

SCIENTIFIC REPORTS



OPEN

Long non-coding RNAs involved in the regulatory network during porcine pre-implantation embryonic development and iPSC induction

Liang Zhong^{1,2}, Haiyuan Mu², Bingqiang Wen², Wei Zhang², Qingqing Wei², Ge Gao³, Jianyong Han²  & Suying Cao^{1,4}

Long non-coding RNAs (lncRNA) play a key role in the orchestration of transcriptional regulation during development and many other cellular processes. The importance of the regulatory co-expression network was highlighted in the identification of the mechanism of these processes in humans and mice. However, elucidation of the properties of porcine lncRNAs involved in the regulatory network during pre-implantation embryonic development and fibroblast reprogramming to induced pluripotent stem cell (iPSC) has been limited to date. Using a weighted gene co-expression network analysis, we constructed the regulatory network and determined that the novel lncRNAs were functionally involved in key events of embryonic development during the pre-implantation period; moreover, reprogramming could be delineated by a small number of potentially functional modules of co-expressed genes. These findings indicate that lncRNAs may be involved in the transcriptional regulation of zygotic genome activation, first lineage segregation and somatic reprogramming to pluripotency. Furthermore, we performed a conservation and synteny analysis with the significant lncRNAs involved in these vital events and validated the results via experimental assays. In summary, the current findings provide a valuable resource to dissect the protein coding gene and lncRNA regulatory networks that underlie the progressive development of embryos and somatic reprogramming.

Long non-coding RNA (lncRNA) is a type of non-protein coding transcript longer than 200 nucleotides that represents a large and diverse class of non-coding RNA. The advances of past several years in RNA sequencing (RNA-Seq) and computational methodology have enabled an unprecedented analysis of these transcripts. Genome-wide analyses have identified more than 9,000 human long intergenic noncoding RNAs (lincRNA)^{1–4} and more than 10,000 long noncoding transcripts in mice^{5,6}. Recently, thousands of lncRNAs have been identified in other model organisms as well as domesticated animals, such as *C. elegans*⁷, *Drosophila*⁸, zebrafish^{9,10}, chicken¹¹, cattle^{12–14} and pigs^{15–18}. However, only a few lncRNAs have been functionally annotated, and the majority remains to be characterized. Several lncRNA databases have been created to provide an integrative annotation of lncRNAs^{19–23}.

The pig is not only a commercial domesticated animal but also an ideal experimental animal for medical and preclinical models that shares similar characteristics with humans in physiology, anatomy and the occurrence of disease²⁴. While the other model animals, such as mice, cannot exactly be used to evaluate the long-term effects of cell differentiation or cell therapy, and human research may result in ethical issues in the developmental process^{25,26}. The difference in pre-implantation development in mice and humans has made pigs invaluable for

¹The Animal Science and Technology College, Beijing University of Agriculture, Beijing, China. ²State Key Laboratory for Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China. ³State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Center for Bioinformatics, Peking University, Beijing, China. ⁴Beijing Key Laboratory of Traditional Chinese Veterinary Medicine, Beijing University of Agriculture, Beijing, China. Correspondence and requests for materials should be addressed to J.H. (email: hanjy@cau.edu.cn) or S.C. (email: 20137602@bua.edu.cn)

basic investigations in which it may serve as an excellent model system to investigate developmental biology, reprogramming and embryonic cell fate commitment. However, to date, no authentic porcine embryonic stem cells have been derived; thus, a comprehensive analysis of pre-implantation embryos and induced pluripotent stem cells (iPSCs) would facilitate the understanding of the mechanism of pluripotency in pigs. Moreover, iPSCs provide tremendous opportunities for regenerative medicine. Many of the molecular mechanisms that underlie tumorigenesis and self-renewal in iPSCs have been elucidated in recent years.

Embryogenesis is the process in which an embryo forms and develops, and it is initiated by the fertilization of the oocyte with a sperm cell. The zygote subsequently undergoes several rapid rounds of mitotic division without significant growth (a process referred to as the cleavage stage) and cellular differentiation, which lead to the development of a multicellular embryo. Interestingly, porcine transcriptome data have indicated that zygotic genome activation (ZGA) occurs during the period from 4-cell stage to 8-cell stage, and there are shared molecular determinants and pathways among pigs, mice and humans involved in first lineage segregation and primitive endoderm differentiation²⁷. However, little is known regarding the potential signaling elements that constitute the regulatory network involved in porcine pre-implantation embryo development, especially the interaction with non-coding RNA during this vital period.

An investigation of the expression patterns of lncRNAs and mRNA at a particular cell-stage may identify novel functional connections and expand the mechanistic knowledge of the developmental process. In this study, we report the genome-wide characterization of iPSCs and all stage-embryonic lncRNAs and define a stringent set of 207 (563 transcripts) novel lncRNA genes from RNA-seq data of porcine embryos. We validated our data set using genomic features of lncRNA, including transcript length, exon number, evolutionary conservation and spatiotemporal expression specificity. A weighted gene co-expression network analysis indicated that lncRNAs are expressed in a strong developmental stage-specific manner, and many lncRNAs are associated with developmental regulatory genes. Furthermore, to derive authentic porcine embryonic stem cells, an intrinsic issue was whether the lncRNAs of piPSCs are regulated by identical or comparable mechanisms in the blastocyst stage. Our genome-wide annotation of embryonic lncRNAs may improve our understanding of the molecular mechanisms that underlie porcine embryogenesis as well as reprogramming and provide several regulatory networks during developmental process.

Results

Pipeline of lncRNA identification. To identify potentially functional lncRNAs and the regulatory network involved in porcine pre-implantation embryos and iPSCs, we first aligned clean reads to the porcine genome (susScr2) using TopHat and assembled them subsequently into 621,085 transcripts to form 227,625 loci in all samples using Cufflinks, which include 27 RNA-seq samples composing of oocyte, 7 consecutive embryonic stages and iPSC. Each of the cell type has biological replicates. Besides, porcine fibroblasts have been sequenced to compare with iPSC. To obtain high-quality novel transcripts, we used Ensembl, NCBI RefSeq and UCSC gene annotations to eliminate known genes and removed the lowly expressed transcripts. The expression level of the lncRNAs was relatively low; therefore, a cutoff threshold was defined for the novel transcripts. According to the peak of Gaussian distribution at the kernel density estimation, we selected transcripts to cut off the FPKM (fragments per kilobase of exon per million fragments mapped) of all samples (Supplementary Fig. 1). Then we obtained the highly confidential transcript which could be used to identify lncRNA with a stringent pipeline. Thus, the remaining transcripts were defined as novel lncRNAs and were used for the downstream functional analysis (Fig. 1).

Transcriptome profiling of porcine pre-implantation embryos and iPSCs. We used porcine oocytes, *in vivo* embryos, somatic cell nuclear transfer (SCNT) embryos, fibroblasts, and iPSCs to conduct RNA-seq. iPSCs were generated from fibroblasts and passaged 20 times. The embryos were at several critical stages of pre-implantation development, namely, zygote, 2-cell, 4-cell, 8-cell, morula and blastocyst. To elucidate the differences between the inner cell mass (ICM) and trophectoderm (TE), we physically segregated the TE and ICM. All embryo samples were collected according to stringent morphological criteria and sequenced.

