyst (LB; Figure 1A and Figure S1A). These stages were identified

based on their morphological characteristics (Figure 1A). We referred to Manual of the International Embryo Transfer Society (<https://www.iets.org/Publications/IETS-Manual>) as a standard for labeling the developmental stages of the embryo. Blastocysts with an intact zona pellucida whose cavity was less than half of the total volume of the embryo were classified as EB stage, corresponding to the code 5. Blastocysts with thin zona pellucida whose cavity completely occupies the total volume of the embryo were classified as MB stage, equal to the

code 7. The hatched blastocysts escaping from the zona pellucida were classified as LB stage, corresponding to the code 8. We then performed single-cell RNA sequencing on cells isolated from these embryos. Eventually, we analyzed 286 single-cell transcriptomes after excluding cells that did not meet data quality thresholds (Figure S1B).

On average, 7450 genes were expressed in each cell (Figure S1B). To delineate the first cell fate decision program in bovine preimplantation embryos based on transcriptomic identity, we employed a Uniform Manifold



**FIGURE 1** Single-cell transcriptomic analysis of peri-cavitating embryos in cattle. (A) Collection of embryos for single-cell RNA-seq at the morula, early blastocyst (EB), middle blastocyst (MB), and late blastocyst (LB) stages. (B) A Uniform Manifold Approximation and Projection (UMAP) plot of all cells (286 in total) reveals cell heterogeneity with seven distinct clusters. Colors indicate developmental stages. (C) Lineage-specific genes are displayed in UMAP plots.

Approximation and Projection (UMAP) approach. The UMAP analysis grouped the cells into two dimensions (**Figure 1B**), facilitating a clear visualization of lineage segregation. We annotated each cell cluster by using known lineage-specific genes (**Figure 1C**). Trophectoderm (TE) markers included *CDX2*, *KRT8*, *GATA3*, *GATA2*, and *DAB2*, while *SOX2*, *GSC*, *HNF4A*, *STAT3*, *NANOG*, *TDGF1*, and *PRDM14* were used as epiblast (EPI) markers. *PDGFRA* and *GATA4* were used to classify primitive endoderm (PE) cluster.<sup>26</sup> Notably, *POU5F1/OCT4* was not included as a pluripotency/epiblast marker due to its universal expression across all bovine morula and blastocyst cells, aligning with observations in human and monkey counterparts.<sup>27</sup>

Our analysis identified seven distinct clusters: Morula, EB, MB ICM, MB TE, LB EPI, LB PE, and LB TE. Interestingly, despite the EB cluster not segregating into different lineages, some EB cells exhibited expression of TE-specific genes like *CDX2*, *GATA2*, *GATA3*, and *DAB2*, while others showed expression of pluripotency genes such as *GSC*, *HNF4A*, and *STAT3* (**Figure 1C**). These findings suggest that the first lineage segregation occurs after the blastocyst cavity formation in bovine embryos.

### 3.2 | TE and ICM begin to separate after blastocyst formation with TE fate specified earlier

To ascertain whether the first lineage segregation occurs at the EB stage, we conducted a pseudotime analysis during bovine preimplantation development. By associating developmental stage annotations with the analysis, we charted the developmental trajectories (**Figure 2A** and **Figure S2A**). The result revealed that the first obvious bifurcation point lies within EB cluster. This result suggests that the first lineage decision is made at the EB stage. This is in contrast to mouse embryos where the first lineage segregation occurs at the morula stage,<sup>28</sup> but consistent with human embryos where it occurs at the blastocyst stage.<sup>29</sup> The first lineage segregation event was further corroborated by analyzing the expression pattern of known TE and ICM markers along the developmental paths (**Figure S2B**).

To validate the stage at which the first lineage segregation occurs, we performed RNA velocity analysis which considers both spliced and un-spliced mRNA counts to

predict potential directionality and speed of cell state transitions. The analysis represents these transitions as vectors, with long vectors indicating rapid differentiation events. RNA velocity analysis revealed a segregation tendency at the morula stage, followed by a more pronounced tendency at the EB stage (**Figure 2A,B**). Interestingly, we found velocity pattern toward TE fate (longer vectors) was stronger than those toward ICM fate (shorter vectors), indicating an earlier specification of TE cells than ICM cells.

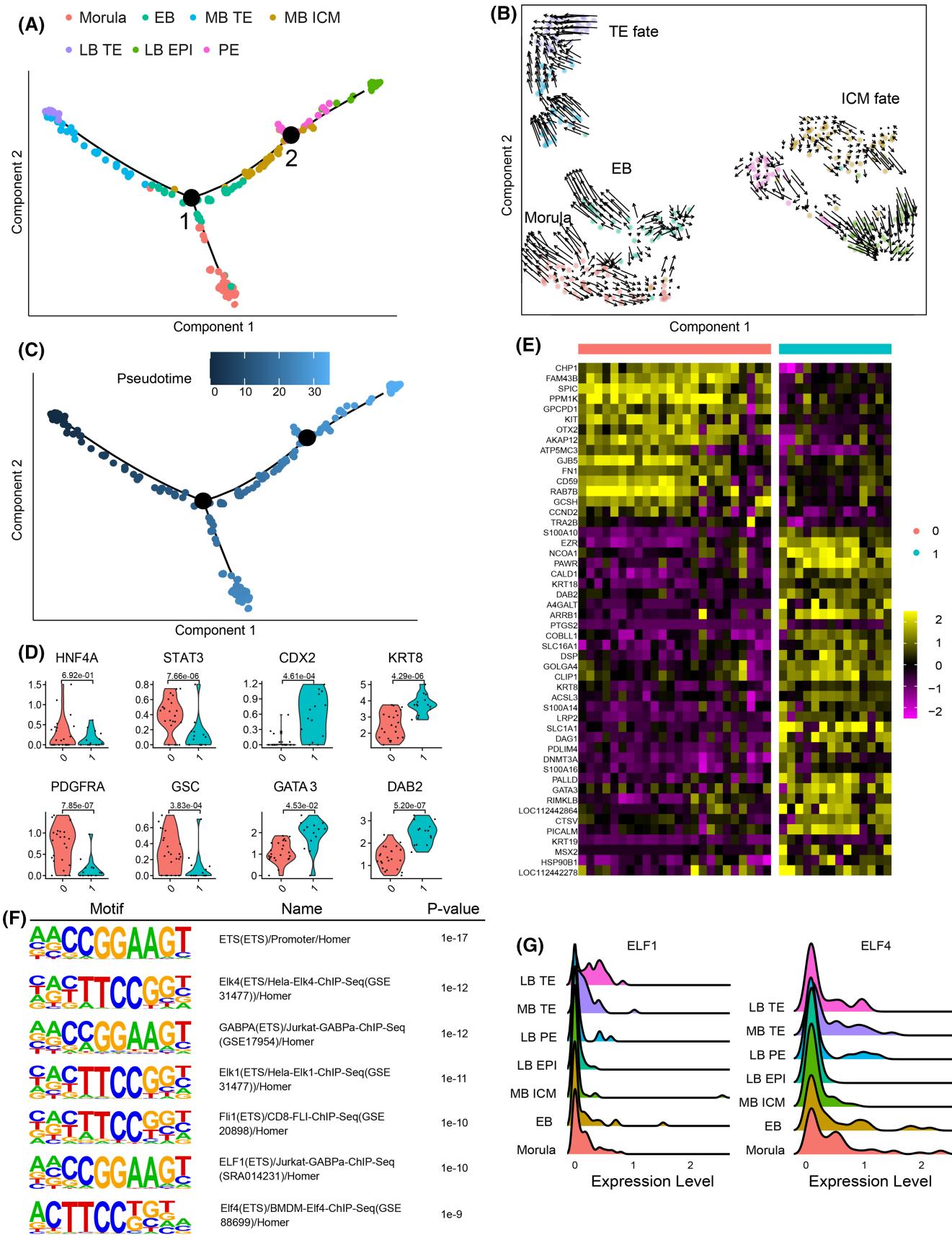
We next examined the pseudotime value to verify the timing of TE and ICM fate decisions, which represents the level of transcriptomic change from the root cell. The closer proximity of TE cells to the origin of trajectory than ICM cells indicated an earlier decision for TE fate than ICM fate (**Figure 2C**), aligning with observations in mouse embryos.<sup>28</sup> To investigate whether TE and ICM fates are asynchronous, we used the UMAP approach to divide EB cells into two clusters. Results show that Cluster 0 had higher expression of ICM markers while Cluster 1 had higher expression of TE markers (**Figure 2D**). We then performed differential expression analysis on these two clusters and selected the top 50 differentially expressed genes, 35 of which were highly expressed in cluster 1 (**Figure 2E**), which also indicated that TE fate emerges earlier than ICM at EB stage.

