Single-cell omics uncover gene dynamics shaping embryonic and extra embryonic lineages in pig blastocysts

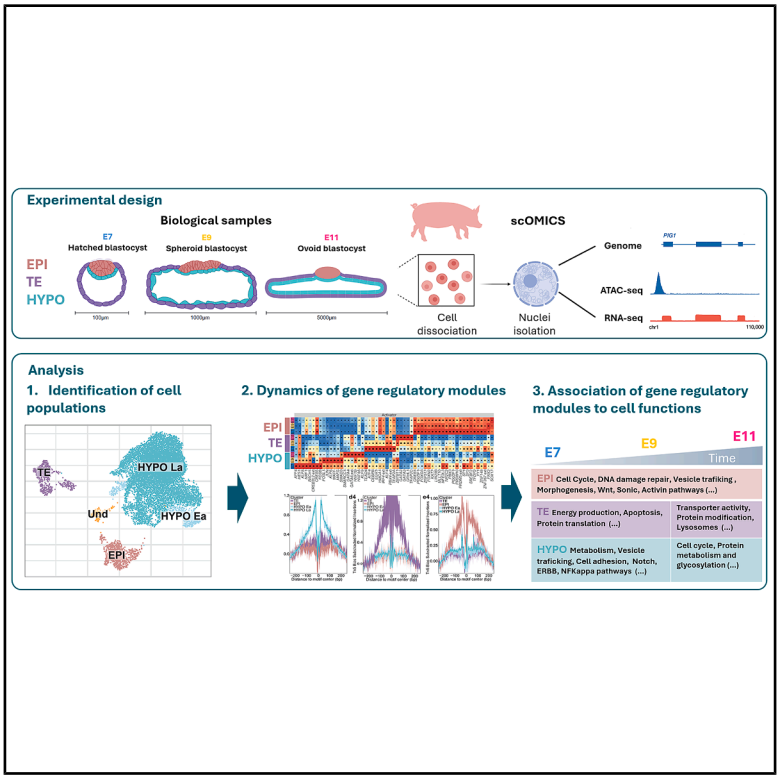
Adrien Dufour, Marie-Noëlle Rossignol, Patrick Manceau, Yoann Bailly, Stéphane Ferchaud, Marie-José Mercat, Ali G Turhan, Sarah Djebali, Sylvain Foissac, Jérôme Artus, et al.

iScience

Article

Single-cell omics uncover gene dynamics shaping embryonic and extra embryonic lineages in pig blastocysts

Graphical abstract



Highlights

- Single-cell multiomics maps embryonic and extra-embryonic lineages in pig blastocysts
- Epiblast and extra-embryonic cells show distinct epigenetic dynamics and regulatory programs
- Rapid shift in gene regulatory modules at the onset of elongation in extra-embryonic tissues
- Epiblast maintains a stable pluripotent regulatory state for days before gastrulation

Authors

Adrien Dufour, Marie-Noëlle Rossignol, Patrick Manceau, ..., Sylvain Foissac, Jérôme Artus, Hervé Acloque

Correspondence

jerome.artus@universite-paris-saclay.fr (J.A.),
herve.acloque@inrae.fr (H.A.)

In brief

Epigenetics; developmental biology; embryology; transcriptomics; model organism

Dufour et al., 2026, iScience *29*, 114519
January 16, 2026
https://doi.org/10.1016/j.isci.2025.114519

CellPress

iScience

CellPress
OPEN ACCESS

Article

Single-cell omics uncover gene dynamics shaping embryonic and extra embryonic lineages in pig blastocysts

Adrien Dufour,[1] Marie-Noëlle Rossignol,[1] Patrick Manceau,[2] Yoann Bailly,[2] Stéphane Ferchaud,[2] Marie-José Mercat,[3] Ali G. Turhan,[4] Sarah Djebali,[5] Sylvain Foissac,[5] Jérôme Artus,[4,6,*] and Hervé Acloque[1,6,7,*]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy en Josas, France
[2]INRAE, GenESI, La Gouvanière, 86480 Rouillé, France
[3]IFIP Institut du Porc, La Motte au Vicomte, BP 35104, 35651 Le Rheu Cedex, France
[4]Université Paris-Saclay, Inserm, UMRS 1310, 7 rue Guy Môquet, 94800 Villejuif, France
[5]GenPhySE, Université de Toulouse, INRAE, INP, ENVT, 31320 Castanet-Tolosan, France
[6]These authors contributed equally
[7]Lead contact
*Correspondence: jerome.artus@universite-paris-saclay.fr (J.A.), herve.acloque@inrae.fr (H.A.)
https://doi.org/10.1016/j.isci.2025.114519

SUMMARY

Late blastocyst development before implantation is a unique feature of ungulates, during which the epiblast proliferates and maintains pluripotency while extra-embryonic tissues expand dramatically, elongating to several tens of centimeters. The mechanisms coordinating these processes are not well understood. We performed single-cell omics profiling of porcine blastocysts from the hatched stage (E7) through early (E9) and late ovoid stages (E11). From 15,370 cells, we identified distinct embryonic and extra-embryonic populations with characteristic chromatin accessibility profiles. We reconstructed gene regulatory networks using enhancer-based eRegulons and validated them through motif occupancy analysis. Extra-embryonic tissues showed strong shifts in gene regulatory module activity at the onset of elongation, reflecting major transitions in morphogenesis and differentiation and the activation of pathways linked to cell morphology, proliferation, metabolism, trafficking, and biomolecule transport. In contrast, epiblast cells retained a stable transcriptional and regulatory identity from day 7 to day 11, immediately preceding the onset of gastrulation.

INTRODUCTION

Pigs are one of the main sources of animal products for humans and one of the most widely consumed meat worldwide. They also serve as an alternative to rodent models for studying early embryo development[1] and as an attractive biomedical model for human pathology.[2] Pig early embryonic development follows a pattern similar to other mammals. After fertilization, the embryo undergoes several cleavage divisions, followed by compaction and the formation of the blastocyst, composed of an inner cell mass (ICM) and an outer trophectoderm (TE). By 6-day post-fertilization (dpf), the ICM differentiates into the hypoblast or primitive endoderm (HYPO) and the epiblast (EPI). Unlike in rodents or primates, the porcine blastocyst begins to elongate from 7-dpf, transitioning through ovoid, tubular, and filamentous stages. During this process, the TE covering the epiblast (EPI), known as Rauber's layer, disappears.

We and others have explored the development of the pig pre-implantation embryo at the transcriptomics single-cell (scRNA-seq) level, before and during gastrulation.[3–6] These studies have highlighted the role of IL6-JAK-STAT signaling in ICM and TE formation. They also have shown that the transition of pluripotency status from a naive-like to a prime-like state is associated with changes in active signaling pathways and pluripotency marker expression. Indeed, naive-like pluripotency at embryonic-day-5 (E5) is characterized by high activity of IL6-STAT3 and PI3K-AKT signaling pathways associated with the expression of naive pluripotency markers (*KLF4*, *ESRRB*, and *STAT3*). Conversely, from E7 onwards, pluripotency shifts to a primed-like state marked by increased activity of TGFβ-SMAD2/3 signaling pathway associated with the expression of primed pluripotency genes (*NANOG*, *DNMT3B*, and *OTX2*). This transition is also accompanied by epigenetic changes through DNA and histone modifications, including X chromosome inactivation in female EPI cells, and a metabolic switch from OXPHOS to glycolysis.[3–5]

Interestingly, our previous single-cell study identified new subpopulations within extra-embryonic cell lineages. These include cells that secrete molecules required for implantation, such as interleukin-1 beta (*IL1B*), and a putative trophoblast stem cell population expressing *LRP2*, which later contributes to the development of the embryonic placenta.[5] Our data also confirmed that extra-embryonic subpopulations play key roles in the metabolism of fatty acids, vitamins, nutrient transport and conceptus elongation.[5] By integrating this dataset, we identified several transcription factors (TFs) and their associated

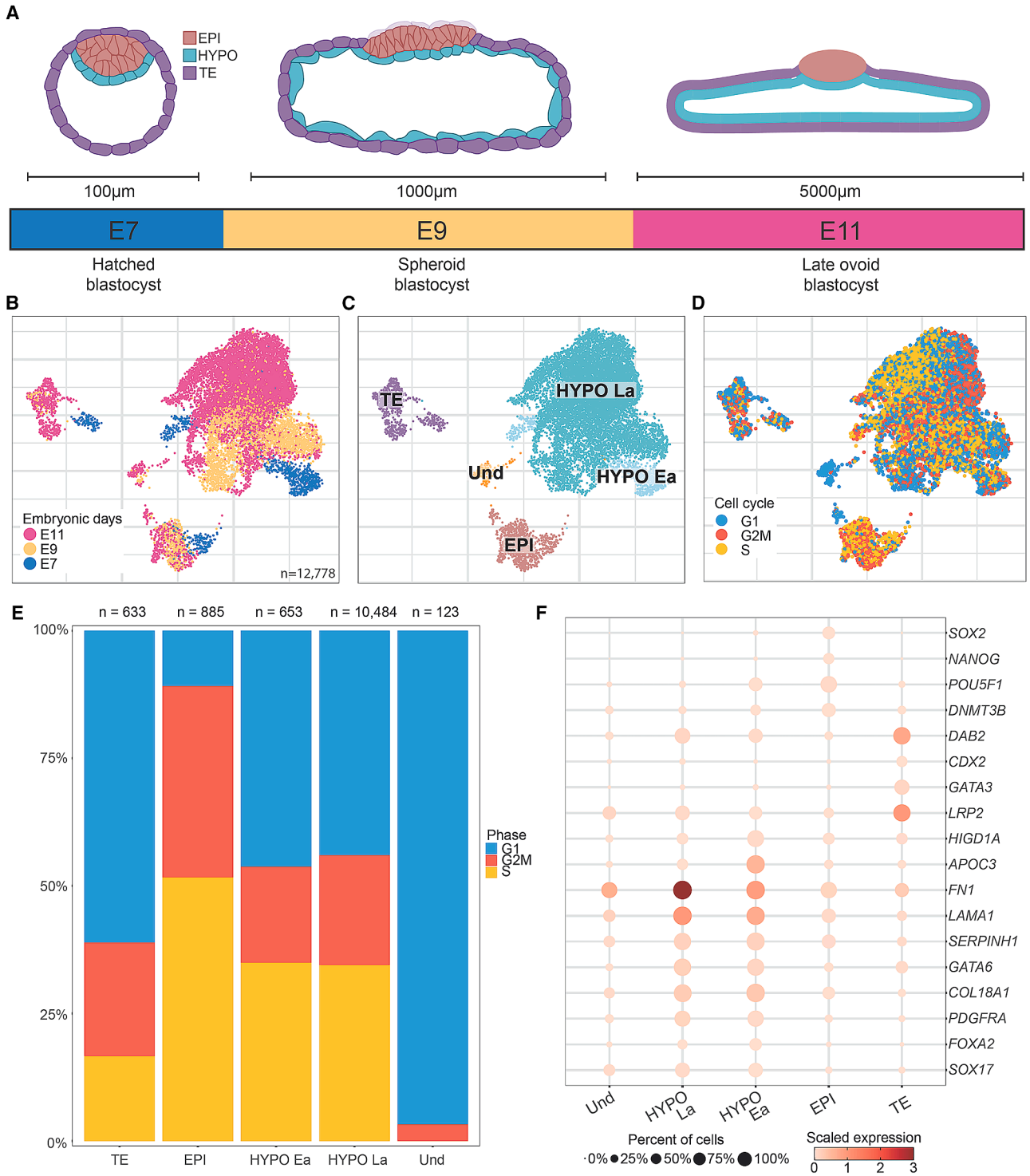


Figure 1. Single-cell multiomics sequencing of pig pre-implantation embryos

(A) Schematic view of pig embryo morphology for embryonic day (E) E7, E9, and E11. Cells from the EPI, TE, and HYPO are represented, respectively, in red, purple, and light blue.