First, we analyzed the known coding genes in each sample. Approximately 12,000 to 14,000 known Ensembl genes were detected (FPKM > 0.1) out of 20,459 Ensembl genes (release-66). On average, the detected expression genes represented 14,275 (69.7%); thus, more than half of the known porcine transcripts were expressed, which suggests that our transcriptomes were reliable for different downstream analysis (Supplementary Fig. 2).

To investigate the relationships among the developmental stages, we performed hierarchical clustering and a principal component analysis (PCA) of the known genes individual embryonic stages and iPSCs (Fig. 2A,B). A distinct cluster formed as a result of different pre-implantation embryonic stages and cell types. There were 2 clusters that were mainly separated by the 8-cell stage, a group of earlier than the 8-cell stage samples in which oocyte, zygote and 2-cell stages were more similar to each other and a second group of 4-cell and 8-cell stage samples. The consecutive stages agglomerated together in the analysis as expected, and from 4-cell stage to 8-cell stage embryos were separated from early samples as a result of degradation of maternal mRNA after fertilization and ZGA between the 4-cell and 8-cell stages, which has been previously reported²⁷. During the morula and blastocyst stages, blastomeres began to differentiate, which resulted in the first cell fate determination in embryogenesis, and progressively segregated into the ICM and TE. Furthermore, piPSCs were similar to porcine blastocysts, which could be explained by the pluripotency during the lineage segregation, and a small number of blastomeres would develop into the pluripotent ICM (Supplementary Fig. 3). However, another notable difference occurred between the *in vivo* embryos and SCNT embryos; there was a lag during the embryonic development to the 8-cell stage, particularly at the 4-cell and 8-cell stages, because the reconstruction of the alternated gene expression of the SCNT embryos has a critical impact on embryogenesis²⁸.

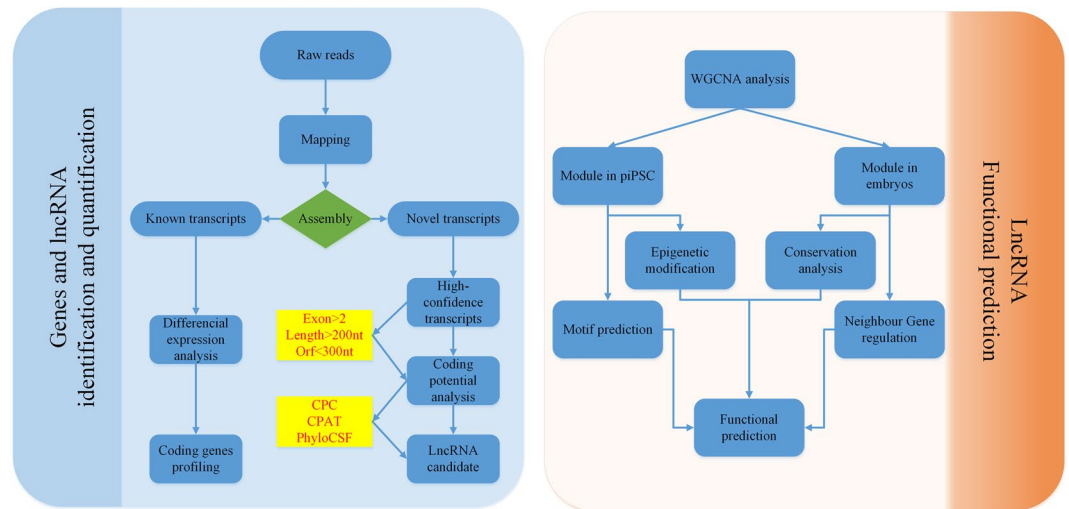


Figure 1. Pipeline for the identification of novel potentially functional lncRNAs in porcine pre-implantation embryos and iPSCs. The left figure is the pipeline for identification of lncRNAs from transcriptomes of porcine pre-implantation embryos and iPSCs. Transcripts were classified as known protein-coding genes and novel transcripts. The novel high-quality transcripts were subsequently filtered by sequence feature, coding potential, phylogenetic conservation as well as dissimilarity to known protein domains, and the remaining transcripts were defined as lncRNAs. The right figure is the pipeline for investigation of functional lncRNAs. To identify potentially functional lncRNAs during porcine embryo development and reprogramming to pluripotency, a co-expression network was constructed with protein coding genes and lncRNAs by WGCNA. The lncRNAs in modules that correlated with important embryonic stage and cell type were chosen to perform epigenetic modification, conservation analysis and experimental verification. Finally, the potentially functional lncRNAs in specific regulatory co-expression network were elucidated.

The results that the same embryonic stage exhibits a greater similarity than between different stages were also supported by the PCA plot, which indicates that the embryos in a given developmental stage were gathered closely and the continuous stage also presented the trends for embryogenesis. Intriguingly, the samples of 4-cell and 8-cell stages were dispersed in the PCA plot though they were relatively similar in the hierarchical cluster results, which suggest that complicated regulation emerged during that period. Porcine ZGA occurs at the 4-cell stage, and as many as 1791 genes exhibited differential expression between the 4-cell stage and 8-cell stage. Therefore, we suggest that the 4-cell and 8-cell stages play an important role in the course of pre-implantation embryogenesis.

Genomic features and stage-specific expression of lncRNAs. Previous studies have shown that the expression levels of lncRNAs occur in a cell-type-specific manner and are significantly lower than protein coding transcripts; moreover, lncRNAs were shorter and less conserved than protein coding transcripts^{1,6}. Those reports have suggested that lncRNAs have temporal- and spatial-specific expression, which plays important roles during embryogenesis¹⁰. However, lncRNAs and their biologically relevant processes in porcine embryonic development are poorly understood. Deciphering the repertoire and expression profiles of lncRNAs, particularly the temporal and spatial expression patterns, would provide a vital step towards a better understanding of the gene regulatory network within the developmental processes. We performed a hierarchical clustering analysis for lncRNAs (Fig. 3A); however, in contrast to the protein coding genes, the dendrogram of embryonic samples was mainly classified into 2 groups, which were separated by the 4-cell to 8-cell stage. Potentially as a result of ZGA, the 4-cell to 8-cell stage embryos were more inclined to vary in the expression of protein coding genes and lncRNAs.

To identify the stage-specific expression of lncRNAs, we performed differential expression analysis for pairwise comparison of all 10 type of samples. Furthermore, we calculated the temporal specificity scores using the Jensen-Shannon distance method (Fig. 3B). The 4-cell and 8-cell stage embryos were the most variable in the expression of lncRNAs, and there were 17 up-regulated lncRNAs (\log_2FC of 8-cell versus 4-cell > 1 , adjusted- $P < 0.05$) in contrast to only 7 down-regulated lncRNAs (\log_2FC of 8-cell versus 4-cell < -1 , adjusted- $P < 0.05$) from the 4-cell to 8-cell stage. Interestingly, we obtained the opposite results in the expression patterns of the protein coding genes (Supplementary Tables S3 and S4); the 4-cell and 8-cell stage embryos had more down-regulated differentially expressed genes (963 genes, $\log_2FC < -1$ and adjusted- $P < 0.05$) than up-regulated genes (828 genes, $\log_2FC > 1$ and adjusted- $P < 0.05$), which suggests that the degradation of maternal transcripts and zygotic genome activation tend to occur in parallel within developmental activities.