The aforementioned results suggest that the first lineage segregation occurs at the EB stage. Therefore, we conducted a motif analysis using 278 genes with relatively high expression levels ( $|Foldchange| > 1$ ) at the EB stage compared to other stages to identify genes that regulate initial lineage segregation (**Figure 2F**). Among these results, we discovered ETS family member ELK1, an essential transcription factor ELK1 in human embryonic stem cells.<sup>30</sup> There are also other ETS family members including ELF1 and ELF4 exhibiting higher expression levels in TE cells compared to ICM cells (**Figure 2G**), suggesting they may regulate the formation of TE lineage. Our findings suggest that the segregation of ICM and TE occurs post-cavitation at the EB stage, with TE fate specified earlier.

### 3.3 | Emergence and establishment of pluripotency in bovine embryos

To investigate the emergence of pluripotent cells, we examined cells at the EB, MB ICM, and LB EPI stages

**FIGURE 2** TE and ICM begin to separate after blastocyst formation with TE fate specified earlier. (A) Pseudotime analysis reveals that the peri-cavitating embryos are arranged into a major trajectory with two bifurcations. (B) RNA velocity vectors projected from the UMAP analysis. (C) Pseudotime projection allows for the visualization of all cell clusters. (D) Two clusters of EB cells can be distinguished by the expression level of ICM and TE markers. (E) The top 50 differentially expressed genes between two clusters of EB cells are displayed in a heatmap. (F) Genes with relatively high expression levels in the EB stage are used to identify known motifs. (G) ELF1 and ELF4 expression levels are displayed in a ridge plot.



(Figure 3A). We referenced pluripotency markers from other species due to limited studies on bovine embryos.<sup>31,32</sup> We checked the expression pattern of naive and primed pluripotency genes. *POU5F1/OCT4* and *SOX2* expression remained consistent across all pluripotent stages (Figure 3B), while *NANOG* expression was first observed in MB ICM. Naive pluripotency markers such as *KLF4*, *KLF5*, *ESRRB*, and *TBX3* were detected in EB cells and gradually decreased in MB ICM and LB EPI. *DPPA3*, *MAEL*, *DUSP14*, and *STAT3* showed high expression level in both EB and MB ICM cells (Figure 3C). Primed pluripotency markers *SOX11*, *NANOG*, *PRDM14*, *DNMT3A*, *DNMT3B*, *MEIS2*, and *DUSP6* were upregulated in LB EPI, while *TET1* and *OTX2* were downregulated in LB EPI. *SALL2* and *FGF2* expression were only detected in LB EPI (Figure 3D).

We used K-means clustering to group genes based on their expression profiles. Genes with high expression levels in EB cells (cluster 3, 4) include naive pluripotency markers. Genes upregulated in MB ICM and LB EPI (clusters 12, 13, 16) include primed pluripotency markers, DNA methyltransferases, and genes indicative of glycolytic metabolism (Figure 3E).

We aimed to identify novel pluripotency markers in bovine embryos using our dataset. In cluster 13, we noticed surface markers including *CD9*, *CD24*, *CD46*, *CD59*, *CD63*, *CD 177*, and *CD 200* (Figure 3E). Surface markers have been reported in human naive and primed pluripotent cells.<sup>33</sup> To verify whether CDs can serve as pluripotency markers in bovine embryos, we checked the expression level of all cells, including pluripotent cells and TE cells. *CD200* was specifically expressed in LB EPI cells and could be a primed pluripotency cell surface marker in bovine embryos (Figure 3F).

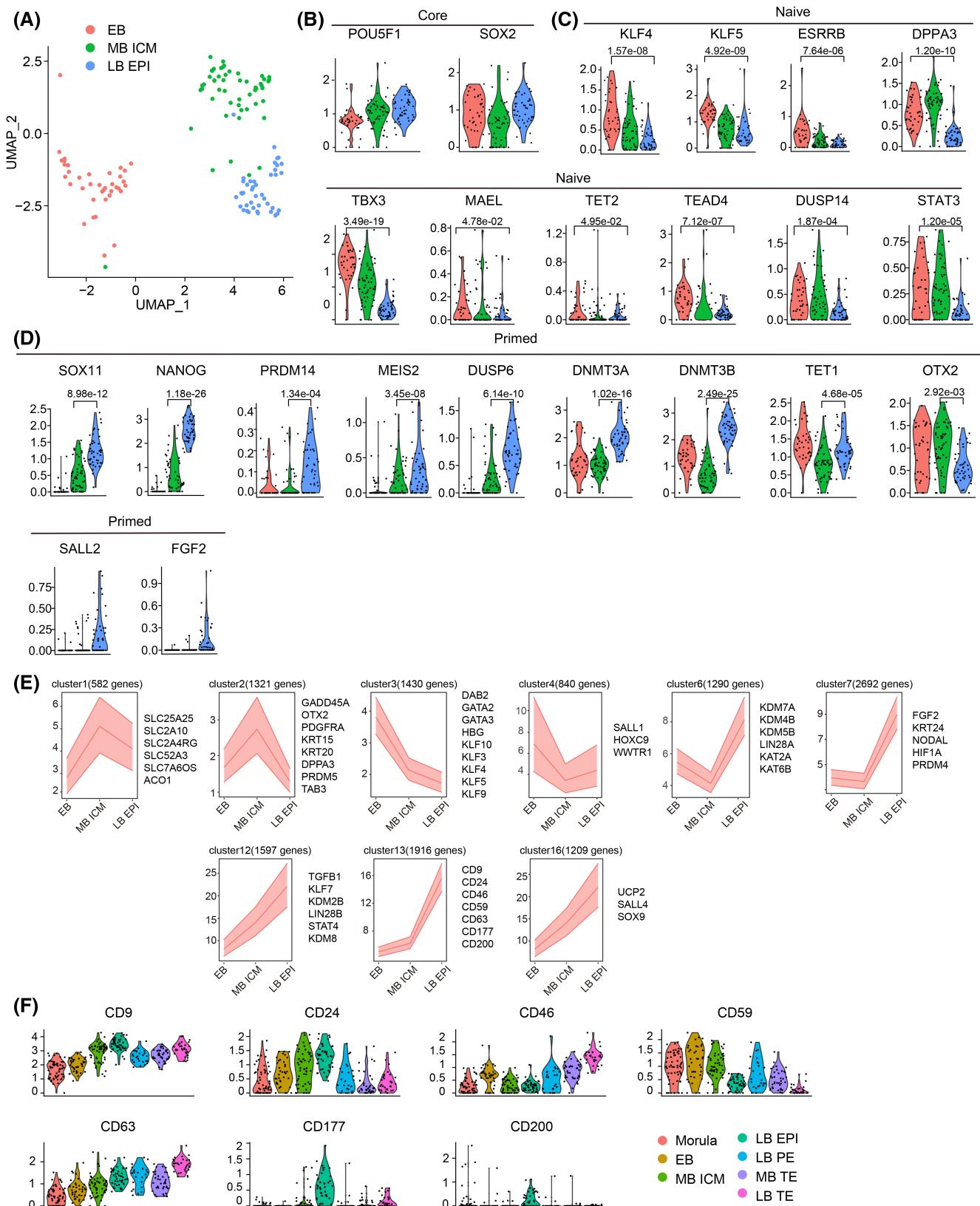
### 3.4 | Glycolytic metabolism is upregulated along with the first cell fate decision

Gene Ontology and KEGG enrichment analysis between these stages revealed that the transition from naive to primed pluripotency is associated with metabolism (Figure S3; Data S1). Glycolysis and oxidative phosphorylation (OxPhos) play a crucial role in balancing biomass generation with energy production in early embryos (Figure 4A). Consequently, we analyzed enzymes and regulatory genes involved in glycolysis and OxPhos by calculating module scores (Figure 4B). The expression of OxPhos-associated genes remained stable from the morula to the middle blastocyst stage and then slightly decreased at the late blastocyst stage. Meanwhile, the

expression of glycolysis-related genes steadily increased. We also observed no significant differences between lineages at the same developmental stage, indicating that glycolysis and OxPhos metabolic processes are not associated with cell lineages (Figure 4B). To validate these findings, we examined gene expression levels in individual cells, consistently reflecting OxPhos and glycolysis module scores (Figure 4C). Notably, *ESRRB* that stimulate OxPhos during naive pluripotency maintenance<sup>34</sup> was upregulated in EB but later downregulated in MB ICM and LB EPI (Figure 4D). Enzymes involved in the tricarboxylic acid (TCA) cycle and OxPhos, such as *IDH1*, *ACO2*, and *UQCRC2*, followed a similar trend. EPI-specific genes were enriched in numerous terms related to metabolism and oxygen dosage processes, including HIF-1 signaling pathways (Data S1). The mitochondrial electron transport chain (ETC) comprises complexes I-IV, as well as electron transporters ubiquinone and cytochrome c (Figure 4A). Our results demonstrated that the expression of electron transfer complex IV (cytochrome c oxidase) genes (10/12 genes) slightly decreased during epiblast maturation (Figure 4E), suggesting a reduction in mitochondrial metabolism.<sup>35</sup> In summary, there is a gradual transition of pluripotency from naive to primed state along with a dynamic metabolic change from EB to LB stage in bovine embryos.