(B) UMAP representation of all single-cells passing quality control colored by embryonic stage (E7 in blue, E9 in orange, and E11 in purple).

(C) UMAP representation of the assigned population: EPI (red), HYPO Late (turquoise), HYPO Early (light blue), TE (purple), and Und (undetermined, orange).

(legend continued on next page)

regulons,[7] as well as ligand-receptor pairs related to these populations. This integration allowed us to predict the upstream signaling pathways governing gene regulatory modules.

While these studies provide a better understanding of early pig embryonic development, they have not explored the epigenetic changes occurring during these stages. Such modifications are thought to be critical during this developmental window, as they likely regulate gene expression dynamics, influence lineage commitment, and establish the epigenetic landscape necessary for proper embryo implantation and subsequent development. Modifications, like *de novo* DNA methylation in the epiblast, are suggested by the increased expression of genes such as *DNTM3B*, *DNMT3A*, and *HELLS* in early pig blastocysts.[4] These questions have been addressed in other mammals, such as mice, using bulk ATAC-seq, histone profiling, and scATAC-seq.[8] More recently, single-cell multiomics (combining RNA-seq and ATAC-seq in the same cells) have emerged as a powerful approach for investigating gene regulatory networks (GRNs), with new analytical methods such as SCENIC+ recently applied to the fly brain,[9] though they remain underutilized in mammalian embryos.

Here, we produced and analyzed a single-cell multiomics dataset of pig preimplantation embryos across key developmental stages, from hatched blastocyst (E7) to spheroid/early ovoid blastocyst (E9) and late ovoid blastocyst (E11) (Figure 1A). We characterized 15,370 embryonic cells, which we clustered into distinct embryonic and extra-embryonic cell populations associated with specific chromatin accessibility landscapes. For each cell population, we then inferred GRNs based on enhancer-driven gene regulation modules (eRegulons)[10] and validated them through motif occupancy visualization. We detected known and new eRegulons for each cell population, and observed a strong dynamic in eRegulons activity within extra-embryonic tissues (TE and HYPO) at the onset of blastocyst elongation, while the activity of gene regulatory modules remained relatively constant in the epiblast.

RESULTS

Joint transcriptomic and epigenomic profiling of embryonic and extra-embryonic lineages during pig blastocyst development

We generated a paired transcriptomic and chromatin accessibility profile using the 10× Multiome technology from single nuclei of pig pre-implantation embryos collected at E7, E9, and E11 (Figure 1A and Table S1). A total of 12,778 cells passed quality control for both data modalities (Figures 1B and S1; Table S2). We identified four cell clusters that were classified as TE, HYPO early/late and EPI according to the expression of specific lineage markers (Figure 1C). To confirm our assignment, we plotted, on the same uniform manifold approximation and projection (UMAP), another assignment based on population signatures obtained from other scRNA-seq datasets obtained from pig pre-implantation embryos[5,6] (Figure S2). Within the hypoblast, we identified two clusters: early hypoblast (HYPO Ea) and late hypoblast (HYPO La). The two HYPO clusters express at similar levels *SOX17*, *FOXA2*, *PDGFRA*, *GATA6*, *COL18A1*, and *SERPINH1* (Figure 1F). HYPO Ea cells express markers such as *APOC3* and *HIGD1A* (Figure 1F), which we previously associated to early and intermediate hypoblast (HYPO Ea and HYPO In populations; Figure S2C)[5] and are exclusively present at the earliest E7 and E9 stages. HYPO La cells express higher levels of *FN1* (Figure 1F), similarly to more mature hypoblast (HYPO Ma) cell populations (Figures S2B and S2C).[5] Similarly, for TE, assignment based on population signatures confirm the existence of previously identified subpopulations, notably the *LRP2* expressing cell population (TE Lr) in E7 and E9 embryos, an intermediate TE population in E7 embryos (TE In2) and a more mature TE population (TE Ma2) in E11 embryos (Figure S2). Strikingly, the proportions of cells within the different populations (8% EPI, 87% HYPO, and 5% TE) differ from our previous study (4% EPI, 44% HYPO, and 52% TE)[5] as well as from other scRNA-seq studies on pig pre-implantation embryos.[6] This discrepancy likely originates from technical bias introduced during nuclei isolation procedure,[11] particularly to the difficulty to efficiently isolate TE nuclei.

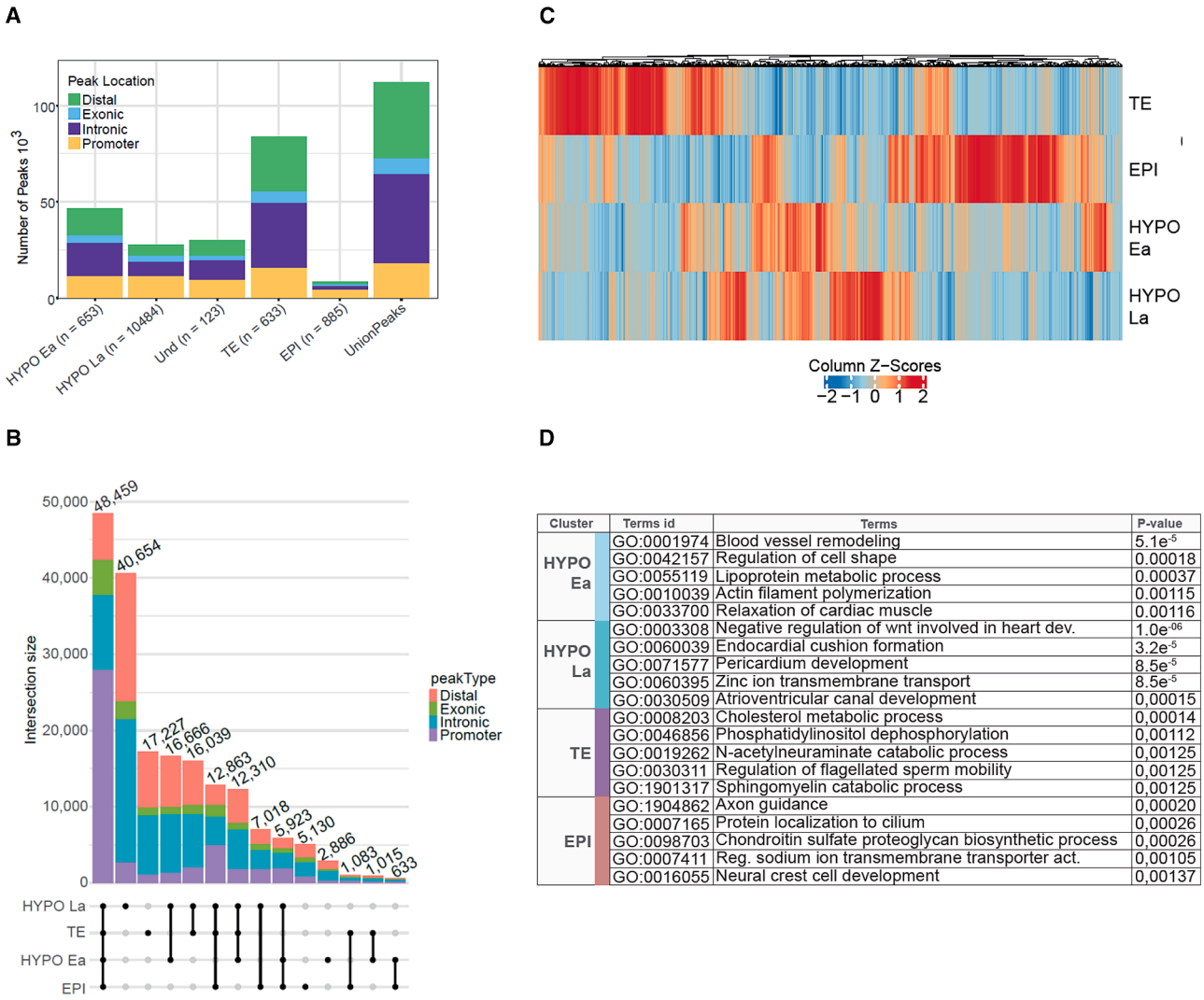
By performing cell cycle assignment using CellCycleScoring Seurat function (see "STAR Methods"), we observed that Hypo Ea and Hypo La populations exhibit similar cell cycle profiles, with approximately 50% of cells in proliferative phases (G2/M and S) and the remaining 50% in G1 (Figures 1D and 1E). In this dataset, our clustering analysis identified a single TE cluster characterized by the expression of *LRP2*, *GATA3*, *CDX2*, and *DAB2* (Figure 1F). However, our cell assignment analysis clearly reveals the presence of distinct subpopulations previously described as TE LRP2 (TE Lr) and TE Mature2 (TE Ma2) (Figures S2B and S2C).[5] TE cells show reduced proliferation compared with the other populations, with less than 40% of cells in G2/M and S phases (Figures 1D and 1E). EPI cells correspond to the epiblast signature (yellow, Figures S2B and S2C) and are associated with the expression of known pluripotency markers such as *NANOG*, *POU5F1*, *SOX2*, and *DNMT3A* (Figure 1F). EPI cells display a strong proliferative profile, with over 80% of cells in G2/M and S phases (Figures 1D and 1E), confirming our previous observation.[5] One small population was named undetermined (Und) and does not express any markers of embryonic or extra-embryonic tissues. This population shows evidence of cell cycle exit, with most cells accumulating in G1 phase.

Each cell population exhibited distinct epigenetic landscapes. In the whole dataset, we identified 108,810 chromatin accessibility peaks (Data S1), located either within genes (intronic and exonic), promoter regions (500 bp upstream and 100 bp downstream) or distal regions (Figure 2A). Forty-four percent of these peaks are

(D) UMAP representation of cell cycle assignment of each cell with respect to the three different phases of the cell cycle: Gap1 (G1, in blue), Synthesis (S, in yellow), Gap2 and mitosis (G2/M, in red). Cells have been assigned to the cell cycle phase obtaining the best score.

(E) Proportion of cells in G1 (blue), G2/M (red), and S (yellow) in each cluster.

(F) Dot plot of gene expression in cell clusters for selected markers. Red color gradient represents the relative expression level across clusters and dot size represents the percentage of cells in that cluster expressing the gene.



Cluster	Terms id	Terms	P-value
HYPO Ea	GO:0001974	Blood vessel remodeling	$5.1e^{-5}$
	GO:0042157	Regulation of cell shape	0.00018
	GO:0055119	Lipoprotein metabolic process	0.00037
	GO:0010039	Actin filament polymerization	0.00115
	GO:0033700	Relaxation of cardiac muscle	0.00116
HYPO La	GO:0003308	Negative regulation of wnt involved in heart dev.	$1.0e^{-06}$
	GO:0060039	Endocardial cushion formation	$3.2e^{-5}$
	GO:0071577	Pericardium development	$8.5e^{-5}$
	GO:0060395	Zinc ion transmembrane transport	$8.5e^{-5}$
	GO:0030509	Atrioventricular canal development	0,00015
TE	GO:0008203	Cholesterol metabolic process	0,00014
	GO:0046856	Phosphatidylinositol dephosphorylation	0,00112
	GO:0019262	N-acetylneuraminate catabolic process	0,00125
	GO:0030311	Regulation of flagellated sperm mobility	0,00125
	GO:1901317	Sphingomyelin catabolic process	0,00125
EPI	GO:1904862	Axon guidance	0,00020
	GO:0007165	Protein localization to cilium	0,00026
	GO:0098703	Chondroitin sulfate proteoglycan biosynthetic process	0,00026
	GO:0007411	Reg. sodium ion transmembrane transporter act.	0,00105
	GO:0016055	Neural crest cell development	0,00137

Figure 2. Epigenomic landscape of pig pre-implantation embryos

(A) Barplot of peak quantification according to genome location in each cluster and resulting union of peaks.