In our study, we found that lncRNA show less conservation than mRNA even in randomly selected genomic background, which display with curve of accumulative PhastCons scores in Fig. 3C. We also determined that the porcine novel lncRNAs exhibit a fewer number of exons and are shorter in sequence length compared with protein coding genes. As indicated in Fig. 3D,E, the distribution of the transcript exon number and length indicates that mRNA has 7.5 exons and 1378 nucleotides in length on average, whereas lncRNA has 3.3 exons and 1157 nucleotides in length on average; these findings indicate a significant difference between mRNA and

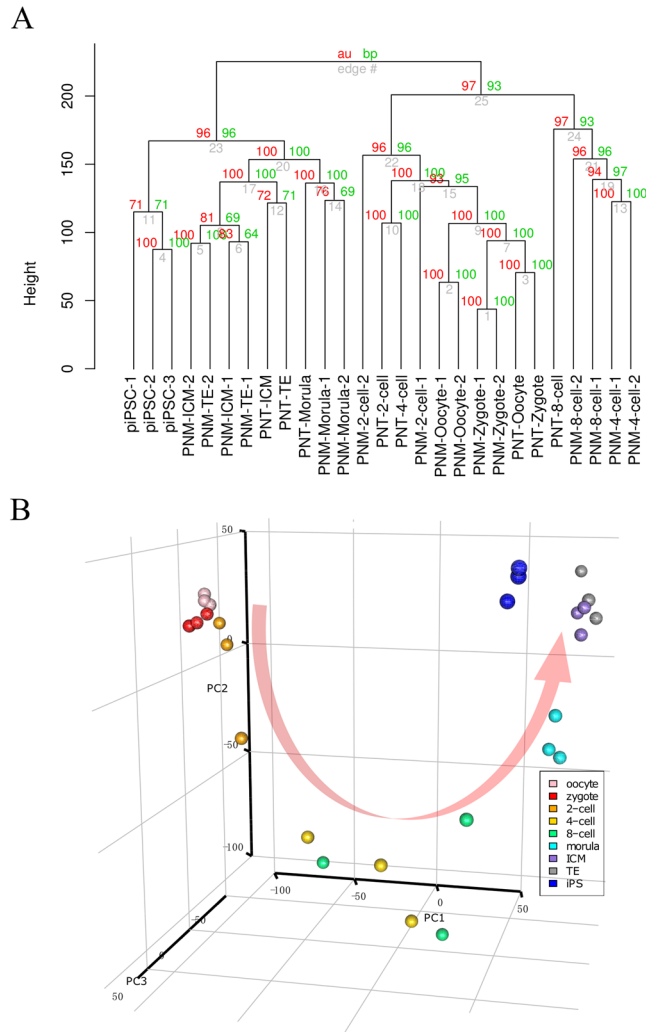


Figure 2. Global profiling of the porcine pre-implantation embryo and iPSC transcriptome. **(A)** Hierarchical cluster analysis of known genes. WPGMA (Weighted Pair Group Method with Averaging) is used as the agglomeration method. AU (Approximately Unbiased) in red color represents p-value and BP (Bootstrap Probability) value in green color. AU p-value is computed by multiscale bootstrap resampling method. The prefixes PNM and PNT of embryonic samples represent porcine *in vivo* embryos and porcine somatic cell nuclear transfer (SCNT) embryo samples, respectively. **(B)** Principal component analysis (PCA) of the porcine pre-implantation embryo and iPSC transcriptome. PC1, PC2 and PC3 represent the top three dimensions of the contribution of gene expression among these samples, which account for 48.9%, 10.4% and 8.0%, respectively.

lncRNAs (exon number: 7.5 versus 3.3, respectively, $P = 3.798 \times 10^{-29}$; length: 1378 versus 1157, respectively, $P = 1.833 \times 10^{-8}$, Welch two-sample t-test). Interestingly, the porcine lncRNAs were similar to the genomic features of human (approximately 1000 nt in length and 2.9 exons on average), mouse (approximately 1300 nt in length and 2.3 exons on average) and zebrafish (approximately 1113 nt in length and 2.8 exons on average) lncRNAs according to the NONCODEv4 database²³. The similarity of the characterization of multi-species lncRNA suggests the porcine lncRNA results were close to those in other model organisms.

Construction of a regulatory network using WGCNA. We subsequently applied a weighted gene co-expression network analysis (WGCNA)²⁹ to investigate the co-expression relationships between the protein coding genes and lncRNAs that were associated with pre-implantation embryonic development and somatic cell reprogramming to pluripotency at a system level. WGCNA identified 25 significant discrete modules of co-expressed genes and lncRNAs (Supplementary Figs 4–6). We focused on 11 of 25 modules that were highly correlated ($r > 0.6$, $P < 0.05$) and associated with embryo development as well as stem cell maintenance in specific stages. Most notably, the “darkred” ($n = 98$, $r = 0.82$, $P = 2 \times 10^{-7}$) and “purple” ($n = 184$, $r = 0.88$, $P = 2 \times 10^{-9}$) modules were involved in the transition from oocyte to zygote, and the “blue” ($n = 713$, $r = 0.95$, $P = 8 \times 10^{-14}$) module was involved in reprogramming to pluripotency. However, for the remaining significant modules that participated in the pre-implantation embryo development, we determined that several modules showed coordinated changes across several successive stages. These findings suggested that the identified coordinated modules were inclined to relate specifically to the developmental events of pre-implantation, such as ZGA and lineage segregation.