### 3.5 | SOX2 knockout disrupts the first cell fate decision

We sought to unravel the molecular circuitry governing pluripotency with a focus on the transcription factor *SOX2*, which is the most notable marker restricted into ICM cells in bovine and human embryos as determined before.<sup>12,36</sup> Its knockout or dysregulation leads to defective differentiation, especially in the second cell fate decision.<sup>12,37</sup> However, it remains unknown if *SOX2* affects the first lineage segregation event in non-rodent models. scRNA-seq was performed in *SOX2* knockout embryos, which was achieved by creating premature stop codon using the cytosine base editing tool (BE3; Figure 5A and Figure S4A) as demonstrated before.<sup>12</sup> Immunofluorescence analysis shows that *SOX2* is absent in KO morula (Figure 5B). The sequencing reads spanning *SOX2* in all detected cells exhibit predicted editing, further confirming the successful *SOX2* KO in the embryo (Figure S4B). After quality control, 38 and 43 cells collected from 4 WT and 4 KO morula, respectively, were analyzed. In UMAP analysis, WT and *SOX2* KO cells formed two distinguishable clusters, suggesting a global difference in transcriptome (Figure 5D). There are 412 differentially expressed genes, among



**FIGURE 3** Progression of pluripotency in bovine blastocysts. (A) UMAP plot with EB, MB ICM, and LB EPI clusters. Colors indicate developmental stages. (B) Violin plots of the expression of core pluripotency genes SOX2 and POU5F1/OCT4. (C) Violin plots of the expression of selected naive pluripotency marker genes. (D) Violin plots of the expression of selected primed pluripotency marker genes. (E) K-means clustering to group genes based on their expression profiles. (F) Violin plots of expression of surface markers in pluripotent lineages.

which 241 were significantly downregulated in the SOX2 KO group (Figure S4D).

In the SOX2 KO morula, the transcript levels of pluripotency genes including *FN1*, *OTX2*, *PDGFRA*, and *SPIC* are significantly reduced, whereas *KRT18*, *GATA3*, and *KRT8*, considered as TE markers, increased (Figure 5C). Other genes expressed in TE-like *CDX2* and *TEAD4* showed no obvious difference between genotypes (Figure 5C). Furthermore, WT cells exhibit differentiation tendencies toward ICM (pre-ICM) or TE (pre-TE) and can be divided into two clusters by Seurat, which shows different expression of markers mentioned above (Figure S4C). This difference is not caused by the heterogeneity between embryos because of the randomly distribution of pre-ICM and pre-TE cells within each embryo (Figure S4C, left). We then found no significant difference about the expression of TE marker between pre-TE and SOX2 KO cells but significant difference about the expression of ICM marker between pre-ICM and SOX2 KO cells (Figure 5E). All of these results suggest SOX2 is required for the initiation of the first cell fate decision program in bovine embryos.

GO and KEGG enrichment analysis revealed that multiple metabolism pathways were significantly disrupted in SOX2 KO cells (Data S2). These results indicated a downregulation of OxPhos metabolism with aberrant upregulation of carbon metabolism, biosynthesis of amino acids, cysteine and methionine metabolism, glycolysis/gluconeogenesis, and PPAR signaling pathway (Data S2), which are involved in biomass generation and energy metabolism.

Previous research suggests that the shift from bivalent respiration to glycolytic metabolism and intact mitochondria function during peri-implantation development is critical for mammalian early development.<sup>38,39</sup> In our quest to understand the metabolic disorder in SOX2 KO cells, we compared the differentially expressed genes between SOX2 KO/WT and pre-ICM/pre-TE. We noted mitochondrial genes were significantly downregulated in the SOX2 KO cells (Figure 5F, left). This differential expression was not observed between WT cells, suggesting that the inhibition of this gene is not due to lineage variation (Figure 5F, right). Interestingly, when we conducted a PCA with glycolysis genes, the WT and SOX2 KO cells segregated and the majority of glycolytic enzymes were found to be upregulated in the SOX2 KO cells (Figure 5G,H). These findings suggest that the developmental response to SOX2 depletion-induced stress observed in this study is primarily driven by metabolic changes.

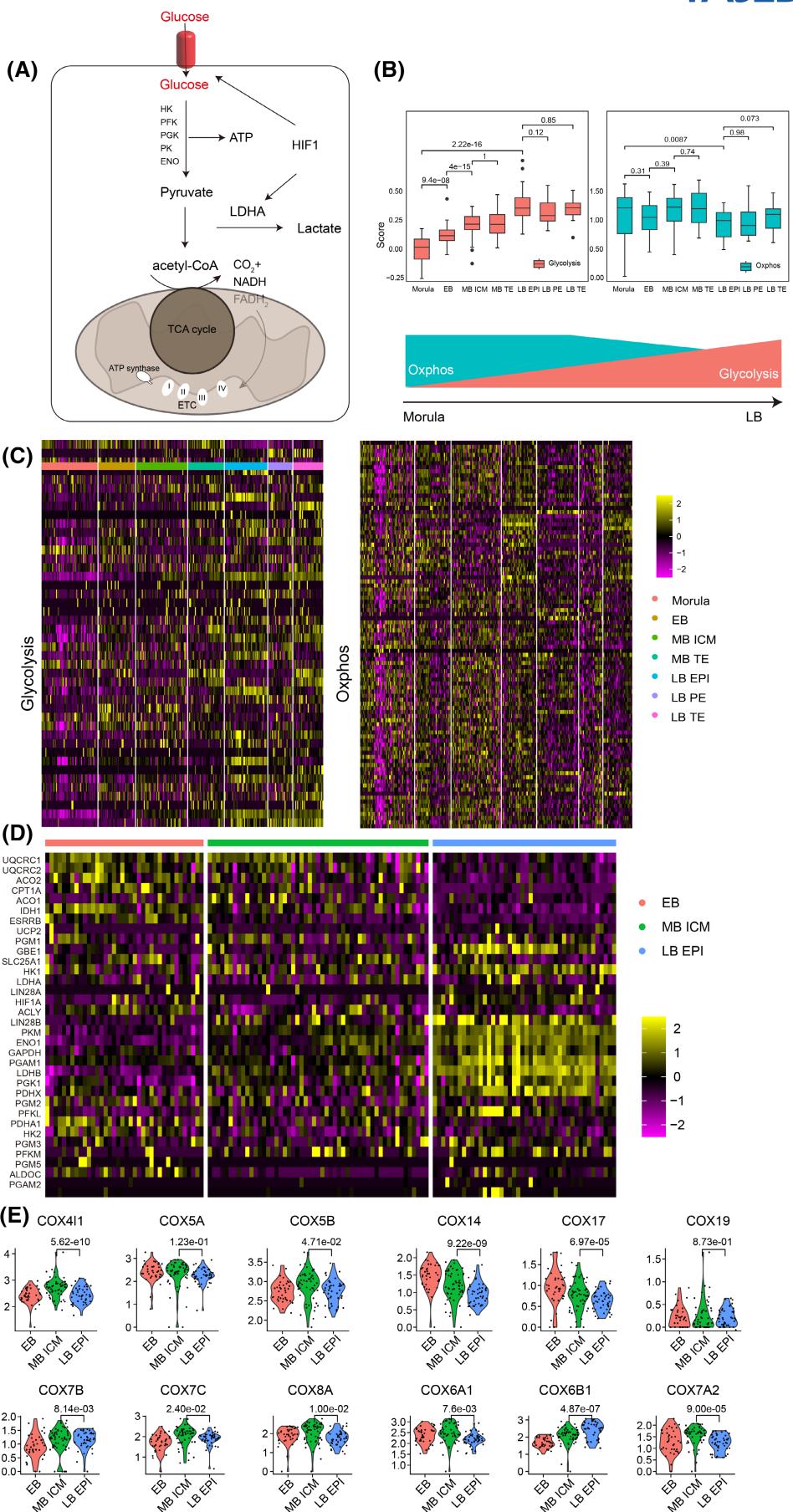
### 3.6 | SOX2 knockout cells acquire TE-like identity but deviate from normal TE