(B) Upset plot comparing the number of peaks overlap between cell populations.

(C) Heatmap reflecting differential promoter accessibility in each cell cluster (500 bp upstream and 100 bp downstream of the promoter).

(D) GO: biological process enrichment in the significantly enriched DEGs in each cluster. The top 5 are shown. p-values originate from hypergeometric tests and are adjusted for multiple testing.

common to all populations while 15% are unique to HYPO and TE and 5% in EPI (Figure 2B). This observation prompted us to investigate to what extent the embryonic epigenetic landscape differs from that of differentiated somatic cells. By comparing our profiles of chromatin accessibility peaks with those from other somatic tissues,[12] we observed that the majority of these peaks are shared across cell types, with only a small proportion being specific to blastocyst tissues (Figure S3 and Table S3).

We then determined whether blastocyst-specific regions reflect unique regulatory features associated with pluripotency and early lineage commitment. Using these peaks, we identified 14,016 differentially accessible peaks (DAPs) between the cell clusters including 980 located on promoter regions, highlighting specific epigenomic landscapes of each cell population (Figure 2C and Table S4).

By integrating differentially expressed genes (DEGs) with differentially accessible promoter regions (DAPRs) across cell populations, we identified a strong correlation between chromatin accessibility profiles and the expression of established lineage markers (Figures 3 and S4; Tables S5 and S6). This analysis also provides functional insight into *cis*-regulatory elements active in pig blastocyst cells. Notably, regions of open chromatin were associated with elevated expression of key genes such as *APOC3*, *FN1*, *HIGD1A*, and *LAMA1* in the HYPO; *DAB2*, *GATA3*, and *CDX2* in the TE; and *SOX2*, *NANOG*, and *POU5F1* in the EPI (Figures 3 and S4).

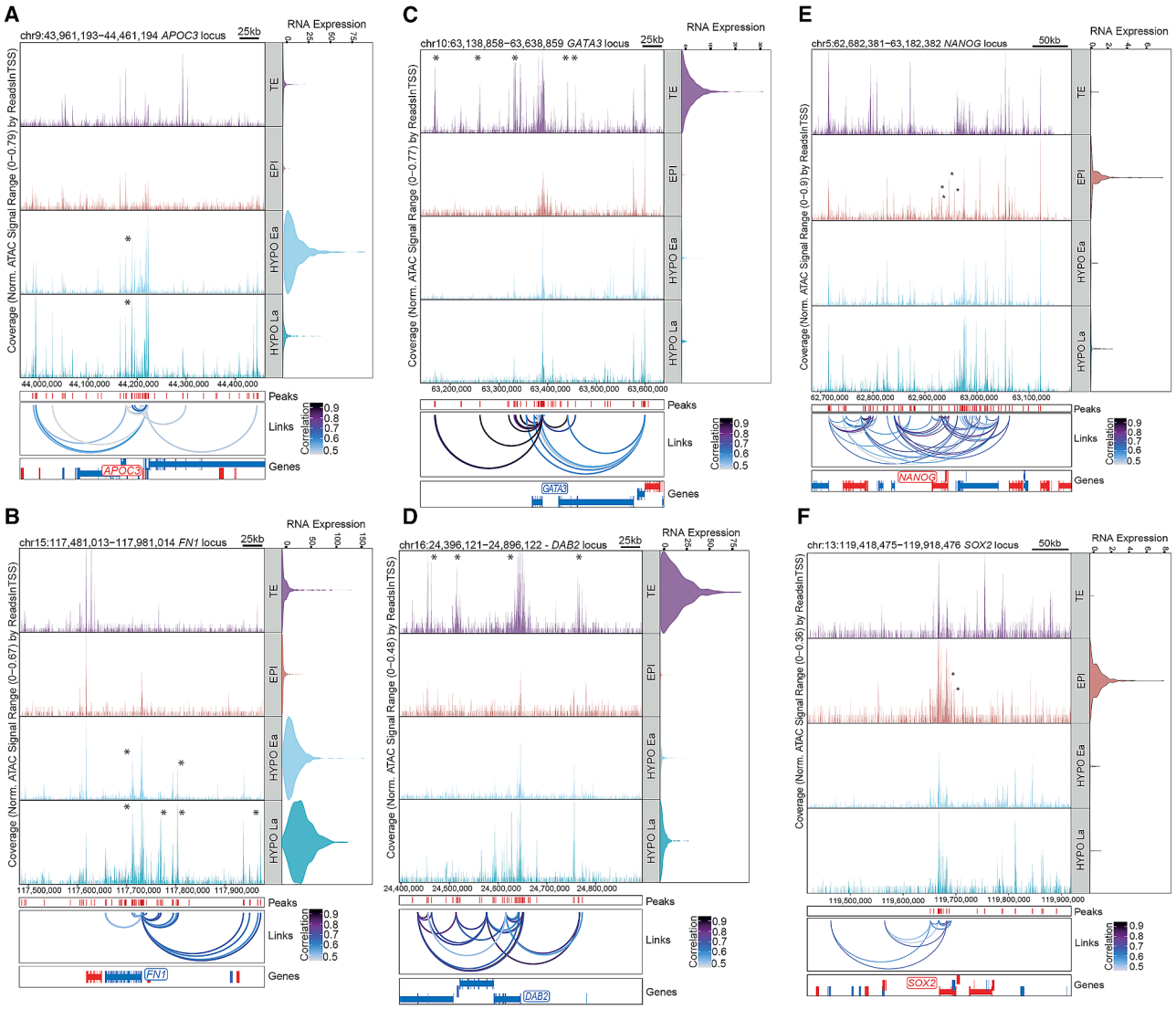


Figure 3. Profiles of peak-to-gene linkages for HYPO, TE, and EPI markers in pig embryos

(A–F) Coverage plots showing aggregated single-cell ATAC signals around *APOC3* (A), *FN1* (B), *GATA3* (C), *DAB2* (D), *NANOG* (E), and *SOX2* (F) loci within the different cell types. On the right of each coverage plot, violin plots show their expression levels for each cell type. Below each coverage plot, loops predicted by ArchR indicate peak-to-gene linkages identified on the full dataset (colored by importance score). Stars indicate peak-to-gene links with a correlation score over 0.8.

We then performed functional enrichment analysis using gene ontology (GO) terms, based on DEGs and DAPRs. We observed distinct regulatory landscapes across embryonic lineages (Figure 2D; Tables S7 and S8). In the EPI, we identified a set of shared GO terms associated with neuronal development, morphogenesis, and cell differentiation, suggesting an early transcriptional priming toward ectodermal fates. We also observed shared GO terms associated with signaling pathways known to be regulated in pluripotent stem cells, including Wnt, BMP, and Notch. In contrast, several signaling pathways exhibited open promoter regions without corresponding gene expression, implying a poised yet transcriptionally inactive state. These included key regulators of early embryonic development such as pathways controlled by neurotrophin, interleukin-1, ephrin receptor, retinoic acid receptor, and epidermal growth factor receptor signaling (Tables S7 and S8), which may reflect epigenetic readiness for future activation. In the TE, GO enrichment predominantly highlighted terms related to energy production, molecular transport, and metabolic processes, consistent with the high biosynthetic and proliferative demands of this lineage. In contrast, in the HYPO, we observed a dynamic change in pathways and functions. During early hypoblast development (HYPO Ea), we observed enrichment for MAPK signaling-related terms, including "negative regulation of MAPK cascade," "positive regulation of ERK1 and ERK2 cascade," "regulation of JNK cascade," and again the "negative regulation of Notch

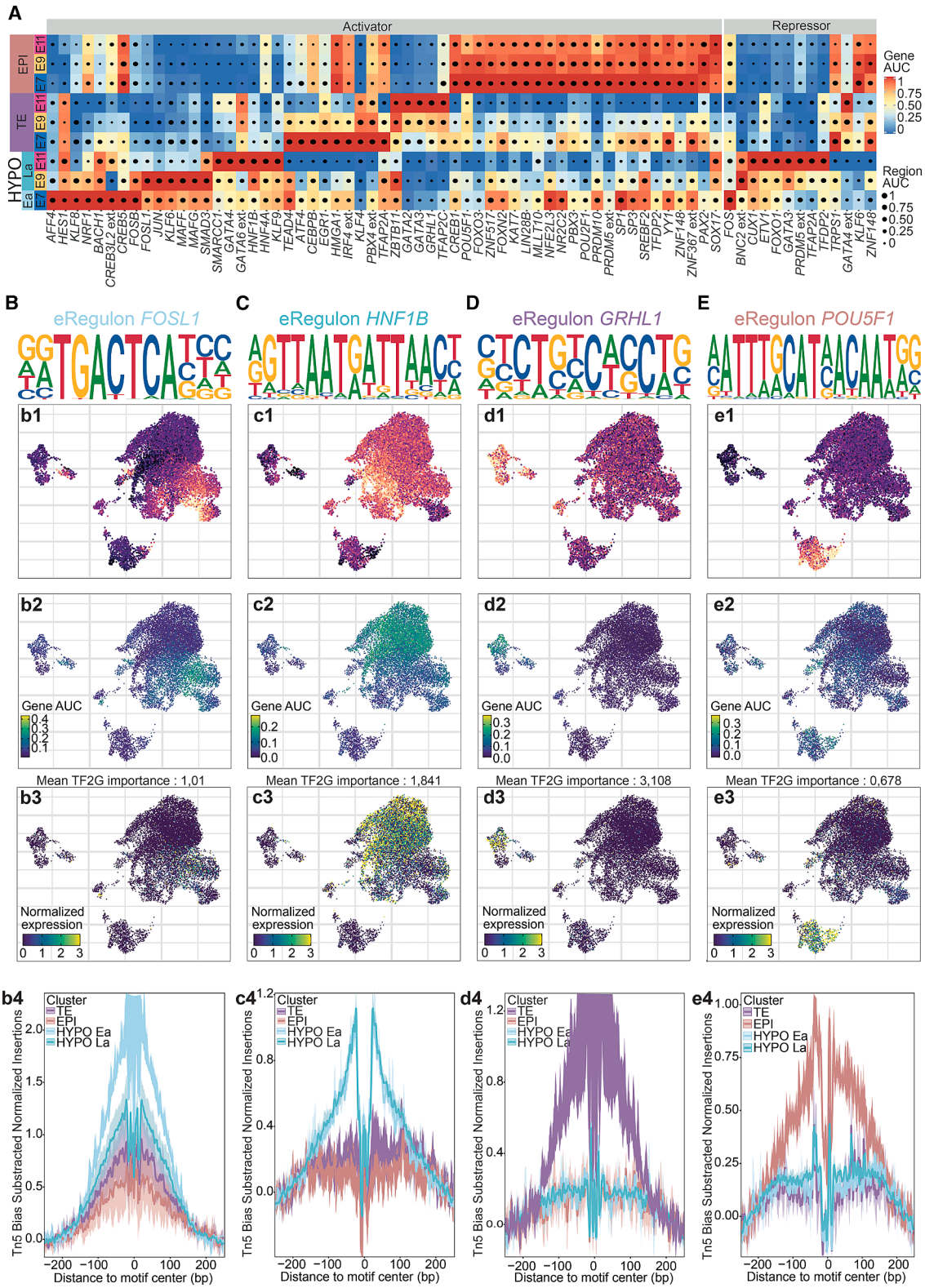


Figure 4. Active gene regulatory modules (eRegulon) of embryonic and extra-embryonic cell populations

(A) Heatmap of eRegulons identified in each cell population and embryonic stage, classified either as activator or repressor. Dot size represents scaled region AUC and color gradient represents target gene expression AUC. TFs name with ext (extended) represents TFs whose annotation is inferred based on motif similarity.