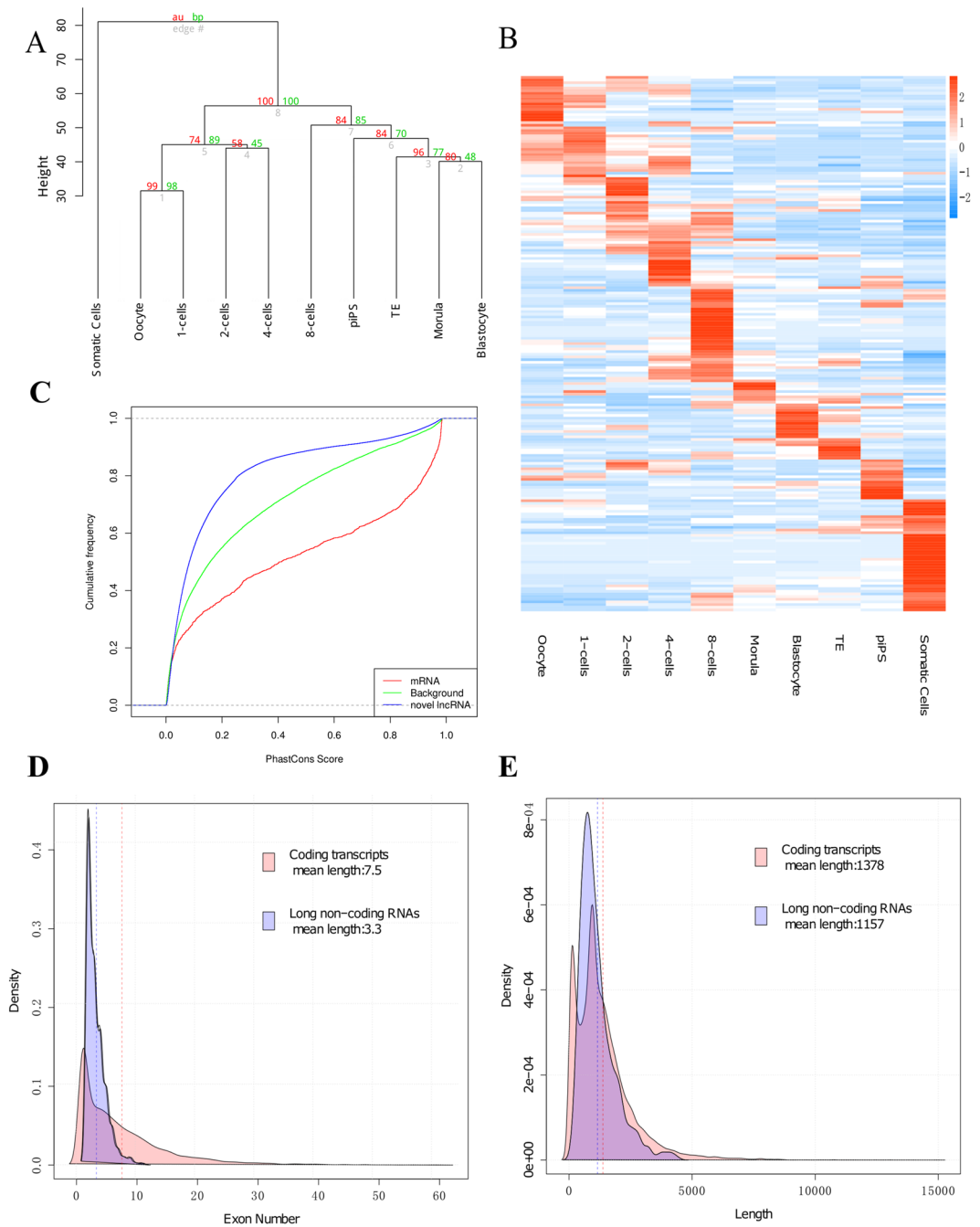


Figure 3. Genomic features and stage-specific expression of lncRNAs. **(A)** Hierarchical cluster analysis of novel long non-coding RNAs. WPGMA is used as the agglomeration method. AU (Approximately Unbiased) in red color represents p-value and BP (Bootstrap Probability) value in green color. AU p-value is computed by multiscale bootstrap resampling method. We combined all samples in the same stage for analysis. **(B)** Heatmap of stage-specific expression of the novel lncRNAs, which was calculated by the Jensen-Shannon distance. **(C)** The cumulative curve of the average phastCons score of the novel lncRNAs (blue), mRNA (red) and the genomic background (green). Genomic background was randomly selected from the whole genome. **(D)** and **(E)** Length and exon number distribution between porcine lncRNAs and protein coding genes; the blue portion represents the lncRNAs, and the pink portion represents the protein coding genes.

Regulatory network of zygotic genome activation during 4-cell and 8-cell stages. To verify whether the identified modules were biologically functional during the processes of pre-implantation embryonic development, we initially performed an enrichment analysis on the two stage using gene ontology (GO) and KEGG pathways. The lightcyan module was specifically and positively related in the two stages compared to the other modules that were significant in the 4-cell or 8-cell stage, although this module particularly showed a strong correlation and significance in the 8-cell ($r=0.71$, $P=3 \times 10^{-5}$) more than the 4-cell ($r=0.39$, $P=0.04$)

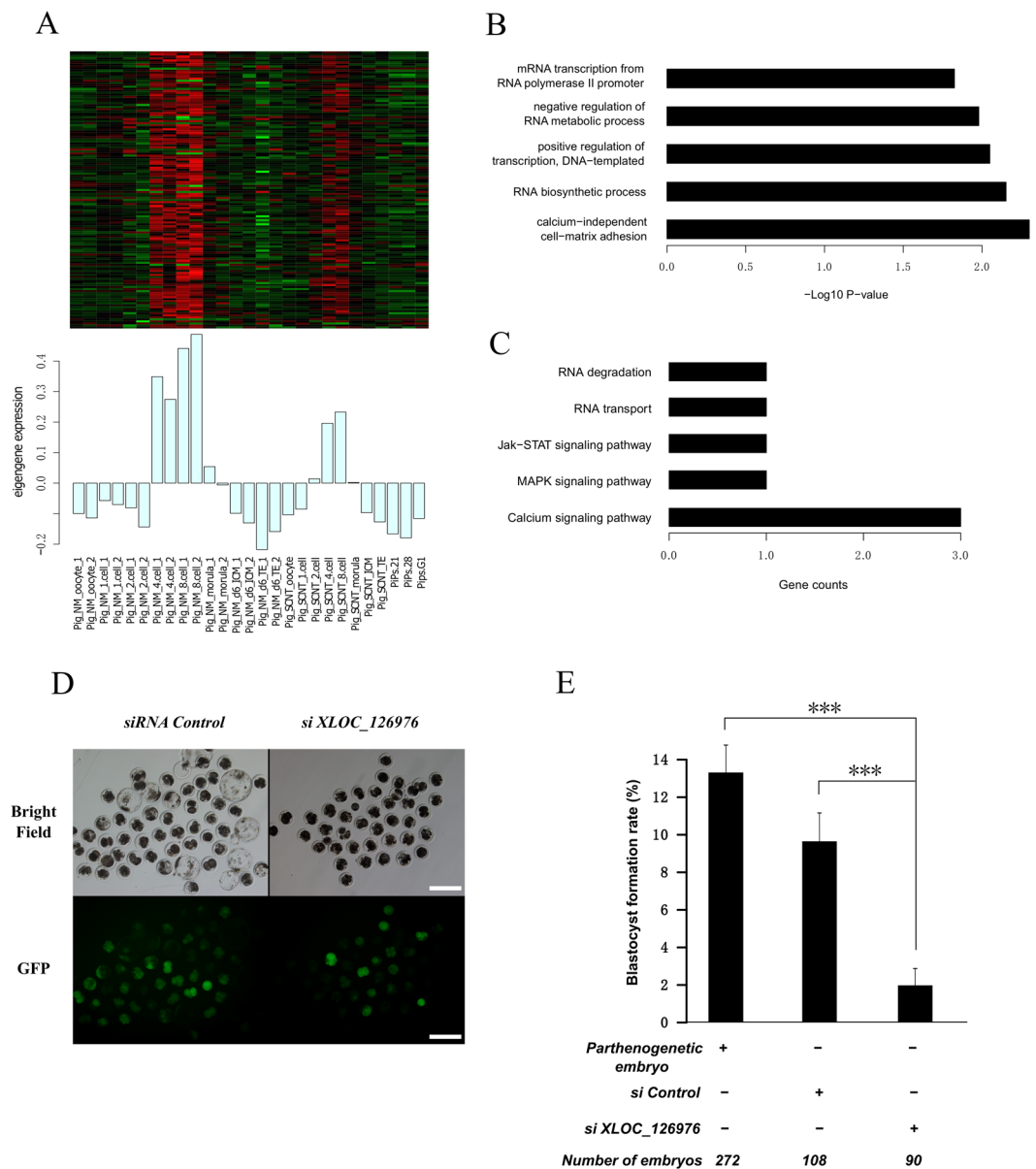


Figure 4. Co-expression network during ZGA stage. **(A)** Heatmap and bar plot of module lightcyan for the zygotic genome activation. The heatmap indicates the expression levels of genes in this module; all data were normalized with the z-score, and the bar plot indicates the module eigengene relative expression level. **(B)** Gene ontology enrichment for this module. **(C)** KEGG pathway enrichment for this module. **(D)** and **(E)** siRNA injection and the blastocyst formation rate. **(D)** The development state of the blastocyst after the siRNA of XLOC_126976 and the negative control were injected into parthenogenetically activated embryos. **(E)** The statistics for the blastocyst formation rate.