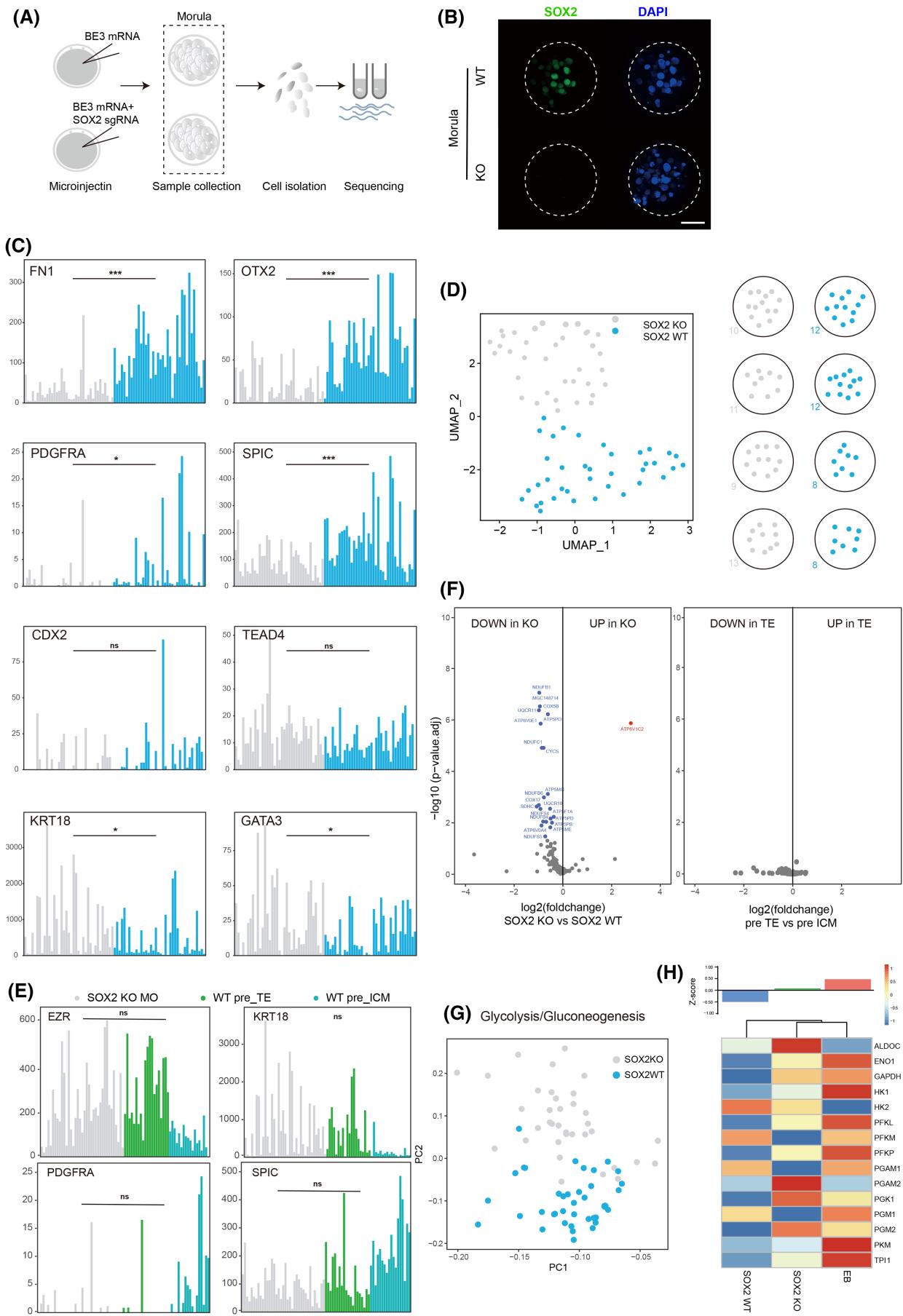
To further investigate the impact of SOX2 KO on ICM lineage specification, we collected single cells from early-stage blastocysts, which included 35 cells from 3 SOX2 KO embryos and 44 cells from 6 WT embryos (Figure 6A,B). We then performed UMAP analysis to reveal distinct clustering patterns (Figure 6C, Figure S5A). To further verify successful SOX2 KO, we examined all reads spanning the editing sites of SOX2 within each cell, showing 100% predicted editing (Figure S5B). WT cells distinctly separated from SOX2 KO cells at the blastocyst stage, with SOX2 KO cells forming a unique cluster, while WT cells divided into TE/ICM clusters based on the previously mentioned marker expression (Figure 6C, Figure S5A–C). These observations suggest that the initial specification of ICM lineage is impaired following SOX2 KO.

Through the combination of pseudotime and diffusion component analysis,<sup>25</sup> we found that SOX2 loss drives cells toward TE (Figure 6D). To determine the similarity between TE and SOX2 KO cells, we utilized MuSiC, a method for resolving the relative proportion of distinct cells by deconvoluting heterogeneous populations, to compute the fraction of identity.<sup>24</sup> Remarkably, the similarity between TE and SOX2 KO cells reached ~60%, significantly higher than the ~40% observed between TE and ICM (Figure 6E). To confirm these findings, we selected the top 100 highly expressed genes in TE (Figure 6F) and ICM (Figure S5D) separately and examined their expression in SOX2 KO cells. The results revealed that SOX2 KO cells expressed TE markers, such as *KRT8*, *KRT18*, *GATA3*, and *DAB2* (Figure 6F), while several ICM markers, including *OTX2*, *SPIC*, and *PDGFRA*, were downregulated (Figure S5D). Notably, some highly expressed ICM genes remained expressed in SOX2 KO cells, possibly due to activation by pathways independent of SOX2.

The aforementioned findings suggest that following SOX2 KO, cellular identity becomes more similar to TE, but still diverges from normal TE (Figure S5D). However, we still observed downregulation of some genes expected to be expressed in MB and LB TE after SOX2 KO (Data S3), such as *CLDN4*, which is crucial for the formation of tight junctions between TE cells,<sup>40</sup> and *IL6* and *PLAC8*, which are associated with implantation potential through promoting trophoblast invasion.<sup>41,42</sup> These results indicate that SOX2 is not necessary for the initial specification of TE but is required for normal TE commitment in bovine embryos.

**FIGURE 4** Metabolic dynamics during progression of pluripotency. (A) Glycolysis and OxPhos metabolism. ETC, electron transport chain; TCA, tricarboxylic acid cycle. (B) Boxplot of scores of glycolysis and OxPhos modules in cell clusters. (B,C) Heat map display genes involved in OxPhos and glycolysis in individual cells. (D) Heat map and violin plots display selected genes involved in OxPhos and glycolysis in pluripotent lineages. (E) Expression of electron transfer complex IV (cytochrome c oxidase) genes.





**FIGURE 5** SOX2 knockout disrupts the first cell fate decision. (A) Experimental scheme to explore effects of SOX2 knockout at morula stage. (B) Immunofluorescent detection of SOX2 signal at morula stage. Scale bar: 50  $\mu$ m. (C) Box plots display gene expression levels of selected ICM and TE markers (\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ ; NS: no significance). (D,G) Left: UMAP plot shows clusters of WT and SOX2 KO cells. Right: Collection of SOX2 knockout embryos for single-cell RNA-seq sequencing; blue: wildtype (WT) cells; gray: SOX2 knockout (KO) cells. (E) Box plot showed lineage marker genes expression of wildtype and SOX2 knockout cells. (F) Volcano plot shows gene expression levels in cells ( $p < .05$ , log2 fold change  $> 1$ , key genes are annotated). (G,H) PCA plot and expression levels of glycolysis genes in WT and SOX2 KO cells.

We next conducted KEGG enrichment analysis on genes downregulated in SOX2 KO compared to ICM or TE. Strikingly, enrichment of the OxPhos was observed in downregulated genes between SOX2 KO and WT ICM, but not WT TE, suggesting that SOX2 regulation of OxPhos might be specific to ICM. Additionally, we did not observe upregulation of OxPhos in ICM compared to TE, indicating that the results are indeed a consequence of SOX2 KO rather than lineage differences (Figure 6G, Data S3). This result suggests that energy metabolism differs among distinct cells and that OxPhos may be involved in the regulatory network of SOX2 to ICM lineage specification.