(legend continued on next page)

signaling pathway." As development progressed to the late hypoblast stage (HYPO La), GO term enrichment shifted toward BMP and TGF-β signaling, suggesting a transition in dominant signaling pathways between E7 and E11. This temporal shift likely reflects functional maturation and lineage specification within the hypoblast.

Comprehensive regulatory profiling of embryonic and extra-embryonic cell populations

We next explored the GRNs at play into those populations. To do this, we used SCENIC+[10] to identify enhancer-driven regulons (eRegulons). These eRegulons represent functional gene modules regulated by enhancers. Each eRegulon is defined by a TF (which gives the eRegulon its name) and its associated set of target genes. Each target gene is co-expressed with the TF and has at least one accessible *cis*-regulatory element containing the TF's DNA-binding motif. We identified 57 activator TFs whose target genes show both chromatin accessibility and induced expression, as well as 13 repressor TFs, characterized by closed chromatin at their target regions and suppressed gene expression (Table S9). For each eRegulon, we visualized the expression levels of target genes (blue to red scale) and their chromatin accessibility (dot size) across clusters and developmental stages (Figure 4A). Compared to our previous SCENIC study on pig pre-implantation embryos,[5] we identified fewer regulons but with higher confidence and 61% of TFs identified here were identified in both studies.

This analysis revealed key TFs with well-established roles in lineage specification, including *GATA2*, *GATA3*, and *TFAP2C* in TE, *GATA6*, *FOSL1*, and *HNF4A* in HYPO, and *POU5F1*, *LIN28B*, and *ZNF148* in EPI. In addition, we identified new eRegulons, including *GRHL1*, *BNC2*, *IRF4*, *JUN*, *PAX2*, *PRDM5*, and *TRPS1*. We also observed that some eRegulons regulated overlapping sets of genes due to their high motif similarity. These include the GATA TFs family (*GATA6*, *GATA4*, and *GATA3*), TFs linked to the AP-1 complex (*ATF*, *JUN*, *BATF*, and *FOS*)[13] as well as two other families of TFs sharing a common set of genes: *MAFF*, *FOSL1*, and *BACH1* for the first set and *PRDM5*, *LIN28B*, and *PBX3* for the second set (Figure S5).

To confirm the relevance of the identified eRegulons, we assessed motif enrichment within DAPs (Table S10). For each cell cluster, we selected one of the most enriched motifs (*FOSL1* for Hypo Ea, *HNF1B* for Hypo La, *GRHL1* for TE, *POU5F1* for EPI) and visualized their enrichment scores on a UMAP projection (Figures 4b1–4e1). In parallel, we displayed their expression and the expression levels of their respective target genes (Figures 4b2–4e2 and 4b3–4e3). To further support the regulatory activity of these TFs, we inferred their footprint profiles around their respective DNA binding motifs by analyzing Tn5 activity bias around their motif sequences (see "STAR Methods"). The resulting TF footprints for FOSL1, HNF1B, GRHL1, and POU5F1 are shown in Figures 4b4–4e4. Together, these analyses confirm the high specificity of the identified eRegulons across distinct cell populations, reflected in both TF activity and target gene expression.

This analysis also revealed enriched motifs for TFs that have not been identified by our SCENIC+ analysis but are known to be lineage-specific markers, like *CDX2* in TE or *MTA3* and *BCL11A* in EPI (Figure S6).

Our analysis shows that the activity of eRegulons detected in the epiblast from E7 to E11 remained relatively stable, with some nuances, since most of them showed a slight decrease in activity from E7 to E11 (such as *POU5F1*), whereas *SOX17* and *TFAP2A* showed a slight increase in activity between E7 and E11. This sharply contrasts with extra-embryonic tissues (HYPO and TE) where eRegulon activity drastically changed between E7 and E11. Interestingly, this regulatory dynamic occurs concomitantly with the drastic morphological changes that occur in these tissues at the start of elongation and should have an impact on the biology of these cells in terms of differentiation and cellular functions.

To understand whether this regulatory dynamic is associated with distinct biological functions, we performed GO and KEGG enrichment analyses on the target genes of each eRegulon, to identify relevant biological processes and pathways (Tables S11, S12, and S13). Surprisingly, most eRegulons are enriched for genes with functions specific to the biology of each cell population. In the EPI cluster, for instance, eRegulons predominantly target genes involved in embryonic development, particularly neural tissue formation (e.g., *LIN28B*, *PBX3*, *PRDM5*, and *FOXO3*), as well as in the regulation of nuclear genome activity and DNA repair (*POU2F1*, *SP3*, *YY1*, *ZNF148*, *ZNF367*, and *FOXN2*). Several eRegulons also control genes related to key signaling pathways, including Wnt (*PRDM5* and *LIN28B*), Sonic Hedgehog (*LIN28B*), RAP1 (*PAX2*), and Activin (*CREB1*). Additionally, some eRegulons appear more specialized, such as *MLLT10* regarding cell cycle regulation, *SREBF2* for cell metabolism, and *SP1* for intracellular trafficking (Tables 1 and S11).

In the TE, at early stages (E7 and E9), eRegulons are predominantly associated with terms linked to fundamental biological processes, like developmental processes (*KLF4* and *CEBP*), gene regulation (*KLF4*) and energy production (HMGA1). By E11, however, active eRegulons are associated with terms that shift toward functions related to more specialized cell functions, including lysosomes biology (*GATA3*, *GRHL1*, and *TFAP2C*), transporter activity (*TFAP2C*), metabolism (*GRHL1*, *ZBTB11*, and *GATA3*), and protein translation (*ZBTB11*) (Tables 1 and S12). We also observed a transition of the activity of *KLF4*, *GATA2*, and *GATA3* eRegulons. *KLF4* expression and eRegulon activity is restricted to TE cells from E7 embryos, predominantly

(B–E) Validation of eRegulons *FOSL1* (B), *HNF1B* (C), *GRHL1* (D), and *POU5F1* (E).
(b1–e1) UMAP visualization of top motif enrichment in differentially accessible genes (DAGs) for *FOSL1*, *HNF1B*, *GRHL1*, and *POU5F1* between each cluster. Gradient represents motif *Z* score calculated from ChromVar at the top motif PWMs used for the TFs.
(b2–e2) UMAP visualization of AUC of each cell for the eRegulons *FOSL1*, *HNF1B*, *GRHL1*, and *POU5F1*.
(b3–e3) UMAP visualization of *FOSL1*, *HNF1B*, *GRHL1*, and *POU5F1*. Mean of TF2G (TF-to-gene) scores are provided for each TF.
(b4–e4) TFs footprint of *FOSL1*, *HNF1B*, *GRHL1*, and *POU5F1* in each cluster at their known binding motifs. Footprints are inferred by assessing Tn5 activity bias around the DNA binding motif.

Table 1. Summary of the functional enrichment analysis of active target genes inferred from eRegulon analysis across embryonic lineages in porcine blastocysts at days E7, E9, and E11

Overview of the functional enrichment analysis of active target genes from eRegulon

	Spheroid blastocyst E7	Early ovoid blastocyst E9	Late ovoid blastocyst E11
EPI	● Neurogenesis: LIN28B, PBX3, PRDM5 ● Developmental processes: FOXO3 ● Gene regulation and chromatin modification: POU2F1, SP3, YY1, ZNF148, ZNF367 ● DNA damage and repair: FOXN2, ZNF148, ZNF367 ● Cell cycle and mitosis: MLLT10 ● Vesicular transport and intracellular trafficking: SP1 ● Post-translational modifications: KAT7 ● Metabolism and anabolism: SREBF2 ● Wnt pathway: PRDM5, LIN28B; Sonic pathway: LIN28B; RAP1 pathway: PAX2; Activin signaling: CREB1		
TE	● Protein translation and associated structures: ATF4, HMGA1 ● Developmental processes: CEBPB, KLF4 ● Mitochondrial functions and energy production: HMGA1 ● Gene regulation and apoptosis: KLF4		● Role of lysosomes and vacuoles in cellular degradation: GATA3, GRHL1, TFAP2C ● Protein modification and degradation: ZBTB11 ● Enzymatic activity involved in nucleic acid processing, small molecule metabolic process: ZBTB11, GATA3 ● Retinol metabolism: GRHL1 ● Transmembrane transporter activity: TFAP2C
HYPO	Pleiotropic functions (gene regulation, metabolism, cell cycle …): HES1, NRF1, BACH1, JUN, MAFF Regulation of apoptosis: FOSB, MAFF Vesicle trafficking: KLF6 Cytoskeleton, ECM, and cell adhesion: KLF6 Phosphorus metabolic process: SMAD3 Notch and ERBB pathways: JUN; PI3K/AKT pathway: KLF6; NF kappa B pathway: MAFF; TLR cascade and IL1/IL17 pathways: MAFF		Cell cycle: SMARCC1 Protein metabolism and glycosylation: SMARCC1, GATA4, HNF4A

Each functional category is associated with eRegulons Transcription Factors. Colored text highlights distinct functional themes and signaling pathways identified within each lineage.

within the previously described LRP2-positive TE cells[5] while *GATA2* and *GATA3* eRegulons activity peak in LRP2-positive TE cells from E9 embryos and begins to decrease by E11 as cells progress toward a more mature TE state (Figures S2 and S7).

A similar dynamic is observed during HYPO development, where eRegulon activity reflects cellular transitions. In HYPO Ea, active eRegulons are mainly associated with terms linked to morphogenesis and associated signaling pathways (*JUN*, *KLF6*, and *MAFF*) that mirror HYPO formation. In contrast, eRegulons active in HYPO La are predominantly linked to cell metabolism and protein glycosylation (*SMARCC1*, *GATA4*, and *HNF4A*) (Tables 1 and S13).

To complete our analysis on the regulatory landscape and its association with biological functions, we investigated peak-to-gene associations to identify potential *cis*-regulatory elements driving gene expression in each cell population. A total of 15,991 peak-to-gene regulatory links were identified (see "STAR Methods" and Table S6). To visualize these regulatory interactions, we selected links exhibiting a variance score above 0.25 and grouped the corresponding peaks into five different k-means clusters, representing the four main cell populations (Figure 5). For each cluster, the top 1,000 highly regulated genes (HRGs) with the highest number of associated peaks were subjected to functional enrichment analysis (Figure 5 and Table S14).

This analysis reinforces the GO annotations previously performed on DEGs and genes from active eRegulons, as shown in Figure 1G and Table 1. The first two clusters of HRGs, predominantly expressed in EPI, are enriched for GO terms related to neurogenesis and morphogenesis. The third HRG cluster, specific to the TE, is enriched for pathways involved in lipid metabolism and transport. Additionally, this cluster also shows enrichment for GO terms related to immune response processes (Figure 5). In the fourth cluster, specific to HYPO, HRGs are associated with terms linked to cell adhesion processes mediated by integrins, while HRGs from cluster 5 genes are associated with terms linked to endoderm development and negative regulation of Wnt signaling pathways. Notably, among these HRGs, GATA6—a key transcriptional regulator of hypoblast lineage specification—was found to be associated with two distal enhancers strongly correlated to its expression in HYPO (asterisks in Figure S8), mirroring similar enhancer—gene associations observed for GATA3 in TE and SOX2 in EPI (Figure 3).

DISCUSSION

This study provides the first comprehensive annotation of the regulatory landscape in the pig blastocyst, characterizing each distinct cell population by linking gene expression with chromatin accessibility.