stage (Fig. 4A). We identified 134 protein coding genes and 4 novel lncRNAs, which implies that lncRNAs may be involved in biological processes with protein coding genes. And we also search the neighborhoods of the 4 lncRNAs' genomic locus, but unfortunately few genes were found and also independent with the co-expressed genes. Therefore we assumed the lncRNA worked in *trans* action. The lightcyan module was enriched for functional terms associated with the regulation of maternal product degradation, zygotic transcription and protein synthesis, including RNA biosynthetic processes (odds ratio = 2.17, $P = 0.007$) and positive regulation of transcription, DNA-template (odds ratio = 4.25, $P = 0.009$), and snRNA transcription (odds ratio = 100.9, $P = 0.01$). Furthermore, the comprised genes were enriched in analogous KEGG pathways, such as RNA degradation, RNA transport, ribosome biogenesis in eukaryotes and ubiquitin-mediated proteolysis. In contrast, the modules not related to the 4-cell or 8-cell stages had no significant enrichment in ZGA functions (see Supplementary Tables S5 and S6 for gene ontology and the pathway analysis associated with the darkred WGCNA module) (Fig. 4B).

Intriguingly, the most enriched functional term was the cellular response to lipoteichoic acid (odds ratio = 82.1, $P = 5 \times 10^{-4}$), which may play a role in the mediation of adherence and recognition process.

Moreover, the second highest KEGG pathway enriched in this module was the calcium signaling pathway (odds ratio = 5.08, $P = 0.028$), which may be critical for the activation of zygotic genome transcription (Fig. 4C). The porcine pre-embryonic development was relatively controversial and confused; however, it corresponded to previous research. When the lncRNA that may regulate this process was knocked down, the embryos failed to further development to blastocysts, which was verified by the injection of siRNA into porcine embryos via parthenogenetic activation, particularly *XLOC_126976* in this module, which significantly reduced the blastocyst formation rate (Fig. 4D and Supplementary Fig. 7). These findings suggested that the module was associated with maternal product degradation and ZGA; therefore, a regulatory network was constructed to illustrate the relationships among protein coding genes and lncRNAs with respect to these biologically relevant processes (Supplementary Fig. 8). We filtered and visualized the lncRNAs and protein coding genes for which the topological overlap of the network connections was greater than the default thresholds. Thus, the module demonstrates a clear relationship between lncRNAs and protein coding genes in the regulation of ZGA during the specific stage.

The regulatory network of first lineage segregation during the blastocyst stage. To investigate the molecular mechanism responsible for first lineage segregation, we selected the tan module (which was positively co-expressed during the morula to blastocyst stages) to perform the GO and pathway enrichment analyses, in this module ICM ($r = 0.64$, $P = 3 \times 10^{-4}$) was more significant than TE ($r = 0.55$, $P = 0.003$). In particular, compared to the morula stage, the tan module was more significantly associated with the blastocyst stage, which was separated into the ICM and TE (Fig. 5A). There were 151 protein coding genes and 2 lncRNAs in this module that were highly enriched for terms of biological processes related to cell differentiation (odds ratio = 3.04, $P = 0.02$), cell fate determination (odds ratio = 9.80, $P = 0.02$) and cellular developmental process (odds ratio = 42.89, $P = 0.03$) (Fig. 5B). Moreover, the hedgehog, MAPK and Wnt signaling pathways were also enriched in this module to mediate the lineage differentiation and choice. Interestingly, we determined that metabolism was significantly altered compared with other embryonic stages and was more different than the same components of the mouse blastocyst stage embryo, including the metabolism of fatty acids and glucose. In the module, the biosynthesis and metabolism of glycerolipid and glycosphingolipid were enriched, which may be accounted for more substantial volume of lipid droplets in the porcine embryo compared to the mouse embryo. These findings suggested that the potential function of this module may regulate the initial phase of the first cell fate determination in which two distinct lineages are formed in the blastocyst stage, namely, the ICM and TE (Fig. 5C and Supplementary Tables S7 and S8).

The expression levels of *Cdx2* and *Nanog* were not significantly differentially expressed between the ICM and TE, probably because the structural and functional prediction is not well-established compared to humans and mice. Therefore, we attempted to take advantage of lncRNA as a potential marker to distinguish the ICM and TE. Furthermore, we determined that the two lncRNAs in the tan module were also specifically expressed in TE. Based on the qPCR results, we determined that *XLOC_155935* was significantly different ($P = 0.007$, Welch two-sample t-test) in the ICM, which was the same as the RNA-seq result ($P = 0.02$) (Fig. 5D and Supplementary Fig. 9). Moreover, supposing that *XLOC_155935* was a lineage-specific marker, we could have a better indicator of first lineage segregation, which generates the TE and ICM.

Regulatory network of reprogramming to pluripotency. No validated porcine embryonic stem cell (ESC) lines have been established since the 1990s⁴⁰; however, substantial efforts had been made to generate porcine pluripotent stem cells, such as putative pESCs and piPSCs. Porcine fibroblasts were used for reprogramming into iPSCs with four defined factors (*Oct4*, *Sox2*, *Klf4* and *c-Myc*). The piPSC colonies stained positive for alkaline phosphatase activity and had the ability to express surface markers of pluripotent stem cells (Supplementary Fig. 10). Moreover, compared to the porcine embryo data in the hierarchical cluster and PCA analyses, the piPSCs exhibited a high level of similarity to the porcine ICM that were used for the isolation of ES cell lines. These findings suggested that the piPSCs were successfully reprogrammed to pluripotency.

Furthermore, to investigate the potential roles of lncRNAs during the process of reprogramming, we performed a functional prediction enrichment analysis for the most significant module in iPS, the blue module, which contains 713 transcripts (702 protein coding genes and 11 lncRNAs) (Fig. 6A). Genes in the blue module were enriched for functional terms associated with pluripotency, such as the positive regulation of stem cell proliferation (odds ratio = 15.83, $P = 3.26 \times 10^{-5}$), embryo development (odds ratio = 20.97, $P = 7.17 \times 10^{-5}$), stem cell differentiation (odds ratio = 2.43, $P = 0.0003$), stem cell development (odds ratio = 2.07, $P = 0.001$) and stem cell maintenance (odds ratio = 7.37, $P = 0.011$) (Fig. 6B). We also displayed a GOChord plot of the terms relevant to embryo development and stem cells between piPSCs and fibroblasts, which indicated that among the GO terms, the most up-regulated genes were over-represented in piPSCs (Supplementary Fig. 11). The pathway of this module was enriched in the Glycolysis/Gluconeogenesis, Wnt, MAPK, Jak-STAT, VEGF, p53 and TGF-beta signaling pathways (Fig. 6C). Energy metabolism shifted from oxidative phosphorylation to glycolysis with reprogramming to pluripotency and there is also a broad consensus of opinion regarding the Wnt, MAPK, Jak-STAT, and TGF-beta signaling pathways, which play critical roles in self-renewal, the maintenance of pluripotency and the identification of key molecules in the regulation of stem cell networks. We subsequently filtered the lncRNA and protein coding gene interactions based on the topological overlap matrix and constructed the piPSC-specific regulation network (Fig. 6E and Supplementary Tables S9 and S10).