### 3.7 | The second lineage segregation occurs at middle blastocyst stage

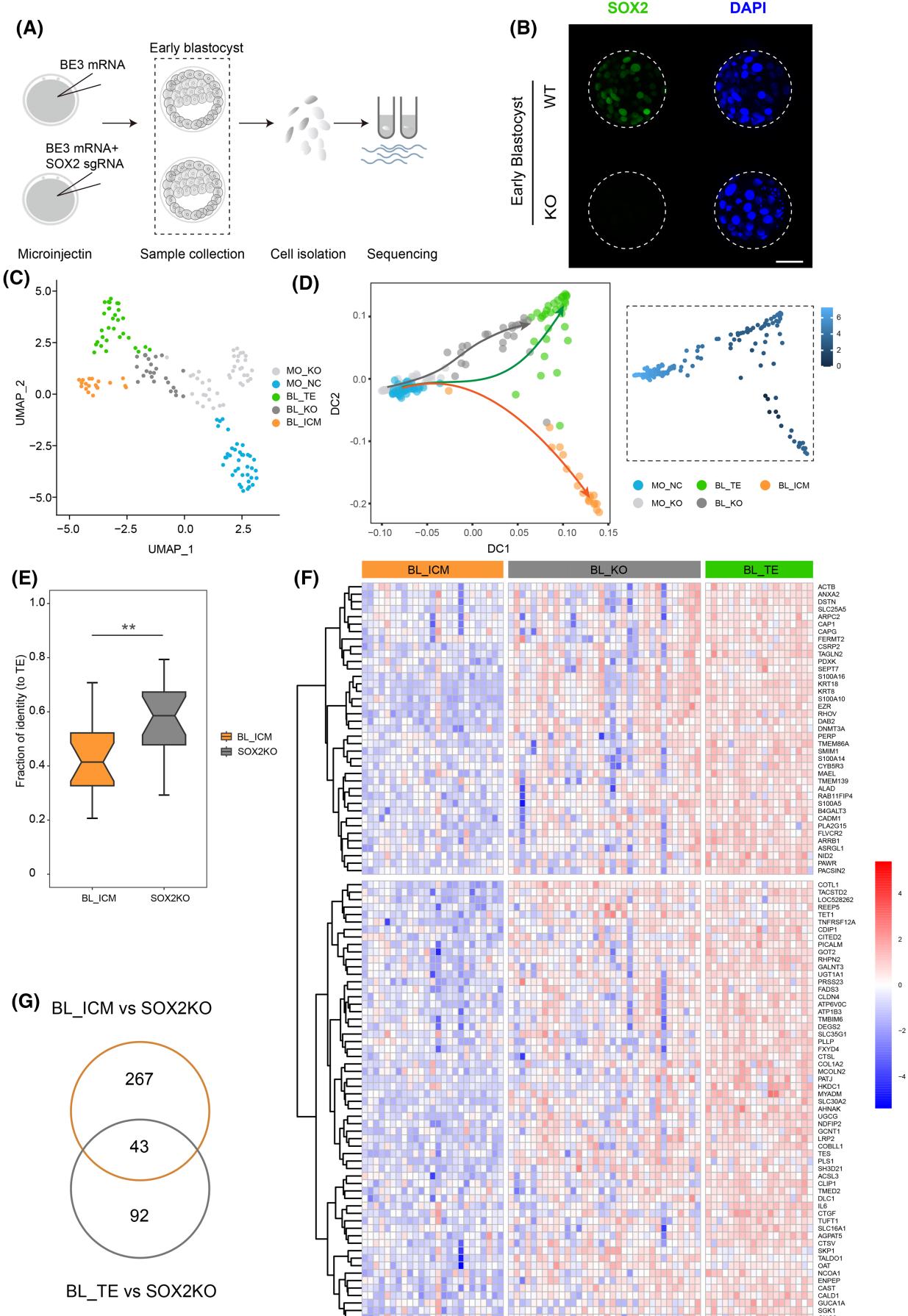
Pseudotime analysis highlighted the bifurcation point where cells make their second fate decision, identified as MB ICM and PE cells. These cells occupy the second branch point and part of the EPI bifurcation (Figure 2A). Consequently, we examined the expression of known EPI markers (*NANOG*, *TDGF1*, *PRDM14*) and the PE marker (*GATA4*) along the developmental paths (Figure S2B). We employed RNA velocity, a method of assessing non-spliced mRNA to infer cell fate,<sup>43</sup> in order to validate the timing of second cell fate decision. The RNA velocity analysis indicated that cells had a transcriptional trending velocity, with some PE cells showing a shift toward EPI gene expression (Figure 2B). This could be attributed to the fact that both lineages differentiate from MB ICM cells, and the PE lineage emerges from this stage when embryos are collected at E9, resulting in few differentially expressed genes between ICM cells and PE cells (Figure S6B). We assessed the expression level of *PDGFRA*, *GATA4*, *GATA6*, *COL4A1*, *SPARC*, and *RSPO* in bovine embryos, which are recognized markers in mouse and human PE cells.<sup>26</sup> As expected, we observed abundant expression of these genes in PE (Figure S6C), suggesting the conservation of these PE-associated genes across human, bovine, and mouse embryos. GO and KEGG analysis revealed that PE-specific genes are enriched in the PI3K-AKT signaling pathway (Figure S3; Data S1), which is known to stimulate other

pathways in mouse and porcine embryos,<sup>44</sup> such as PI3K-AKT and MAPK.

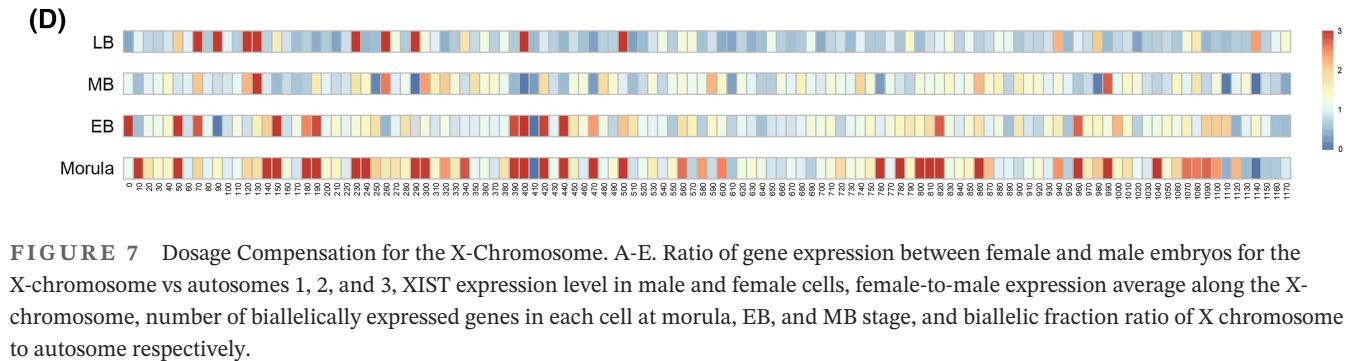
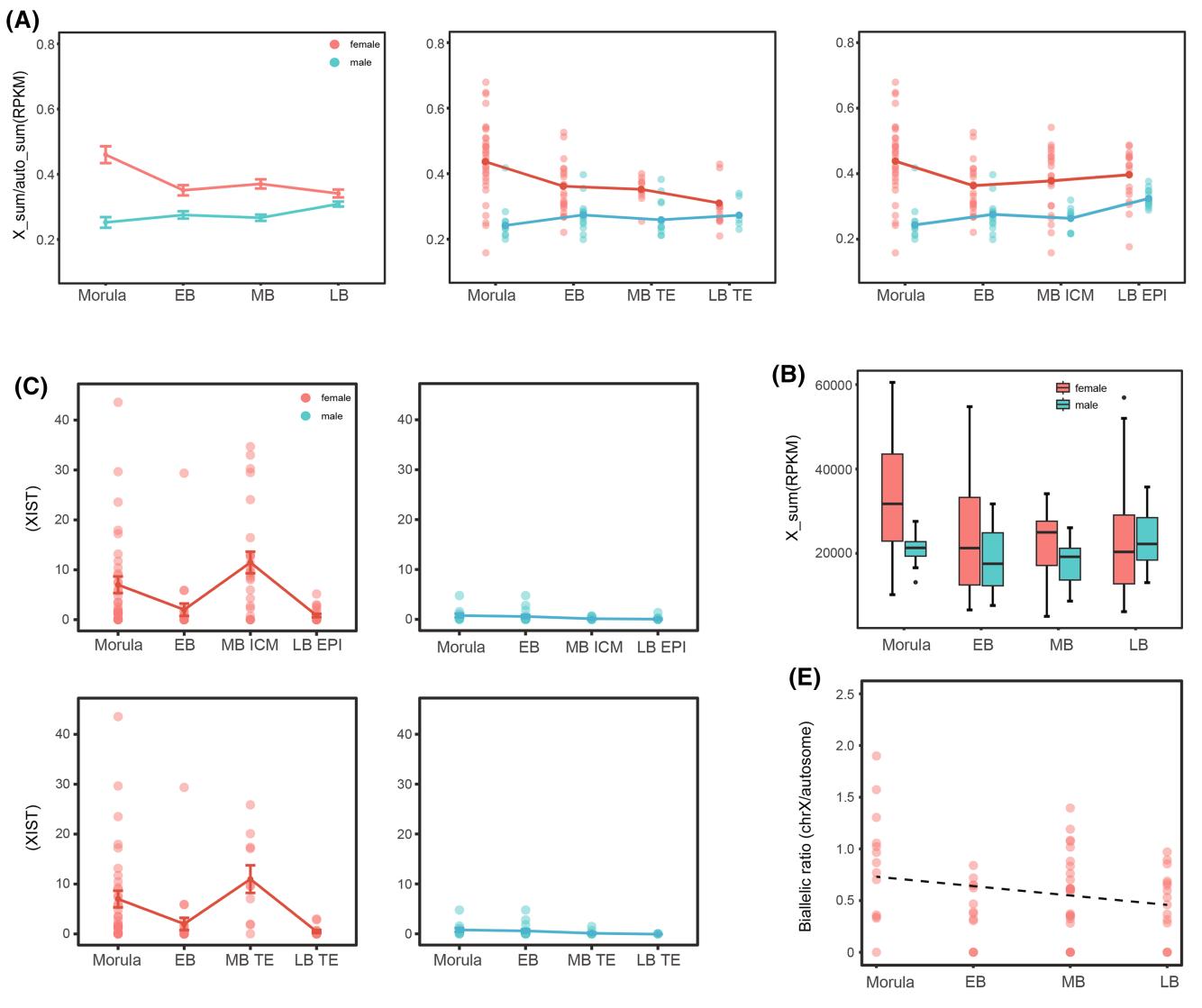
### 3.8 | Dosage compensation of X chromosome in bovine embryos