This study completes our knowledge of the dynamics of the epigenome and gene regulation in mice and humans[14,15] by focusing on a developmental stage that is specific to many other mammals, where blastocyst development continues for

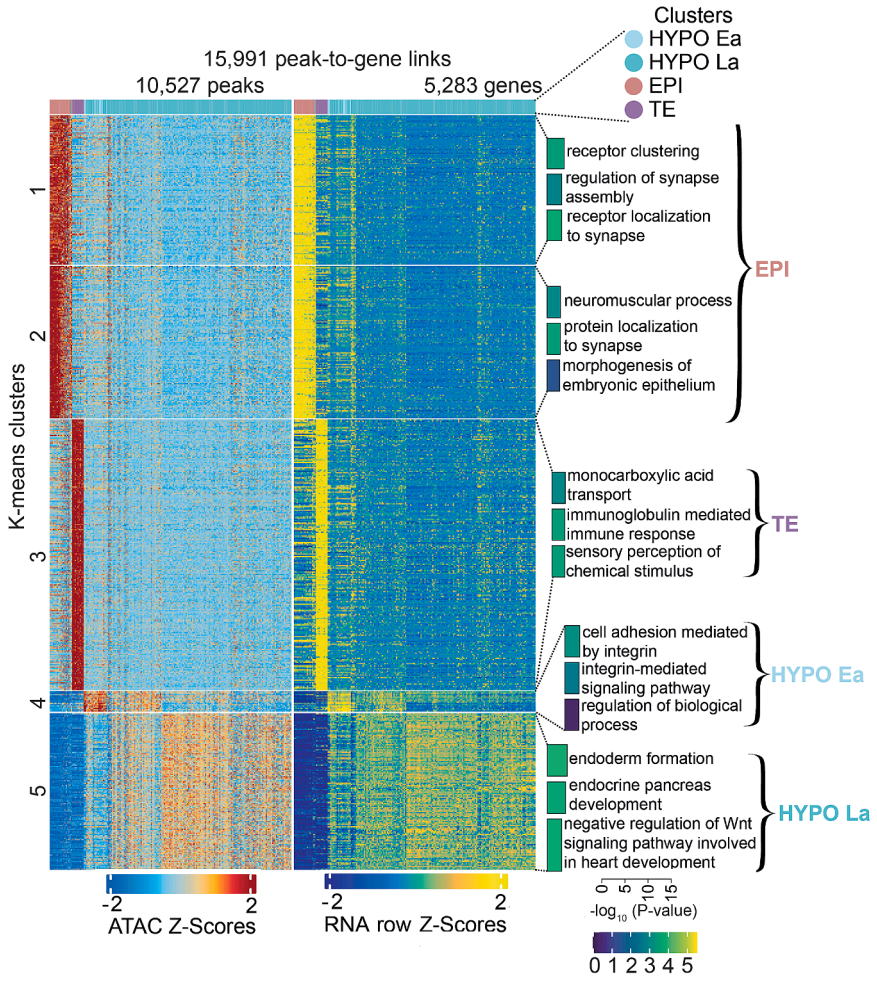


Figure 5. Relation between gene accessibility and gene expression within cell population

Heatmap representing chromatin accessibility on the left and gene expression on the right for 5,000 peak-to-gene linkages and clustered using the *k*-means clustering algorithms. In each cluster, the 1,000 most HRGs (e.g., which have a high number of peaks linked to them) were used for GO: Biological process enrichment and the top 3 terms for each cluster are shown. Color gradient represents the log of the fold enrichment among genes in peak-to-gene linkage.

several days prior to implantation.[1] Our analyses enabled us to annotate numerous regulatory elements whose chromatin accessibility correlates with the expression of key TFs defining the distinct cell populations of the blastocyst (HYPO, EPI, and TE). These results will support comparative studies on GRNs—both conserved and divergent—across mammalian species, with potential insights into species-specific traits and associated evolutionary strategies.

Our findings reveal two distinct regulatory dynamics between embryonic and extra-embryonic tissues. In the epiblast, despite the divergence of pre-implantation development in pigs from that of rodents and primates beyond the hatched blastocyst stage, the core regulatory gene networks, including *POU5F1*, *LIN28B*, *PRDM5*, *YY1*, and *POU2F1*,[11,16–19] are conserved and show limited variation between embryonic day 7 and day 11. This also includes regulators of known signaling pathways involved in pluripotency, such as *CREB1*, which regulates the MEK/ERK pathway,[20] *FOXO3*, which regulates the PI3K/Akt[21] and *ZNF148*, a Notch1 repressor.[22] These observations suggest that epiblast pluripotency is relatively stable during this 4-day window and resembles the pluripotent state described in other mammals. Given that gastrulation in pigs begins around embryonic day 11.5 at the tubular stage,[6] it is likely that major changes in GRNs and epigenomic states occur only at this tubular blastocyst stage, just beyond the ovoid stage examined in this study.

In the epiblast, among the identified eRegulons, several are linked to neural tissue development and neurogenesis, functions also found in the functional annotations of HRGs. This supports previous observations in mice, where mouse embryonic stem cells were shown to exhibit a default proneural state in absence of extrinsic signals.[23] Interestingly, we identified *MLLT10* as an eRegulon involved in cell cycle regulation, confirming prior findings in which *DOT1*, a known partner of *MLLT10*, plays a role in controlling the cell cycle and in the differentiation of both mouse and human pluripotent stem cells.[24–26] We also identified *SREBF2* as an eRegulon, with its target genes primarily associated with the regulation of cellular metabolism, particularly lipid metabolism. Its role in pluripotent stem cell biology remains largely unexplored.

In contrast, the regulatory dynamics in extra-embryonic tissues are markedly different. We observed a rapid evolution in eRegulon activity during the studied time window, coinciding with the onset of extra-embryonic tissue elongation. In both TE and HYPO, certain eRegulons are predominantly active in the hatched blastocyst prior to elongation, while others become specifically activated during the initiation of the elongation process.

In the hypoblast lineage, we identified two cell clusters in this dataset, classified as early (HYPO Ea) and late (HYPO La) hypoblast. Compared to our previous study, fewer subpopulations were detected using clustering approaches but an alternative assignment based on population-specific signatures confirmed the presence of the previously described subpopulations.[5] HYPO Ea comprises cells from early and intermediate hypoblast, while HYPO La includes visceral and mature hypoblast cells (Figure S2). In HYPO Ea numerous eRegulons from the AP-1 TF complex, including *FOSL1*, *BACH1*, *FOSB*, and *JUN*,[27] are associated with cell proliferation. Other eRegulons active in HYPO Ea, including *KLF8*, *NRF1*, *CREB*, *KLF6*, and

MAFF, as well as those active in HYPO La, such as *GATA4, HNF4A* and *KLF9*, were previously identified in the visceral hypoblast subpopulation.[5] UMAP visualisation of target gene expression for *KLF6*, *FOSB*, *SMAD3* and *GATA6* using the dataset from Dufour et al.[5] revealed that these genes are not exclusively expressed in the visceral hypoblast, although they appear upregulated in this population (Figure S9). Similarly, UMAP visualisation of target genes from regulons previously identified in the visceral hypoblast[5] using SCENIC, such as *FOSL, KLF9, E2F1, SREBF2*, confirmed that their expression is not restricted to this population, even though some, including *E2F1*, *SREBF2*, and *FOSL1*, are upregulated in visceral hypoblast cells (Figures 4B and S10).

While most of these eRegulons function as pleiotropic regulators, some exhibit more specialized roles. For instance, *KLF6* predominantly regulates target genes involved in cell adhesion, extracellular matrix interactions, and vesicle formation and trafficking, whereas *FOSB* primarily controls the expression of genes linked to apoptosis.

In contrast, eRegulons associated with classical hypoblast markers from the *GATA* and *HNF* families, as well as *SOX17*[28] show increased activity in the late hypoblast. This transition is accompanied by extensive remodeling of chromatin accessibility and enhancer activity, leading to a significant upregulation of the target genes of these eRegulons despite the fact that these TFs, such as *GATA4* and *GATA6*, are already expressed earlier as markers of hypoblast identity. In line with hypoblast maturation, we observed biological functions such as protein metabolism and glycosylation, which are characteristic of more differentiated hypoblast cells. These functions are absent in the early hypoblast, even though the TFs that drive them—including *SMARCC1*, *GATA4*, and *HNF4A*—are already expressed in HYPO Ea. This temporal shift between TFs expression and target gene activation, driven by differential chromatin accessibility, has also been observed in the mouse embryo, notably for *GATA3* in the TE.[14]

For the TE, although subpopulations could not be distinguished using clustering approaches, assignment based on population-specific signatures confirmed the presence of previously identified subpopulations, notably the LRP2-expressing cell population (TE Lr) in E7 and E9 embryos and a more mature TE population (TE Ma2) in E11 embryos (Figure S2). Similar to what we found for the hypoblast, we observed a strong dynamic in eRegulon activity across developmental stages. In the earlier, less mature stages, we identified eRegulons associated with morphogenesis (such as *KLF4* and *CEBPB*) and mitochondrial activity (such as *HMGA1*). *HMGA1* has been shown to regulate cell proliferation and migration in human TE,[29] and is itself regulated by the LIN28/let-7 axis, which plays a role in TE elongation in sheep.[30] We also observed that the activity of *KLF4*, *GATA2*, and *GATA3* eRegulons peaks successively in cells assigned to LRP2-expressing TE cell population, which we identified as a putative TE stem cell population, exhibiting a characteristic salt-and-pepper distribution.[5] This is consistent with the previously established function of KLF4 as a key factor in reprogramming human fibroblasts into trophoblast stem cells.[31] This observation further supports a role for these TFs in trophoblast lineage specification, not only during early early ICM-TE segregation,[32,33] but also at the onset of blastocyst elongation.

In agreement with Perez-Gomez et al.,[34] we did not observe specific expression or activity of *TEAD4* in the TE. Instead, the *TEAD4* eRegulon is predominantly active in the HYPO (Figure S11). Our analysis further supports, in another ungulate species, the observation by Perez-Gomez et al.,[34] that TEAD4 is dispensable for TE formation and development in bovine blastocysts, even at later stages, including the onset of elongation, in contrast with its essential role in mice.[35,36]

At later stages, we observed the activation of eRegulons corresponding to well-established TE-specific TFs, including *GATA2*, *GATA3*, and *TFAP2A/C*, as well as novel TFs and regulons not previously described in porcine TE, such as *GRHL1* and *ZBTB11*. Similar to what was observed for the hypoblast, there is a temporal shift between the activation of these eRegulons and the expression of their associated TFs.[14] This shift reflects rapid epigenomic remodeling and changes in the accessibility of regulatory regions in the target genes comprising these eRegulons across the different developmental stages. The functions associated with these eRegulons are characteristic of more differentiated cells and include lysosomal trafficking (*GATA3*, *GRHL1*, and *TFAP2C*), post-translational modifications (*ZBTB11*), and membrane transporter activity (*TFAP2C*). *GRHL1* has been shown to regulate trophoblast differentiation[37] and has been proposed to act as a pioneer TF that primes epithelial enhancers.[38] Interestingly, *GRHL1* is itself regulated by *KLF4*, another TF active in TE at earlier stages.[39]

Finally, as part of the Functional Annotation of ANimal Genomes (FAANG) international initiative, this work contributes to the FAANGSingleCell objectives by delivering single-cell atlases for key tissues in farm animal species.[40] By integrating gene expression and chromatin accessibility at single-cell resolution, the study robustly identifies key regulatory networks active during early developmental stages and will help to pinpoint genetic variants that may influence these regulatory mechanisms, with potential implications for traits of agricultural interest. To ensure the reproducibility and reusability of our datasets, all metadata and data generated in this study have been deposited in the FAANG Data Portal (https://data.faang.org), facilitating access and use by the broader research community.