To validate the expression levels of lncRNAs identified by the RNA sequencing in the piPSC-specific module, all lncRNAs in this module were selected for qPCR analysis. Compared to fibroblasts, the expression of *XLOC_165754* was the most significant lncRNA ($P = 1.58 \times 10^{-8}$, Welch two-sample t-test) and was not expressed in fibroblasts using RNA-seq (Fig. 6D). We subsequently performed a synteny analysis to investigate whether *XLOC_165754* had co-localized and conserved blocks that linked two loci within two sets of chromosomes. *Cap2*, *Nup153*, *Kif13a* and *Dek* were identified as having shared synteny and were conserved genes ordered between

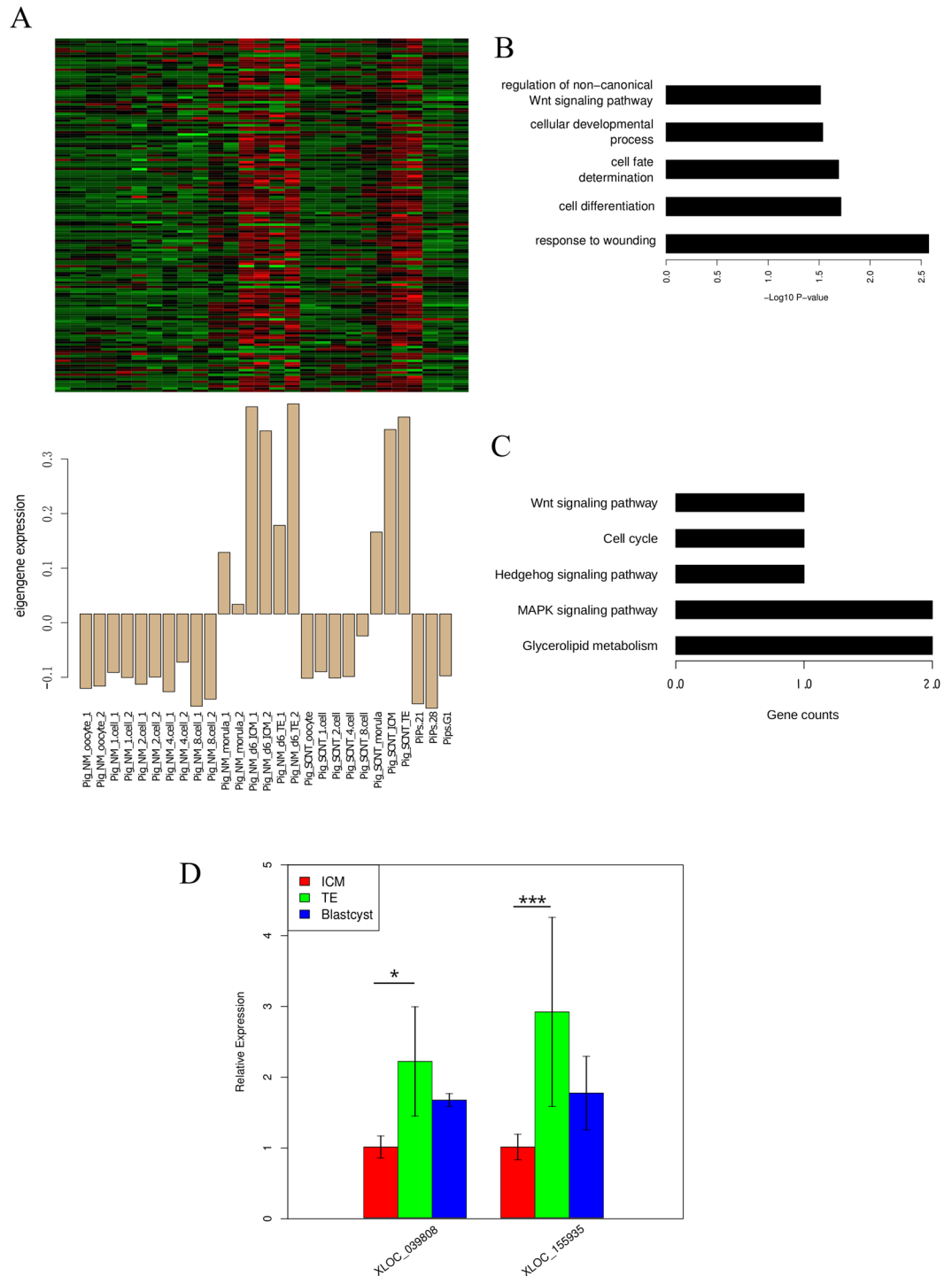


Figure 5. Co-expression network during first lineage segregation. **(A)** Heatmap and bar plot of module tan of the first lineage segregation. The heatmap indicates the expression level of genes in this module; all data were normalized with the z-score, and the bar plot indicates the module eigengene relative expression level. **(B)** Gene ontology enrichment for the module. **(C)** KEGG pathway enrichment for this module. **(D)** qPCR results of the 2 significant lncRNAs that were specifically expressed in the TE, *represent $P < 0.05$ and ***represent $P < 0.01$.

the genomes of porcine and humans or mice with 5 kb upstream and downstream of the syntenic non-coding block (Supplementary Fig. 12). These findings suggested that the iPSC module plays an important role in reprogramming to pluripotency, and XLOC_165754 is the most likely lncRNA to indicate the pluripotent state in pigs through the promotion of reprogramming by the recruitment of transcription factors or epigenetic regulators.