The single-cell data offer a unique opportunity to investigate the dynamic pattern of X chromosome expression in bovine preimplantation embryos. To determine the gender of each cell, we calculate the cumulative expression and number of transcribed genes on the Y chromosome (Figure S7A,B). To discern the differences in X chromosome expression between male and female cells, we analyze the gene expression ratio of the X chromosome relative to autosomes within cells at each developmental stage and cell lineage. Our findings indicate that X-linked genes are progressively equalizing between females and males over time in both ICM and TE lineages, but interestingly, the process of dosage compensation appears to be more pronounced in the TE lineage (Figure 7A). We proceeded to assess the significance between stages in male and female embryos. Results revealed a significant decrease in the total expression of the X chromosome from morula to the EB stage in female embryos. However, the difference between the EB and MB stages was not found to be significant in female embryos. While no significant differences were observed between the MB and LB stages overall, there was a noteworthy increase in the expression of the X chromosome from MB ICM to LB EPI in female embryos. Additionally, no significant differences were detected between stages in male embryos. In summary, the total expression of the X chromosome decreases in females while remaining constant in males from the morula to LB stage, further corroborating the occurrence of dosage compensation (Figure 7B and Figure S7C).

In alignment with X chromosome dosage compensation, the expression of *XIST* is upregulated during development and is significantly higher in most female cells. However, only sporadic expression of *XIST* is detected in some male cells (Figure 7C). The late blastocyst stage was excluded due to low *XIST* expression across all cells compared to other stages, suggesting a predominantly male composition of embryos (Figure S7D).



**FIGURE 6** SOX2 knockout cells acquire TE-like identity but deviate from normal TE. (A) Experimental design to explore effects of SOX2 knockout at early blastocyst stage. (B) Immunofluorescence detection of SOX2 signal at morula stage. Scale bar: 50  $\mu$ m. (C) UMAP plot of early blastocyst cells. Colors represent different genotypes and lineages. (D) Left: Diffusion plot for WT/SOX2 KO cells at morula and early blastocyst stage. Colors represent different genotypes and lineages. Right: Diffusion map and pseudotime expression. MO\_NC, wildtype cells of morula stage; MO\_KO, knockout cells of morula stage; BL\_ICM, wildtype ICM cells of blastocyst stage; BL\_TE, wildtype TE cells of blastocyst stage; BL\_KO, knockout cells of blastocyst stage. (E) Fraction of identity between BL ICM/SOX2 KO and TE at early blastocyst stage using all expressed genes (RPKM >1, \*\* $p$  <.01). (F) Heatmap of TE-high expressed genes in all cells. (G) Numbers of differentially expressed genes between BL ICM/SOX2 KO and BL TE/SOX2 KO.



To explore regional differences in the reduction of female X chromosome expression, we calculated the ratio of female to male expression along the chromosome with equal intervals. This revealed a progressive dosage compensation mechanism occurring across the entire X chromosome, with certain areas maintaining high ratios of expression at the late blastocyst stage (Figure 7D).

To determine whether dosage compensation is accompanied by XCI, we sought to analyze X chromosome expression at an allelic resolution. We were able to identify single nucleotide variants (SNV) directly from the RNA-seq reads and filtered them using previously suggested criteria for each cell.<sup>4</sup> Notably, in female cells, the biallelic fraction ratio of the X chromosome to autosome shows tendency of downregulation with development (Figure 7D,E). The frequency of biallelically expressed X-linked genes in males was noticeably lower than in females, confirming the accuracy of our allelic expression analysis (Figure S7D). Further observation of SNPs within individual genes revealed that XIST is mostly monoallelically expressed in female cells (with only ~10% biallelic expression detected), consistent with the small proportions of two XIST spots in a single nucleus.<sup>45</sup> Overall, these results suggest that dosage compensation of X-linked genes initiates from the morula stage accompanied by the partial inactivation of the X chromosome and the progress of X chromosome dosage compensation is more extensive within TE lineage.

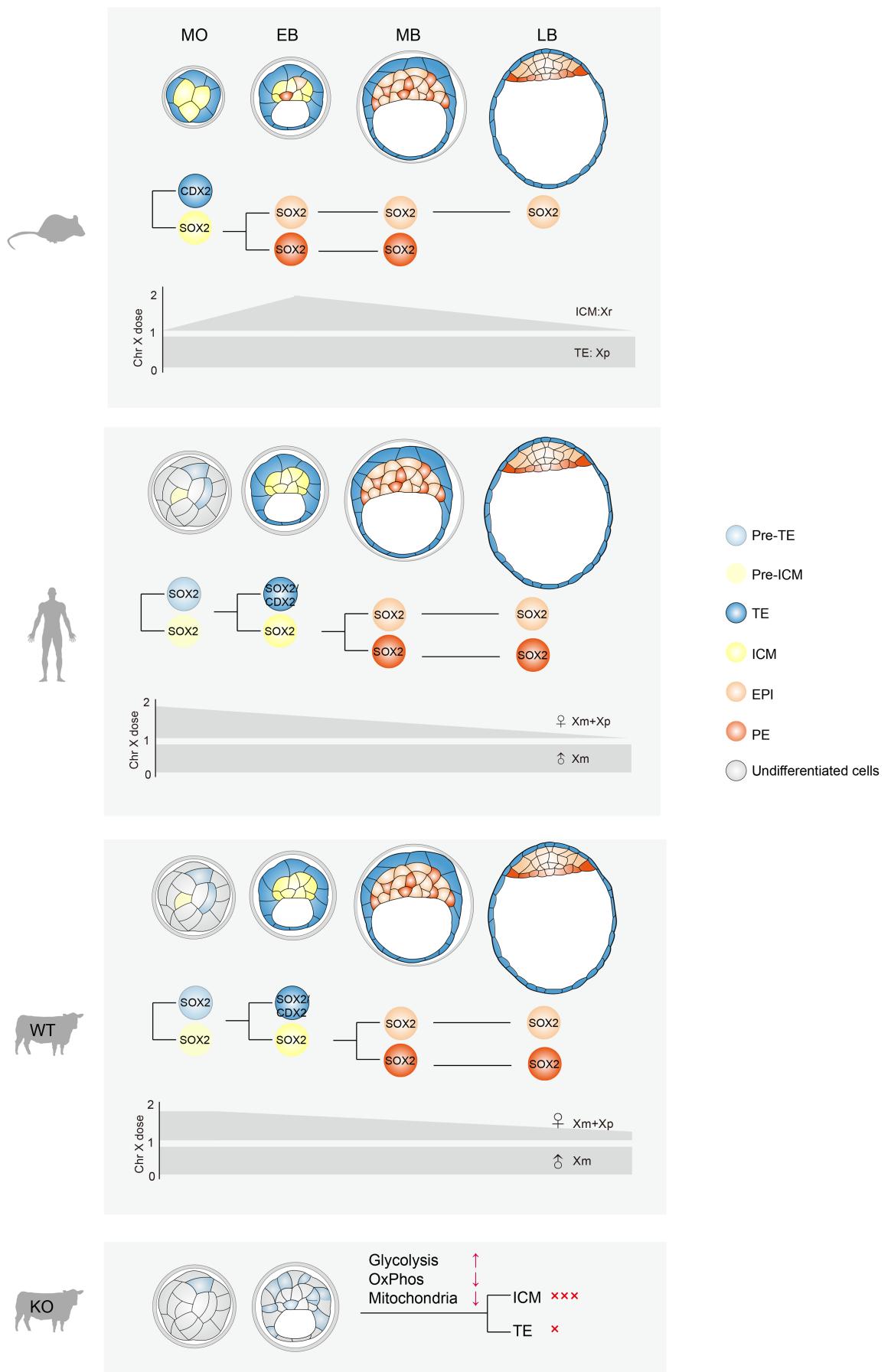
## 4 | DISCUSSION

In this study, we obtained the transcriptome atlas of bovine peri-cavitating embryos through single-cell transcriptomics and determined the precise timing of the first lineage segregation (TE/ICM) in bovine embryos. Furthermore, by combining base editing and single-cell transcriptomic technology, we demonstrated that SOX2 regulates the first cell fate decision program. In particular, SOX2 is required for the initiation of ICM fate and TE commitment, likely mediated through a metabolic mechanism. In addition, we revealed that X chromosome dosage compensation in bovine embryos began at the morula stage, accompanied by partial XCI. This study also compared the consistency and differences in lineage segregation, pluripotency establishment, and the role of SOX2 and X chromosome

dosage compensation among bovine, mouse, and human embryos (Figure 8).