Altogether, our work using multiomic data at the single-cell level enabled us to identify changes in gene expression associated with epigenetic modifications during the pre-implantation period of pig embryos. These changes are indicative of the maturation of the extra-embryonic tissues during elongation and are accompanied by a dynamic evolution of GRNs with decreasing or increasing activity of TFs in the different populations.

Limitations of the study

Our work is based on *in silico* prediction of the gene regulatory landscape in action in embryonic and extra-embryonic tissues of the pig blastocyst. While our data highlight an highly dynamic profile of the epigenomic landscape and eRegulon activity in these tissues over time, experimental validation will be necessary to elucidate the precise mechanisms driving their

formations and functions during this developmental window. Conducting such experiments in pig embryos is, however, particularly challenging, as it requires considerable time, specific molecular and cellular tools, and careful adherence to ethical constraints associated with large-animal research (3R principles). Therefore, these functional studies must focus on a limited number of key genes and mechanisms. We believe that our dataset offers a valuable resource for the community, enabling future functional investigations into the mechanisms of embryonic and extra-embryonic tissues formation.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Hervé Acloque (herve.acloque@inrae.fr).

Materials availability

All detailed protocols are available in FAANG Data portal: https://api.faang.org/files/protocols/samples/INRAE_SOP_PLUS4PIGS_EMBRYOS_SAMPLING_PROTO2_20230131.pdf; https://api.faang.org/files/protocols/samples/INRAE_SOP_PLUS4PIGS_EMBRYOS_DISSOCIATION_PROTO3_20230131.pdf; https://api.faang.org/files/protocols/samples/INRAE_SOP_PLUS4PIGS_EMBRYOS_DISSOCIATION_PROTO3_20230131.pdf.

Metadata related to the biological samples used in this study have been submitted to BioSamples: https://www.ebi.ac.uk/biosamples/.

Data and code availability

- All raw sequencing data are available in FAANG Data portal (https://data.faang.org/home) and European Nucleotide Archive (ENA) under accessions ENA: PRJEB81663 and ENA: PRJEB83269.
- The code used for the analysis is available at GitHub: https://github.com/INSERM-U1141-Neurodiderot/pig_embryo_multiome.
- Txdb package used for Archr is available at Zenodo: https://doi.org/10.5281/zenodo.10844555 Pig SCENIC+ database is available at Zenodo: https://doi.org/10.5281/zenodo.14833685. Any additional information required to reanalyze the data reported in this paper are available from the lead contact upon request.

ACKNOWLEDGMENTS

We are grateful to the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul: https://doi.org/10.15454/1.5572369328961167E12) for providing help, computing, and storage resources. We are grateful to people from the INRAE experimental farm GenESI: https://doi.org/10.15454/1.5572415481185847E12 who took care of the animals. We are grateful to Sylvain Bourgeais and the people from the Unité Expérimentale de Physiologie Animale de l'Orfrasière (UEPAO) INRAE. We thank the @BRIDGe facility for excellent technical assistance. We thank people from our respective laboratories for their helpful comments and discussion. We thank the foundation Vaincre le Cancer-NRB for the acquisition of the Chromium 10× machine. This work was supported by the ANR PluS4PiGs (ANR-19-CE20-0019) and the ANR STEM4PIGS (ANR-24-CE20-7792). A.D. is funded by the DIM-1HEALTH from Région Ile-de-France, the Animal Genetics Division of INRAE and the IFIP, Institut du porc. A.D. is also the recipient of an EMBO scientific exchange fellowship.

AUTHOR CONTRIBUTIONS

Following the CRediT taxonomy:

Conceptualization, H.A., J.A., A.D., and Sy.F.; data curation, A.D., H.A., Sy.F., and S.D.; formal analysis, A.D., H.A., Sy.F., and S.D.; funding acquisition, H.A. and M.-J.M.; investigation, A.D., M.-N.R., P.M., Y.B., Sy.F., J.A., and H.A.; methodology, A.D., H.A., Sy.F., and S.D.; project administration, H.A. and J.A.; resources, M.-N.R., P.M., Y.B., St.F., and A.T.; software, A.D., Sy.F., and S.D.; supervision: H.A. and J.A.; validation, A.D. and H.A.; visualization, A.D., H.A., and J.A.; writing – original draft, A.D., H.A., and J.A.; writing – review and editing, A.D., H.A., and J.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT and DeepL in order to improve the readability and language of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animals
- METHOD DETAILS
  - Preparation of single nuclei suspension
  - scMultiome library preparation and sequencing
  - ArchR gene and genome annotation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Single-cell multi-omics analysis on the embryos
  - Cluster assignment
  - Peaks assignation for genomic features
  - SCENIC+ analysis
  - Motif & footprints analysis
  - Functional enrichment

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.114519.

Received: June 23, 2025
Revised: October 21, 2025
Accepted: December 18, 2025
Published: December 23, 2025

REFERENCES

1. Artus, J., Hue, I., and Acloque, H. (2020). Preimplantation development in ungulates: a 'ménage à quatre' scenario. Reproduction *159*, R151–R172. https://doi.org/10.1530/REP-19-0348.
2. Griffith, B.P., Goerlich, C.E., Singh, A.K., Rothblatt, M., Lau, C.L., Shah, A., Lorber, M., Grazioli, A., Saharia, K.K., Hong, S.N., et al. (2022). Genetically Modified Porcine-to-Human Cardiac Xenotransplantation. N. Engl. J. Med. *387*, 35–44. https://doi.org/10.1056/NEJMoa2201422.
3. 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.
4. Ramos-Ibeas, P., Sang, F., Zhu, Q., Tang, W.W.C., Withey, S., Klisch, D., Wood, L., Loose, M., Surani, M.A., and Alberio, R. (2019). Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. Nat. Commun. *10*, 500. https://doi.org/10.1038/s41467-019-08387-8.
5. Dufour, A., Kurylo, C., Stöckl, J.B., Laloë, D., Bailly, Y., Manceau, P., Martins, F., Turhan, A.G., Ferchaud, S., Pain, B., et al. (2024). Cell specification

and functional interactions in the pig blastocyst inferred from single-cell transcriptomics and uterine fluids proteomics. Genomics *116*, 110780. https://doi.org/10.1016/j.ygeno.2023.110780.

6. Simpson, L., Strange, A., Klisch, D., Kraunsoe, S., Azami, T., Goszczynski, D., Le Minh, T., Planells, B., Holmes, N., Sang, F., et al. (2024). A single-cell atlas of pig gastrulation as a resource for comparative embryology. Nat. Commun. *15*, 5210. https://doi.org/10.1038/s41467-024-49407-6.
7. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
8. Argelaguet, R., Lohoff, T., Li, J.G., Nakhuda, A., Drage, D., Krueger, F., Velten, L., Clark, S.J., and Reik, W. (2022). Decoding gene regulation in the mouse embryo using single-cell multi-omics. Preprint at bioRxiv, https://doi.org/10.1101/2022.06.15.496239.
9. Janssens, J., Aibar, S., Taskiran, I.I., Ismail, J.N., Gomez, A.E., Aughey, G., Spanier, K.I., De Rop, F.V., González-Blas, C.B., Dionne, M., et al. (2022). Decoding gene regulation in the fly brain. Nature *601*, 630–636. https://doi.org/10.1038/s41586-021-04262-z.
10. Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2023). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat. Methods *20*, 1355–1367. https://doi.org/10.1038/s41592-023-01938-4.
11. Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell *6*, 468–478. https://doi.org/10.1016/j.stem.2010.03.015.
12. Acloque, H., Harrison, P.W., Lakhal, W., Martin, F., Archibald, A.L., Beinat, M., Davey, M., Djebali, S., Foissac, S., Guizard, S., et al. (2022). 550. Extensive functional genomics information from early developmental time points for pig and chicken. In Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP), R.F. Veerkamp and Y. de Haas, eds. (Wageningen Academic Publishers), pp. 2281–2284. https://doi.org/10.3920/978-90-8686-940-4_550.
13. Garces De Los Fayos Alonso, I., Liang, H.-C., Turner, S.D., Lagger, S., Merkel, O., and Kenner, L. (2018). The Role of Activator Protein-1 (AP-1) Family Members in CD30-Positive Lymphomas. Cancers *10*, 93. https://doi.org/10.3390/cancers10040093.
14. Li, M., Jiang, Z., Xu, X., Wu, X., Liu, Y., Chen, K., Liao, Y., Li, W., Wang, X., Guo, Y., et al. (2025). Chromatin accessibility landscape of mouse early embryos revealed by single-cell NanoATAC-seq2. Science *387*, eadp4319. https://doi.org/10.1126/science.adp4319.
15. Liu, L., Leng, L., Liu, C., Lu, C., Yuan, Y., Wu, L., Gong, F., Zhang, S., Wei, X., Wang, M., et al. (2019). An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. Nat. Commun. *10*, 364. https://doi.org/10.1038/s41467-018-08244-0.
16. Sebastiano, V., Dalvai, M., Gentile, L., Schubart, K., Sutter, J., Wu, G.-M., Tapia, N., Esch, D., Ju, J.-Y., Hübner, K., et al. (2010). Oct1 regulates trophoblast development during early mouse embryogenesis. Development *137*, 3551–3560. https://doi.org/10.1242/dev.047027.
17. Habekost, M., Jørgensen, A.L., Qvist, P., and Denham, M. (2019). Transcriptomic profiling of porcine pluripotency identifies species-specific reprogramming requirements for culturing iPSCs. Stem Cell Res. *41*, 101645. https://doi.org/10.1016/j.scr.2019.101645.
18. Chen, G., Schell, J.P., Benitez, J.A., Petropoulos, S., Yilmaz, M., Reinius, B., Alekseenko, Z., Shi, L., Hedlund, E., Lanner, F., et al. (2016). Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. Genome Res. *26*, 1342–1354. https://doi.org/10.1101/gr.201954.115.
19. Wallingford, M.C., Hiller, J., Zhang, K., and Mager, J. (2017). YY1 Is Required for Posttranscriptional Stability of SOX2 and OCT4 Proteins. Cell. Reprogram. *19*, 263–269. https://doi.org/10.1089/cell.2017.0002.
20. Wang, H., Xu, J., Lazarovici, P., Quirion, R., and Zheng, W. (2018). cAMP Response Element-Binding Protein (CREB): A Possible Signaling Molecule Link in the Pathophysiology of Schizophrenia. Front. Mol. Neurosci. *11*, 255. https://doi.org/10.3389/fnmol.2018.00255.
21. Geiselmann, A., Micouin, A., Vandormael-Pournin, S., Laville, V., Chervova, A., Mella, S., Navarro, P., and Cohen-Tannoudji, M. (2025). PI3K/AKT signalling orchestrates ICM maturation and proper epiblast and primitive endoderm specification. Dev. Cell *60*, 204–219.e6. https://doi.org/10.1016/j.devcel.2024.10.001.
22. Wang, N., Li, M.Y., Liu, Y., Yu, J., Ren, J., Zheng, Z., Wang, S., Yang, S., Yang, S.L., Liu, L.P., et al. (2020). ZBP-89 negatively regulates self-renewal of liver cancer stem cells via suppression of Notch1 signaling pathway. Cancer Lett. *472*, 70–80. https://doi.org/10.1016/j.canlet.2019.12.026.
23. Smukler, S.R., Runciman, S.B., Xu, S., and Van Der Kooy, D. (2006). Embryonic stem cells assume a primitive neural stem cell fate in the absence of extrinsic influences. J. Cell Biol. *172*, 79–90. https://doi.org/10.1083/jcb.200508085.
24. Uğurlu-Çimen, D., Odluyurt, D., Sevinç, K., Özkan-Küçük, N.E., Özçimen, B., Demirtaş, D., Enüstün, E., Aztekin, C., Philpott, M., Oppermann, U., et al. (2021). AF10 (MLLT10) prevents somatic cell reprogramming through regulation of DOT1L-mediated H3K79 methylation. Epigenetics Chromatin *14*, 32. https://doi.org/10.1186/s13072-021-00406-7.
25. Pursani, V., Bhartiya, D., Tanavde, V., Bashir, M., and Sampath, P. (2018). Transcriptional activator DOT1L putatively regulates human embryonic stem cell differentiation into the cardiac lineage. Stem Cell Res. Ther. *9*, 97. https://doi.org/10.1186/s13287-018-0810-8.
26. Barry, E.R., Krueger, W., Jakuba, C.M., Veilleux, E., Ambrosi, D.J., Nelson, C.E., and Rasmussen, T.P. (2009). ES Cell Cycle Progression and Differentiation Require the Action of the Histone Methyltransferase Dot1L. Stem Cell. *27*, 1538–1547. https://doi.org/10.1002/stem.86.
27. Wei, Y., Zhang, E., Yu, L., Ci, B., Sakurai, M., Guo, L., Zhang, X., Lin, S., Takii, S., Liu, L., et al. (2023). Dissecting embryonic and extraembryonic lineage crosstalk with stem cell co-culture. Cell *186*, 5859–5875.e24. https://doi.org/10.1016/j.cell.2023.11.008.
28. Artus, J., Piliszek, A., and Hadjantonakis, A.-K. (2011). The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. Dev. Biol. *350*, 393–404. https://doi.org/10.1016/j.ydbio.2010.12.007.
29. Uchikura, Y., Matsubara, K., Muto, Y., Matsubara, Y., Fujioka, T., Matsumoto, T., and Sugiyama, T. (2017). Extranuclear Translocation of High-Mobility Group A1 Reduces the Invasion of Extravillous Trophoblasts Involved in the Pathogenesis of Preeclampsia: New Aspect of High-Mobility Group A1. Reprod. Sci. *24*, 1630–1638. https://doi.org/10.1177/1933719117697254.
30. Ali, A., Stenglein, M.D., Spencer, T.E., Bouma, G.J., Anthony, R.V., and Winger, Q.A. (2020). Trophectoderm-Specific Knockdown of LIN28 Decreases Expression of Genes Necessary for Cell Proliferation and Reduces Elongation of Sheep Conceptus. IJMS *21*, 2549. https://doi.org/10.3390/ijms21072549.
31. Naama, M., Rahamim, M., Zayat, V., Sebban, S., Radwan, A., Orzech, D., Lasry, R., Ifrah, A., Jaber, M., Sabag, O., et al. (2023). Pluripotency-independent induction of human trophoblast stem cells from fibroblasts. Nat. Commun. *14*, 3359. https://doi.org/10.1038/s41467-023-39104-1.
32. Krendl, C., Shaposhnikov, D., Rishko, V., Ori, C., Ziegenhain, C., Sass, S., Simon, L., Müller, N.S., Straub, T., Brooks, K.E., et al. (2017). GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. Proc. Natl. Acad. Sci. USA *114*, E9579–E9588. https://doi.org/10.1073/pnas.1708341114.
33. Artus, J., and Hadjantonakis, A.-K. (2012). Troika of the Mouse Blastocyst: Lineage Segregation and Stem Cells. CSCR *7*, 78–91. https://doi.org/10.2174/157488812798483403.