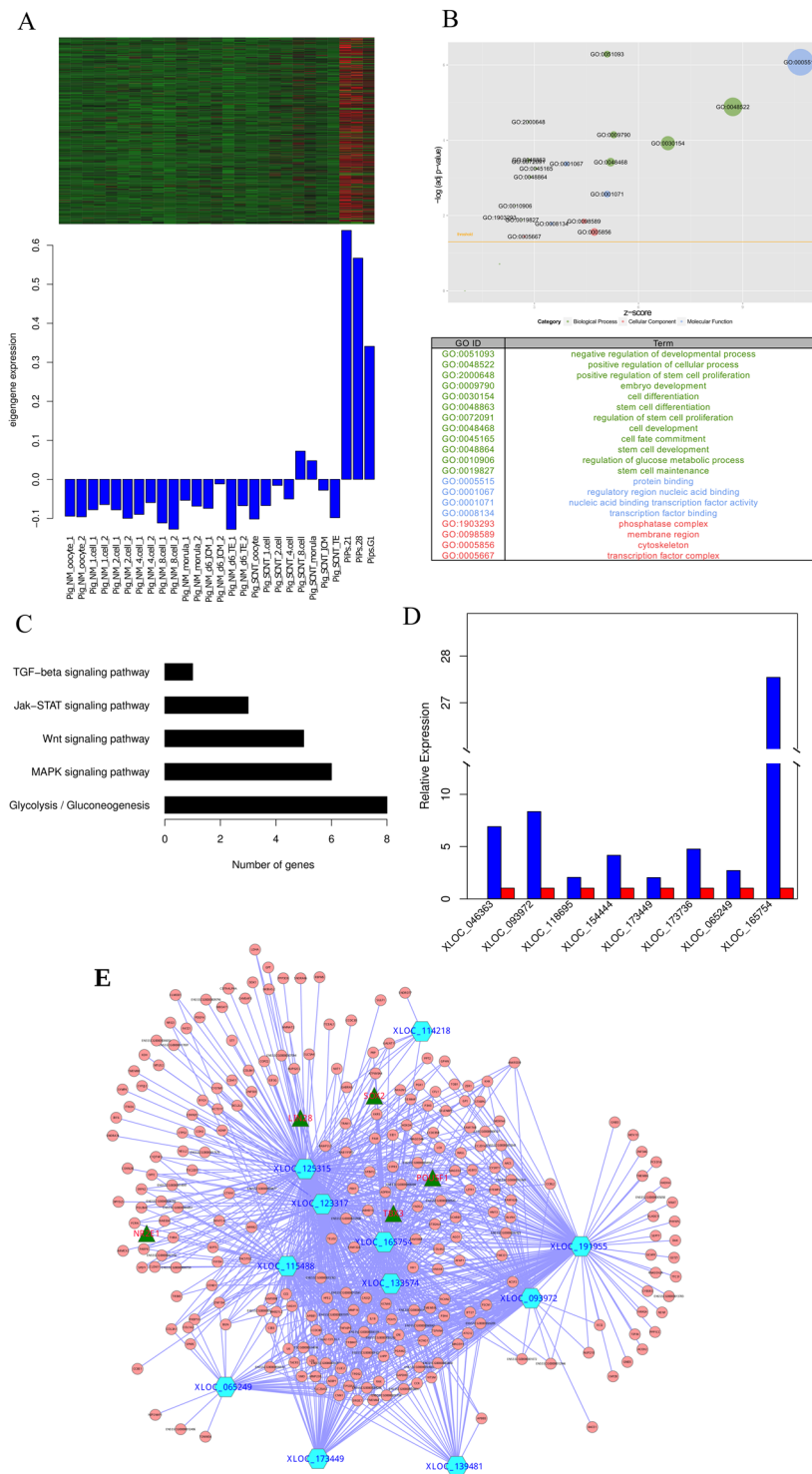


Figure 6. Co-expression network during reprogramming to pluripotency. **(A)** Heatmap and bar plot of module blue of the iPSC. The heatmap indicates the expression level of genes in this module; all data were normalized with the z-score method, and the bar plot indicates the module eigengene relative expression level. **(B)** Gene ontology enrichment for the module. **(C)** KEGG pathway enrichment for this module. **(D)** qPCR results for which the expression was significantly different between iPSC and pEF. **(E)** The lncRNA and protein coding gene regulatory network in iPSCs.

Epigenetic modification and conservation analysis for iPSC-specific lncRNAs. lncRNAs specifically expressed in different cell types indicated that epigenetic modifications may regulate the expression of lncRNAs by affecting their transcription. We determined that *XLOC_165754* was conserved across other species, such

as mice and humans. The phastCons and phyloP scores were calculated for *XLOC_165754*. However, we determined that the conserved elements appeared to follow the coding exons reasonably well while also containing several noncoding regions, and the phyloP scores were fairly consistent with the phastCons scores. Intriguingly, the phyloP scores upstream of *XLOC_165754* were less than zero, which indicates a region of accelerated evolution (Supplementary Fig. 13). To investigate whether the lncRNA expression in iPSCs was activated or inactivated, we used previous ChIP-seq data (GSE36114) to validate our results. The broader peaks extended H3K4me3 at the TSS upstream of *XLOC_165754* and also had relatively broad peaks marked with the transcriptional regions of *XLOC_165754* (Supplementary Fig. 14), which suggested the character of *XLOC_165754* in accordance with previous studies⁶. H3K4me3 was bound up with the activation of lncRNA expression, whereas H3K27me3 was related to repression, which was responsible for the significantly differentiated expression between pEF and piPSCs. However, *Oct4*, *Sox2* and *Nanog* were not marked at the promoter regions of *XLOC_165754*; these three pluripotent transcription factors play important roles in the regulation of reprogramming and the maintenance of pluripotency for humans and mice³¹. These findings demonstrated that epigenetic modifications activated the expression of *XLOC_165754* to affect reprogramming and pluripotency, which may differ from the human or mouse core pluripotent networks and may involve other protein coding genes in porcine.

Discussion

LncRNAs pose intriguing questions in developmental biology, particularly during pre-implantation embryonic development and somatic reprogramming. In this study, we systematically analyzed the lncRNA profiles of porcine pre-implantation embryos and iPSCs using RNA-seq data for the first time. To obtain high confidence lncRNAs, we filtered low-quality assemblies with an optimized FPKM threshold and identified 563 novel transcripts from 207 loci assigned low protein coding potential scores; however, the overall numbers of lncRNAs in this study is relatively small. We subsequently performed a weighted gene co-expression network and predicted the potentially functional regulatory network of stage-specific lncRNAs based on their associations with known protein coding genes involved in pre-implantation embryonic development and reprogramming to pluripotency. Moreover, we examined several potentially functional lncRNAs involved in embryonic developmental events and somatic reprogramming.

The novel lncRNAs in porcine embryos and iPSCs exhibit apparent similarities to other mammalian species. lncRNAs are typically shorter and have fewer exon numbers and lower expression levels than protein coding genes. Furthermore, lncRNAs exhibit more tissue-specific expression¹ and less conserved sequence with other mammalian lncRNAs; however, some lncRNAs have more positional conservation than sequence conservation across vertebrates³². Our findings also indicate these properties, which suggest that the identified lncRNAs are reliable. Due to different analytic methods, the genomic features of identified lncRNAs might show slight divergences. Compared to porcine lncRNAs in previous reports, we found that the lncRNAs in embryos and iPSCs (1157 nt) were shorter than those in endometrium (1454 nt) while longer than those in skeletal muscle (1043 nt). Besides, the number of exon in our result (3.3) was more than those in Wang and Zhao's (~2.4) result^{18,33}. The main reason is that a conservation method was applied in our identification pipeline, which leading to the highly conserved sequence of non-coding transcripts with human and mouse lncRNAs were chosen. Different analytic methods for identifying lncRNAs bring variation despite the genomic feature of lncRNAs in embryos and iPSCs in porcine was quite similar to former reports.