Lineage segregation is a pivotal event during mammalian preimplantation development with specific regulation differing across species, especially the first lineage segregation.<sup>46</sup> In bovine embryos, differentiation within the cell population commences as early as the morula stage, yet TE/ICM lineage segregation is only completed in the early blastocyst stage following blastocyst formation. Conversely, ICM and TE fate segregation transpires during the late morula stage in mouse embryos,<sup>28</sup> while the initial lineage segregation occurs after blastocyst formation in human embryos.<sup>29</sup> This discrepancy suggests that the first cell decision is independent of blastocyst cavity formation. Concurrently, this observation aligns with the fact that CDX2 is detectable in outside cells until the early blastocyst in humans<sup>36</sup> and cattle,<sup>12</sup> whereas it is clearly restricted to the outside cell in morula embryos in mice.<sup>28</sup> Intriguingly, TE fate is determined earlier than ICM, a feature conserved among bovine, mouse, and human embryos. In accordance, TE markers (CDX2 and GATA3) are confined to TE, while pluripotency marker SOX2 is limited to ICM cells in late blastocysts in cattle.<sup>12,36</sup> It is hypothesized that TE may be necessitated earlier in placental animals, potentially to establish a suitable environment for successful embryonic development. The outer trophectoderm in mice undergoes commitment by the middle blastocyst stage.<sup>28</sup> In contrast, cells of the trophectoderm in late-expanded cattle blastocysts are predetermined to maintain their trophectodermal identity, but they have not yet reached a state of definitive commitment to this fate.<sup>47</sup> We observed a clear visualization of MB TE and LB TE segregation in UMAP plot and cell trajectories (Figures 1B and 2A), suggesting the existence of different TE populations in the preimplantation bovine embryos. The species-specific variation in TE lineage may be attributed to the distinction that implantation takes place in mouse embryos post-blastocyst hatching, whereas bovine embryos experience a period of elongation following blastocyst hatching.<sup>48</sup> Moreover, we identified there is indeed a second lineage differentiation in bovine embryos by cell trajectory analysis. Regretfully, we did not determine the precise mechanism of the second lineage segregation. This could be attributed to the fact

**FIGURE 8** Model of the first cell fate decision program among mouse, human and bovine embryos and SOX2's functional significance in bovine embryos. In mice, CDX2 and SOX2 are restricted into trophectoderm and inside cells (progenitor of ICM), respectively, during the first cell fate decision in late morula. In contrast, the first cell fate decision occurs in early blastocysts in human and bovine embryos, with TE being specified first as shown by restricted CDX2 expression while SOX2 is only exclusively expressed in ICM cells at late blastocyst stage. In the absence of SOX2, bovine embryos fail to make the first cell fate decision, resulting in no ICM fate and defective TE fate. This aberrant developmental outcome could be attributed to the disruption of metabolic pathways (glycolysis and oxidative phosphorylation) and mitochondria function upon SOX2 deletion.



that the second lineage segregation has been completed until the preimplantation stage in bovine embryo. Remarkably, we found that MB collected on the same day as LB are more similar to MB collected on earlier days rather than LB. It is noteworthy that middle and late blastocysts could emerge simultaneously on embryonic Day 9, prompting us to collect both MB and LB cells on the same day. Our dimensionality reduction analysis indicated that cells from MB collected on Days 8 and 9 cluster together, suggesting that time post-fertilization and morphological states jointly determine the developmental changes.

The derivation of embryonic stem cells (ESCs), particularly naive state ESCs, has proven difficult in cattle, partly due to limited knowledge regarding pluripotency in bovine early development. Our study uncovers the successive progression of pluripotency in bovine blastocysts. Specifically, we demonstrate a transition from a naive state to a primed state during the development from early blastocyst cells to late blastocysts. We discovered that the expression dynamics of naive pluripotency markers (*KLF4*, *KLF5*, and *TBX3*) in bovine blastocysts follow a similar pattern to those observed in humans and monkeys,<sup>27,49,50</sup> while *Klf2*, *Prdm14*, and *Tfacp2l1* serve as naive pluripotency markers in mouse embryos.<sup>51</sup> These genes regulate gene expression and cell signaling, emphasizing potential functional differences in naive pluripotent cells between mice and other mammals. Furthermore, we identified CD200 as a potential primed pluripotency cell surface marker in bovine embryos (Figure 3F). Correspondingly, surface markers have been reported in human naive and primed pluripotent cells.<sup>33</sup> These findings could aid in the development of an efficient protocol for deriving ESCs in bovine species, potentially transforming breeding strategies in domestic animal production.

SOX2 is the earliest pluripotency marker restricted to inside cells (progenitor of ICM) at the morula stage in mice.<sup>36,37</sup> In contrast, SOX2 is restricted to ICM at the late blastocyst stage in both human<sup>36</sup> and bovine embryos,<sup>12</sup> suggesting differential regulation of ICM fate relative to mice. In mice, SOX2 is not required for the initial expression of OCT4 and NANOG, but is essential for PE development.<sup>37</sup> However, we recently discovered that SOX2 regulates NANOG and OCT4 expression in bovine blastocysts,<sup>12</sup> emphasizing the critical role of SOX2 in establishing and maintaining the EPI lineage in bovine embryos, although its precise function in the first cell fate decision remains enigmatic. In our study, we also observed significant downregulation of pluripotency-related genes and precocious expression of TE markers in embryos following SOX2 KO. Meanwhile, we did not observe a trend toward TE/ICM differentiation within

SOX2 KO embryos, which appeared in WT embryos, from the morula to the blastocyst stage. We show that SOX2 KO not only damages the initiation of ICM fate but also disturbs TE fate in bovine embryos, which are not seen in mouse embryos.<sup>37</sup> Similar species-specific effects have also been observed for another pluripotency gene, OCT4, which shows similar expression dynamics to those of SOX2. OCT4-null human embryos display downregulation of genes associated with all three lineages (TE, EPI, and PE).<sup>52</sup> In both human and bovine embryos, OCT4 is required for NANOG expression in blastocysts.<sup>52,53</sup> Nevertheless, OCT4 KO mouse embryos exhibit ectopic TE gene expression and normal NANOG expression.<sup>54</sup> Thus, it warrants further investigation of the extent to which molecular mechanisms regulating embryogenesis differ among mammals.

Moreover, genes associated with OxPhos, which typically display a gradual upregulation during embryo development,<sup>38</sup> were significantly downregulated in cells lacking SOX2. Notably, we found no significant difference in the overall expression of genes associated with OxPhos metabolism between TE and ICM lineages. However, downregulated genes upon SOX2 KO, compared to ICM rather than TE, showed specific enrichment in OxPhos. This result indicates that SOX2-mediated regulation of OxPhos is exclusive to the ICM. Consequently, disrupted energy metabolism may underlie the failed lineage segregation observed in SOX2 KO embryos. Further research is needed to clarify how SOX2 contributes to the first lineage segregation through energy metabolic mechanism.

X chromosome inactivation and dosage compensation are conserved events during early mammalian embryonic development, but different species exhibit significant diversity in XCI regulation strategies. In mice, two rounds of XCI begin at the 4-cell stage with imprinted X chromosome (Xp) inactivation followed by X chromosome reactivation in ICM and random X chromosome inactivation during the epiblast stage. There was no whole-chromosome X chromosome inactivation observed at E7.0 in human embryos, where completed X chromosome dosage compensation was thought to be caused by X chromosome dampening, that is, simultaneous inactivation of both X chromosomes. Our data showed that X chromosome dosage compensation in bovine embryos began at the morula stage and was accompanied by partial XCI. However, some cells with two XIST spots (10%–20%) were detected in bovine embryos by other studies, indicating that X-linked gene co-suppression (“dampening”) may also occur in bovine embryos. In human embryonic development, such phenomenon has also been observed and occurs at a higher frequency of approximately 80%.<sup>55</sup> In addition, we found that the average expression level of X chromosome genes was higher in ICM cells than TE

cells among bovine embryos. Indeed, it has been reported that the number of XIST spots in NANOG-positive cells of bovine blastocysts is higher than that in CDX2-positive cells. These results suggest that XCI occurs more slowly in pluripotent cells.

In summary, this study defines the first cell fate decision program and X chromosome dosage compensation during bovine early embryonic development by using single-cell transcriptomics. Importantly, we show SOX2 is required for the first cell fate decision likely through a metabolic mechanism. We also compared the differences and similarities among bovine, mouse, and human embryos in these biological events, providing valuable resources for studying mammalian preimplantation development.