34. Pérez-Gómez, A., González-Brusi, L., Flores-Borobia, I., Galiano-Cogolludo, B., Lamas-Toranzo, I., Hamze, J.G., Toledano-Díaz, A., Santiago-Moreno, J., Ramos-Ibeas, P., and Bermejo-Álvarez, P. (2024). The role of TEAD4 in trophectoderm commitment and development is not conserved in non-rodent mammals. Development *151*, dev202993. https://doi.org/10.1242/dev.202993.

35. Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech. Dev. *125*, 270–283. https://doi.org/10.1016/j.mod.2007.11.002.

36. Acloque, H., Ocaña, O.H., and Nieto, M.A. (2012). Mutual exclusion of transcription factors and cell behaviour in the definition of vertebrate embryonic territories. Curr. Opin. Genet. Dev. *22*, 308–314. https://doi.org/10.1016/j.gde.2012.03.004.

37. Kubota, K., Kent, L.N., Rumi, M.A.K., Roby, K.F., and Soares, M.J. (2015). Dynamic Regulation of AP-1 Transcriptional Complexes Directs Trophoblast Differentiation. Mol. Cell Biol. *35*, 3163–3177. https://doi.org/10.1128/MCB.00118-15.

38. Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, I.I., et al. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat. Genet. *50*, 1011–1020. https://doi.org/10.1038/s41588-018-0140-x.

39. Kotarba, G., Taracha-Wisniewska, A., Miller, M., Dabrowski, M., and Wilanowski, T. (2021). Transcription factors Krüppel-like factor 4 and paired box 5 regulate the expression of the Grainyhead-like genes. PLoS One *16*, e0257977. https://doi.org/10.1371/journal.pone.0257977.

40. Clark, E.L., Archibald, A.L., Daetwyler, H.D., Groenen, M.A.M., Harrison, P.W., Houston, R.D., Kühn, C., Lien, S., Macqueen, D.J., Reecy, J.M., et al. (2020). From FAANG to fork: application of highly annotated genomes to improve farmed animal production. Genome Biol. *21*, 285. https://doi.org/10.1186/s13059-020-02197-8.

41. Shen, W.-K., Chen, S.-Y., Gan, Z.-Q., Zhang, Y.-Z., Yue, T., Chen, M.-M., Xue, Y., Hu, H., and Guo, A.-Y. (2023). AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res. *51*, D39–D45. https://doi.org/10.1093/nar/gkac907.

42. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.

43. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

44. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

45. Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. *35*, W193–W200. https://doi.org/10.1093/nar/gkm226.

46. Ober-Reynolds, B., Wang, C., Ko, J.M., Rios, E.J., Aasi, S.Z., Davis, M.M., Oro, A.E., and Greenleaf, W.J. (2023). Integrated single-cell chromatin and transcriptomic analyses of human scalp identify gene-regulatory programs and critical cell types for hair and skin diseases. Nat. Genet. *55*, 1288–1300. https://doi.org/10.1038/s41588-023-01445-4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
INRAE_Plus4PigS_embryo_D7_omic_pool_1 pool of single cell specimens from D7 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811553
INRAE_Plus4PigS_embryo_D9_omic_pool_1 pool of single cell specimens from D9 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811554
INRAE_Plus4PigS_embryo_D9_omic_pool_2 pool of single cell specimens from D9 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811555
INRAE_Plus4PigS_embryo_D11_omic_pool_1 pool of single cell specimens from D11 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811556
INRAE_Plus4PigS_embryo_D11_omic_pool_2 pool of single cell specimens from D11 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811557
INRAE_Plus4PigS_embryo_single_cell_D11_3 single cell and nuclei preparation from D11 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811551
INRAE_Plus4PigS_embryo_single_cell_D11_4 single cell and nuclei preparation from D11 pig embryo	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA115811552
Chemicals, peptides, and recombinant proteins		
DMEM/F12 medium	Thermo Fisher	11320033
IMV Embryo holding media	IMV Technologies	019449
Accutase	Thermo Fisher	A1110501
SPRIselect DNA Size Selection Reagent 5 mL ()	Beckman Coulter	B23317
TrypLE	Thermo Fisher	12605010
Tween^{TM}20 Surfact-Amps^{TM}	Thermo Fisher	28320
Nonidet P40 Substitute	Merck	74385
BSA, acetylated (20 mg/mL)	Thermo Fisher	AM2614
DL-Dithiothréitol solution	Merck	646563
RNasin® Ribonuclease Inhibitor	Promega	N2515
Digitonin 5%	Thermo Fisher	BN2006
Regumate porcine 0.4% w/v Altrenogest	MSD Santé Animale	VM 06376/4097
Critical commercial assays		
Chromium Next GEM Chip J Single Cell Kit	10× Genomics	PN-1000230
Chromium Next GEM Single Cell Multiome Reagent Kit A, 4 rxns	10× Genomics	PN-1000284
Single Index Kit N Set A, 96 rxns	10× Genomics	PN-1000212
Dual Index Kit TT Set A, 96 rxns	10× Genomics	PN-1000215
Deposited data		
Sus scrofa reference genome 11.1 Ensembl 108 annotation	Ensembl	Ensembl: https://www.ensembl.org/Sus_scrofa/Info/Index
Raw single-cell OMICS (scRNAseq+ scATACseq)	This paper	ENA: PRJEB81663; ENA: PRJEB83269
Pig feather file for SCENIC+	This paper	Zenodo: https://doi.org/10.5281/zenodo.14833685

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pig embryo scRNAseq reference	Dufour et al.[5]	ENA: PRJEB60517
Aertslab v10 motif collection	González-Blas et al.[10]	https://resources.aertslab.org/cistarget/motif_collections/v10nr_clust_public/
AnimalTFDB v4.0	Shen et al.[41]	https://guolab.wchscu.cn/AnimalTFDB4/#/
Experimental models: Organisms/strains		
Sus Scrofa/Large White breed	BioSamples: https://www.ebi.ac.uk/biosamples/	BioSamples: SAMEA112465595 to BioSamples: SAMEA112465607
Software and algorithms		
Code used	This paper	GitHub: https://github.com/INSERM-U1141-Neurodiderot/pig_embryo_multiome
CellRanger ARC v2.0.0	10× Genomics	https://www.10xgenomics.com/support/software/cell-ranger-arc/latest
Archer v1.0.3	Granja et al.[42]	GitHub: https://github.com/GreenleafLab/ArchR
SuscrofaTxdb.11.108.july	This paper	Zenodo: https://doi.org/10.5281/zenodo.10844556
Org.Ss.e.g.,.db v3.20.0	Marc Carlson	Bioconductor: https://www.bioconductor.org/packages/org.Ss.eg.db/
Harmony v1.2.0	Korsunsky et al.[43]	GitHub: https://github.com/immunogenomics/harmony
MACS v2.2.7.1	Zhang et al.[44]	GitHub: https://github.com/macs3-project/MACS
Gprofiler2	Reimand et al.[45]	https://cran.r-project.org/web/packages/gprofiler2/index.html
SCENIC+	González-Blas et al.[10]	GitHub: https://github.com/aertslab/scenicplus; Zenodo: https://doi.org/10.5281/zenodo.14833685
TopGO v2.58	Alexa Rahnenführer	Bioconductor: https://bioconductor.org/packages/release/bioc/html/topGO.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

Embryo production

Pig embryos (Sus scrofa, Large White breed) were sampled at the hatched blastocyst stage (7 days post-insemination), at the early ovoid blastocyst stage (9 days post-insemination) and at the late ovoid blastocyst stage (11 days post-insemination). All information on animal samples is summarised in Table S1.

Large White sows used for embryo production were reared and slaughtered according to conventional breeding procedures, at INRAE facilities approved by the French Veterinary Services. The pig embryos analyzed in this study were collected postmortem, at a stage prior to the final third of gestation, and therefore fall outside the scope of European Union Directive 2010/63/EU on the protection of animals used for scientific purposes.

All embryos were produced at the INRAE experimental unit GenESI (Rouillé, France). Metadata related to the biological samples used in this study have been submitted to FAANG Data Portal (https://data.faang.org/home) and BioSamples:https://www.ebi.ac.uk/biosamples/, and are summarised in Table S1.

The estrous cycle of each sow was synchronised using Altrenogest (Regumate), a synthetic progestin, for 18 days. The day after, sows were artificially inseminated, and the insemination was repeated the following day. When the gestational time matched the embryonic stage to be sampled (7, 9, and 11 days post-insemination), the sows were transported from the breeding unit (Rouillé, France) to the slaughterhouse (Nouzilly, France). They were stunned by electronarcosis and bled. The uterus was clamped and rapidly extracted from the abdominal cavity. Then, embryos were collected into two 50 mL tubes by retro-flushing of the uterine horns from the bottom of the uterine horn upwards (ovary) in 100 mL of physiological saline solution. The detailed protocol is publicly accessible on the FAANG Data Portal: https://api.faang.org/files/protocols/samples/INRAE_SOP_PLUS4PIGS_EMBRYOS_SAMPLING_PROTO2_20230131.pdf.

METHOD DETAILS

Preparation of single nuclei suspension

Once recovered, embryos were staged, pooled, and transported in DMEM/F12 medium (E9 and E11) or IMV Embryo holding media (E7) to the molecular biology laboratory in a thermostatically controlled chamber at 38°C. Upon arrival, embryos were transferred to a 4-well dish and staged again under a stereomicroscope. When necessary, embryos of the same stage were pooled together into a drop of DMEM/F12 or IMV Embryo holding media and processed for cell dissociation. The full protocol is publicly accessible on the FAANG Data Portal: https://api.faang.org/files/protocols/samples/INRAE_SOP_PLUS4PIGS_EMBRYOS_DISSOCIATION_PROTO3_20230131.pdf. All embryos (E7 to E11) were incubated in pre-warmed Accutase for 10 min, then in pre-warmed TrypLE for 10 min, followed by mechanical dissociation by repeated pipette aspiration.

Nuclei were isolated following the 10× Genomics protocol: CG000365 · Rev A Nuclei Isolation for Single Cell Multiome. Nuclei for embryonic cells were isolated following the same protocol from 10× Genomics, using the dedicated protocol for low cell input nuclei isolation (Appendix section). For E9 and E11 embryos, around 40,000 cells were used for each preparation. For E7 embryos, all cells (~7,000) were used for one reaction. Each single cell suspension was washed twice with 50 μL of ice-cold PBS containing 0.04% BSA and placed in a 0.2 mL tube. Cells were centrifuged for 1 min at full speed using a minifuge. After the second wash, the supernatant was removed and 45 μL of ice-cold lysis buffer was added (10 mM Tris-HCl pH7.4; 10 mM NaCl; 3 mM MgCl2; 0.1% Tween 20; 0.1% Nonidet P40; 0.5% BSA; 1 mM DTT; 0.5U/μL RNase inhibitor; 0.01% digitonin). Cells were resuspended by pipetting up and down 3 times and were incubated on ice for exactly 3′30. Then, 50 μL of ice-cold Wash Buffer (10 mM Tris-HCl pH7.4; 10 mM NaCl; 3 mM MgCl2; 0.1% Tween 20; 0.5% BSA; 1 mM DTT; 0.5U/μL RNase inhibitor) was added carefully to the tube without disrupting the pellet. Cells were again centrifuged 1 min at full speed using a minifuge. 95 μL of supernatant was then removed without disrupting the nuclei pellet and 95 μL of chilled Diluted Nuclei Buffer was carefully added to the tube. Cells were again centrifuged 1 min at full speed using the labtop minifuge. The supernatant was removed without touching the bottom of the tube to avoid dislodging the nuclei pellet. The nuclei pellet was then resuspended in 7 μL of chilled Diluted Nuclei Buffer.

To quantify nuclei, 2 μL of nuclei suspension were mixed with 8 μL of Diluted Nuclei Buffer and 10 μL of Trypan Blue and counted using a Malassez cell counting chamber. Around 5,000 nuclei/μL have been recovered from each prep. 25,000 nuclei resuspended in 5 μL of Nuclei Dilution Buffer are processed for transposition according to: ChromiumNextGEM_Multiome_ATAC_GEX_User_Guide_RevB.

The full protocol is publicly accessible on the FAANG Data Portal: https://api.faang.org/files/protocols/samples/INRAE_SOP_PLUS4PIGS_.pdf.

scMultiome library preparation and sequencing

Libraries were sequenced on an Illumina NovaSeq6000 and subsequently aligned to the *Sus scrofa* genome (version 11.1). Gene positions were annotated according to Ensembl build 108 and filtered based on biotype (only protein-coding, long intergenic non-coding RNA, antisense, immunoglobulin or T cell receptor). Following this, CellRanger ARC (version 2.0.0) was employed to generate the count matrices based on this annotation.

ArchR gene and genome annotation

The data were processed using the ArchR R package (1.0.3).[42] Gene annotation parameters were developed using the GTF files produced earlier and the Org.Ss.e.g.,.db package. Genome annotation was conducted similarly, applying filters to exclude contigs that were not detected in either the ATAC-seq or the RNA-seq datasets (Zenodo: https://doi.org/10.5281/zenodo.10844556), while also including the mitochondrial chromosome. Due to the presence of short contigs, the downstream and upstream extensions for the gene score matrix function were set to a minimum of 0 and a maximum of 100,000.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-cell multi-omics analysis on the embryos

A quality control was applied and cells were excluded if they had more than 1e+10 or fewer than 1,000 fragments and a minimum enrichment of Transcription Start Site (TSS) of 4. Cells were retained only if they were present in both ATAC-seq and RNA-seq assays. Subsequently, we filtered out cell doublets using ArchR functions addDoubletScores and filterDoublets with default parameters, resulting in a total of 12,778 cells (Table S2). After filtering, we generated Latent Semantic Index (LSI) for both assays.

For ATAC-seq, we used the TileMatrix with the following parameters: a resolution of 2, a sample size of 30,000 cells, n.start set to 10 and the number of fragments as coverage. For RNA-seq, we used the GeneExpressionMatrix with the following parameters: a resolution of 2, a sample size of 30,000 cells, n.start set to 10 and the number of fragments as coverage. We used the number of Unique Molecular Identifiers (UMI) as coverage, selected the 2,500 most variable genes and deactivated the binarization. A new dimensional reduction was then created using the two previously generated LSI with the addCombinedDims function.

Batch corrections for each sample were applied using the Harmony implementation in ArchR on the combined dimensional reduction. A UMAP was generated on the corrected dimensions reduction using 15 neighbors, a minimal distance of 0.8 and a cosine metric. Then a first clustering was conducted at a resolution of 0.1 (Figures 1B and 1C).

Group coverage and peak identification were conducted using ArchR function addGroupCoverage and addReproductiblePeakSet, alongside MACS (v2.2.7.1)[44] with a genome size of 1,341,049,888.

Peak enrichment analyses were performed using the getMarkerFeatures function, utilizing the wilcoxon method along with correction for TSSEnrichment and log10 bias. Additionally, connections between peak and gene expression were established using the addPeak2GeneLinks function in ArchR, applying Harmony[43] for dimensionality reduction and the gene expression matrix. The resulting data was incorporated into the differential peak table (Table S6) when significant correlations were identified.

Cluster assignment

To perform embryo assignment (Figure S2), we used our previous scRNA-seq data from pig embryos (ENA: PRJEB60517).[5] Cell mapping was conducted using ArchR addGeneIntegrationMatrix function with the following parameters: dimensional reduction was carried out using Harmony, cell sample size of 50,000, number of dimensions set to 60 (utilizing all available dimensions), 4,000 variables genes, the cca reduction. Subsequently, we applied the FindAnchorFunction from ArchR with the following parameters: k.anchor = 80, k.filter = 1600, k.score = 200, max.features = 800, and n.trees = 800. Based on these results, we confidently identified clusters of epiblast and trophectoderm. Additionally, we distinguished two clusters of hypoblasts (early and late), though some cells within the late hypoblast cluster displayed a lower confidence score. All identified clusters expressed specific gene markers indicative of their respective populations.[3,4]

Peaks assignation for genomic features

Promoter's maps were generated using ArchR classification of peaks. Peaks identified with ArchR were classified into four categories: introns, exons, distal and promoter for the peak situated within 2,000 bp upstream and 100 bp downstream of the TSS site.

SCENIC+ analysis

To run SCENIC+[10] on the pig genome, we created a custom database. Transcription factor lists were generated following the methodology used for AnimalTFDB 4.0,[41] utilizing the pig Ensembl build 108. The motif-to-transcription factor annotation was adapted to pig by leveraging genes orthology through gprofiler2.[45] The motif database was constructed using scripts from aertslab (github.com/aertslab/create_cisTarget_databases) focusing on the best transcript score for each region of the genomes based on the 10 kb upstream regions.

The pipeline was executed on embryo using a combination of clusters and stages as cell aggregations, excluding categories with fewer than 20 cells (specifically, Cluster Hypo Ea at E7 and E9, and Cluster HYPO La at E7), applying default parameters and adapting annotations for the pig genome. For CisTopic analysis, we retained 16 topics. Following the SCENIC+ results, chromatin and transcriptome scores were calculated using the score_eRegulons function with an auc_threshold set to 0.05. We filtered the resulting regulons by rejecting extended regulons if direct regulons already existed, excluding regulons with fewer than 10 genes and those where region accessibility correlated with transcriptomics. TF-to-gene (TF2G) scores were calculated using SCENIC+. It is based on the Arboreto Python package (v.0.1.6) using gradient-boosting machine regression by predicting raw TF expression from raw gene expression counts and using the importance score of each feature (gene) as the TF-to-gene importance score.[10]

Motif & footprints analysis

Motif analysis was conducted using the aertslab v10 motif collection.[10] Motif annotations were incorporated into the ArchrProject through the addMotifAnnotations function, with parameters set to a cutoff of 5e-05, a width of 7 and utilizing version 2. Motif enrichment was subsequently performed using the peakAnnoEnrichment function from ArchR, with a filter applied for a minimal false discovery rate (FDR) of 0.1 and a log-fold greater than 0.5. The background peak and Deviations Matrix were computed using the addDeviationsMatrix and addBgdPeaks functions from ArchR, employing default parameters.

After obtaining the motif enrichment results, the motif positions were acquired using ArchR getPositions function with default settings. Footprints were generated using the ArchR getFootprints function, and these footprints were visualized through the plotFootpints function, utilizing "Substract" normalization method and a smoothing window that corresponds to the motif length.

Functional enrichment

Functional enrichment of gene expression between clusters (Figure 2D) was assessed using marker genes from RNA-seq, which were obtained via the ArchR getMarkerFeatures function on the GeneExpressionMatrix. The Wilcoxon test was employed, along with a correction for the number of UMIs. Following this, an enrichment analysis was conducted using the gprofiler2 package for all identified genes, focusing on the *Sus scrofa* organism and applying default parameters to the 2,000 most enriched genes.

Similarly, functional enrichment of peak differential accessibility between clusters was determined by linking genes to the corresponding peaks or if no gene was associated, using the nearest gene in the genome. This was also achieved using ArchR getMarkerFeatures on GeneExpressionMatrix using the Wilcoxon test with a correction for the number of UMIs. The resulting gene table underwent an enrichment analysis with the gprofiler2 R package for the *Sus scrofa* organism, using default parameters on the 2,000 most enriched genes.

For peak-to-gene heatmap functional annotations (Figure 5), we implemented methods outlined in GreenLeaf et al.[46] We selected the 1,000 most frequent genes for each k-means cluster, then gene ontology (GO) enrichment was calculated using the TopGO package (Bioconductor: https://bioconductor.org/packages/topGO). The runTest function was applied using GO with specific parameters: Biological Process, mapping to the *Sus scrofa* database via symbol gene names, a minimal threshold of 5 genes, the weight01 algorithm and the Fisher test. Subsequently, the GenTable function was utilized to select the 50 top GO terms.