## AUTHOR CONTRIBUTIONS

Bingjie Hu and Kun Zhang conceived the project and designed research; Bingjie Hu, Hao Jin, Yan Shi, and Haotian Yu performed the research and acquired the data; Bingjie Hu and Hao Jin analyzed the data; Bingjie Hu, Hao Jin, and Kun Zhang wrote the manuscript; Xiaotong Wu and Shaohua Wang revised the manuscript.

## ACKNOWLEDGMENTS

We would like to express our gratitude to all members of the Zhang laboratory for their helpful discussions. This work was supported by the National Natural Science Foundation of China (Nos. 32161143032, and 32072731 to K.Z.), National Key R&D Program of China (2022YFE0100200 to K.Z.), Qinghai Science and Technology Major Program (2021-NK-A5 to K.Z.), and the Foundation of Key Laboratory of Veterinary Biotechnology (No. BD1500010 to K.Z.), Shanghai, P.R. China.

## DISCLOSURES

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

All data are available from the corresponding author. Single-cell RNA-seq data generated in this study have been deposited at Gene Expression Omnibus (GEO) database under accession number GSE239782.

## ORCID

Bingjie Hu  <https://orcid.org/0000-0001-7717-3063>  
Kun Zhang  <https://orcid.org/0000-0002-2324-9381>

## REFERENCES

- White MD, Zenker J, Bissiere S, Plachta N. Instructions for assembling the early mammalian embryo. *Dev Cell*. 2018;45:667-679.
- Rossant J. Genetic control of early cell lineages in the mammalian embryo. *Annu Rev Genet*. 2018;52:185-201.
- Chazaud C, Yamanaka Y. Lineage specification in the mouse preimplantation embryo. *Development*. 2016;143:1063-1074.
- Petropoulos S, Edsgard D, Reinius B, et al. Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell*. 2016;165:1012-1026.
- Singh M, Widmann TJ, Bansal V, et al. The selection arena in early human blastocysts resolves the pluripotent inner cell mass. 2019.
- Stirparo GG, Boroviak T, Guo G, Nichols J, Smith A, Bertone P. Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. *Development*. 2018;145:dev158501.
- Negrón-Perez VM, Zhang Y, Hansen PJ. Single-cell gene expression of the bovine blastocyst. *Reproduction*. 2017;154:627-644.
- Wei Q, Zhong L, Zhang S, et al. Bovine lineage specification revealed by single-cell gene expression analysis from zygote to blastocyst. *Biol Reprod*. 2017;97:5-17.
- Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science*. 2004;303:644-649.
- Mak W, Nesterova TB, de Napoles M, et al. Reactivation of the paternal X chromosome in early mouse embryos. *Science*. 2004;303:666-669.
- Dupont C, Gribnau J. Different flavors of X-chromosome inactivation in mammals. *Curr Opin Cell Biol*. 2013;25:314-321.
- Luo L, Shi Y, Wang H, et al. Base editing in bovine embryos reveals a species-specific role of SOX2 in regulation of pluripotency. *PLoS Genet*. 2022;18:e1010307.
- Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 2014;9:171-181.
- Ma H, Zhai J, Wan H, et al. In vitro culture of cynomolgus monkey embryos beyond early gastrulation. *Science*. 2019;366:eaax7890.
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12:357-360.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923-930.
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36:411-420.
- Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
- Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32:381-386.
- Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods*. 2017;14:309-315.
- Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14:979-982.
- Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576-589.

24. Wang XR, Park J, Susztak K, Zhang NR, Li MY. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun.* 2019;10:380.
25. Angerer P, Haghverdi L, Buttner M, Theis FI, Marr C, Buettner F. destiny: diffusion maps for large-scale single cell data in R. *Bioinformatics.* 2016;32:1241-1243.
26. Schröde N, Xenopoulos P, Piliszek A, Frankenberg S, Plusa B, Hadjantonakis AK. Anatomy of a blastocyst: cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. *Genesis.* 2013;51:219-233.
27. Nakamura T, Okamoto I, Sasaki K, et al. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature.* 2016;537:57-62.
28. Posfai E, Petropoulos S, de Barros FRO, et al. Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *elife.* 2017;6:e22906.
29. Meistermann D, Bruneau A, Loubersac S, et al. Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. *Cell Stem Cell.* 2021;28:1625-1640.e6.
30. Goke J, Chan YS, Yan J, Vingron M, Ng HH. Genome-wide kinase-chromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells. *Mol Cell.* 2013;50:844-855.
31. Boroviak T, Loos R, Bertone P, Smith A, Nichols J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol.* 2014;16:516-528.
32. Bogliotti YS, Wu J, Vilarino M, et al. Efficient derivation of stable primed pluripotent embryonic stem cells from bovine blastocysts. *Proc Natl Acad Sci USA.* 2018;115:2090-2095.
33. Collier AJ, Panula SP, Schell JP, et al. Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. *Cell Stem Cell.* 2017;20:874-890.e7.
34. Ware CB, Nelson AM, Mecham B, et al. Derivation of naive human embryonic stem cells. *Proc Natl Acad Sci USA.* 2014;111:4484-4489.
35. Zhou W, Choi M, Margineantu D, et al. HIF1 $\alpha$  induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* 2012;31:2103-2116.
36. Gerri C, McCarthy A, Alanis-Lobato G, et al. Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature.* 2020;587:443-447.
37. Wicklow E, Blij S, Frum T, et al. HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS Genet.* 2014;10:e1004618.
38. Malkowska A, Penfold C, Bergmann S, Boroviak TE. A hexa-species transcriptome atlas of mammalian embryogenesis delineates metabolic regulation across three different implantation modes. *Nat Commun.* 2022;13:3407.
39. Cheong A, Archambault D, Degani R, Iverson E, Tremblay KD, Mager J. Nuclear-encoded mitochondrial ribosomal proteins are required to initiate gastrulation. *Development.* 2020;147:dev188714.
40. Moriwaki K, Tsukita S, Furuse M. Tight junctions containing claudin 4 and 6 are essential for blastocyst formation in preimplantation mouse embryos. *Dev Biol.* 2007;312:509-522.
41. Li M, Liu D, Wang L, Wang W, Wang A, Yao Y. Expression of placenta-specific 8 in human oocytes, embryos, and models of in vitro implantation. *Fertil Steril.* 2016;106:781-789.e2.
42. Champion H, Innes BA, Robson SC, Lash GE, Bulmer JN. Effects of interleukin-6 on extravillous trophoblast invasion in early human pregnancy. *Mol Hum Reprod.* 2012;18:391-400.
43. La Manno G, Soldatov R, Zeisel A, et al. RNA velocity of single cells. *Nature.* 2018;560:494-498.
44. Ramos-Ibeas P, Sang F, Zhu Q, et al. Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. *Nat Commun.* 2019;10:500.
45. Yu B, van Tol HTA, Stout TAE, Roelen BAJ. Initiation of X chromosome inactivation during bovine embryo development. *Cell.* 2020;9:1016.
46. Gerri C, Menchero S, Mahadevaiah SK, Turner JMA, Niakan KK. Human embryogenesis: a comparative perspective. *Annu Rev Cell Dev Biol.* 2020;36:411-440.
47. Berg DK, Smith CS, Pearton DJ, et al. Trophectoderm lineage determination in cattle. *Dev Cell.* 2011;20:244-255.
48. Valadão L, da Silva HM, da Silva FM. Bovine embryonic development to implantation. In: Bin W, Huai LF, eds. Ch. 6, *Embryology*. IntechOpen; 2018.
49. Blakeley P, Fogarty NM, Del Valle I, et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development.* 2015;142:3613.
50. Boroviak T, Loos R, Lombard P, et al. Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. *Dev Cell.* 2015;35:366-382.
51. Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol.* 2016;17:155-169.
52. Fogarty NME, McCarthy A, Snijders KE, et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature.* 2017;550:67-73.
53. Simmet K, Zakhartchenko V, Philippou-Massier J, Blum H, Klymiuk N, Wolf E. OCT4/POU5F1 is required for NANOG expression in bovine blastocysts. *Proc Natl Acad Sci USA.* 2018;115:2770-2775.
54. Frum T, Halbisen MA, Wang C, Amiri H, Robson P, Ralston A. Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Dev Cell.* 2013;25:610-622.
55. van den Berg IM, Laven JSE, Stevens M, et al. X chromosome inactivation is initiated in human preimplantation embryos. *Am J Hum Genet.* 2009;84:771-779.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Hu B, Jin H, Shi Y, et al. Single-cell RNA-Seq reveals the earliest lineage specification and X chromosome dosage compensation in bovine preimplantation embryos. *The FASEB Journal.* 2024;38:e23492. doi:[10.1096/fj.202302035RR](https://doi.org/10.1096/fj.202302035RR)