This study indicates modules of the regulatory network during embryonic development that are significantly enriched for genes involved in several developmental-specific stages. Due to the importance of embryogenesis, our data demonstrate 2 notable events during this period, including ZGA and first lineage segregation. In the ZGA period, genes were enriched for RNA biosynthesis and degradation. *Taf4* has been shown to mediate the RNA biosynthetic process and also to serve as a co-activator for developmental genes in *Drosophila* cells and vertebrates³⁴. *Kdm3a* significantly altered the expression levels of maternal H3K9me2 and H3K9me3³⁵, and at the transcriptional level, the H3K9 methylation patterns affect the switch of zygotic transcription³⁶. Lipoteichoic acid mediates adherence and recognition between blastomeres, possibly due to the biases between blastomeres that result from an asymmetry of transcriptional trends during the 2- to 16-cell embryo stages³⁷. Moreover, the impact of Ca²⁺ activated a stimulus to recruit specific maternal RNAs, which caused the activation of zygotic genome transcription^{38,39}. Previous studies indicated that the 4-cell to 8-cell stage enriched lncRNAs, such as *XLOC_126976*, may play a role as a regulator that mediates the maternal to zygotic transition. The activation of these lncRNAs appears to be critical for the further development to blastocyst, which was consistent with siRNA injection with parthenogenetically activated embryos.

For the first lineage segregation, genes were enriched for cell fate determination and cell differentiation in this period. *Gata3* is notable in this module, which has been identified as a trophoblast factor and is co-expressed with *Cdx2* in blastocysts. *Gata3* and *Cdx2* may act in parallel pathways to induce the expression of the trophoblast lineage⁴⁰. However, *Cdx2* was not significantly differentially expressed in blastocysts or co-expressed in this module. *Gata2* also plays a role in the determination of cell fate, and the expression level of *Gata2* is higher in the TE than the ICM in bovine blastocysts⁴¹. Thus, *XLOC_155935* may be regarded as a lineage-specific marker to TE instead of the absence of *Cdx2*. Furthermore, the metabolic characters of porcine pre-implantation embryos exhibit similarities and differences with other mammalian species. The similarities are represented in the well-established process of glycolysis⁴² and the increase in oxygen consumption during the blastocyst period⁴³, oxidative phosphorylation was enriched in the blastocyst stage as expected because twice as much oxygen was consumed as in the cleavage stage embryos, which was consistent with previous research³⁰; differences are present in fatty acid metabolism in which porcine embryos contain substantial amounts of lipid droplets, and the genes in this pathway are significantly upregulated, whereas this interesting phenomenon does not occur in mouse embryos²⁷. These findings suggest that lncRNAs during porcine pre-implantation development are interesting given the dramatic transcriptional regularity that occurs in embryogenesis.

In addition, due to the similarity between iPSCs and ESCs derived from blastocyst inner cell masses in humans and mice⁴⁴, we further investigated the lncRNAs and the regulatory network in piPSCs. Based on the principal component analysis and hierarchical clustering analyses, we determined the pattern of similarity between piPSCs and blastocysts, and the functional prediction of the genes in this module was also enriched for embryonic development. The Wnt signaling pathway could maintain the pluripotency in embryonic stem cells while playing a role in pre-implantation development and blastocyst lineage specification^{45,46}. The inhibition of the MAPK signaling pathway could support the self-renewal of ESCs, whereas in mouse embryos, the MAPK signaling pathway promotes TE formation and blastocyst development^{47,48}. LIF is a well-established cytokine for the maintenance of pluripotency and the self-renewal of mouse ES cells via the mediation of the Jak-stat signaling pathway, and SATA5 could also mediate the signals from cytokines that regulate pre-implantation development^{49,50}. However, the VEGF and p53 signaling pathways could induce differentiation of embryonic stem cells in mice and humans, and the p53-p21 pathway served as a barrier in iPSC generation^{51,52}. Furthermore, to our knowledge, piPSCs exhibit disparities in pluripotent states when cultured with different conditions⁵³, and we not only remarkably determined that the surface markers were different but also showed that the iPSC stage-specific lncRNAs were differentially expressed. To examine the reprogramming associated lncRNA, we focused on the most significantly expressed lncRNA in iPSCs, *XLOC_165754*, which is expressed higher in hESC like piPSCs than mESC like piPSCs. Interestingly, with the synteny analysis, we showed that several counterparts up- and down-stream of *XLOC_165754* have a function in the maintenance of stem cell pluripotency^{54,55}. Besides, according to the synteny analysis, we also found that the neighborhood of *XLOC_165754* such as *Nup153* could bind to the transcriptional start site (TSS) of developmental genes and mediate the recruitment of polycomb-repressive complex 1 (PRC1) to maintain stem cell pluripotency by functioning in mammalian epigenetic gene silencing⁵⁴, and in a comparison of the differentiation into cardiac precursors, *Kif13a* is the most highly expressed in the hESC lines⁵⁵. However, no authentic porcine ESCs have been derived, whereas piPSCs will not remain stable without exogenous genes. The co-expression network suggests that with the exception of *Oct4* and *Sox2*, endogenous *Lin28* and *Tbx3* are likely to be crucially involved in the pluripotency network, which could provide clues for the generation of high quality piPSCs.

Many studies have been conducted in mice⁵⁶, humans⁵⁷ and other model organisms¹⁰ to explore how embryonic development and reprogramming to pluripotency are regulated by lncRNAs and mRNAs; however, to date, the mechanism reported with respect to porcine pre-implantation embryos and pluripotent stem cells has been extremely limited. RNA-seq is becoming an attractive approach to profile gene expression levels with respect to decreasing sequencing costs. However, RNA-seq is confined by technical limitations that make it difficult to profile mRNA and lncRNA transcripts in porcine pre-implantation embryos due to the small number of embryonic cells and the identification and analysis of high confidence lncRNAs at relatively lower expression levels. Moreover, a blastocyst that is separated into the ICM and TE via a surgical procedure may result in a cellular response to the damage, which may be responsible for genes enriched with functions in immune or inflammatory responses. Our study identified novel lncRNA candidates; however, the accuracy could be further improved via strand-specific deep sequencing. Moreover, for porcine embryos, it is important to utilize the single cell sequencing technologies capable of providing more unbiased methods to profile the transcriptomes.

In conclusion, we identified high confidence lncRNAs involved in the processes of pre-implantation development and somatic reprogramming. Notable co-expressed modules were identified in several important events during these periods; thus, we predict the potentially functional lncRNAs and their regulatory networks with protein coding genes. Our study indicates that the transcriptional regularities of lncRNAs during embryonic development and reprogramming to pluripotency are highly dynamic and stage-specific, which will enable further investigation of these lncRNAs and may provide clues for the generation of high quality pluripotent stem cells as well as the derivation of embryonic stem cells from porcine blastocysts.

Accession Numbers. Raw reads of porcine embryos and iPSC RNA-seq data has been submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra/>) under accession number SRA326708.

Methods

Biological materials. Young adult female C57 mice (Vital River Laboratories, Beijing, China) and young adult female Nong Da Xiang mini-pigs (China Agricultural University pig farm, Zhuo Zhou, China) were kept at 25 °C, fed *ad libitum* and a 24 h light-dark cycle (12 L:12D). All animal procedures and all of the animal works were approved by the Animal Care and Use Committee of China Agriculture University (Permit Number: SKLAB-2016-05-01). All methods were carried out in accordance with relevant regulations and guidelines of State Key Laboratory for Agrobiotechnology, China Agricultural University.

Embryo collection and RNA isolation. Oocytes were collected via superovulation by an initial intraperitoneal injection of 5 IU Pregnant Mare's Serum Gonadotropin (PMSG) followed 48 h later by 5 IU human chorionic gonadotrophin (hCG). At 14 h after the last intraperitoneal injection oocytes were flushed from the oviducts. The *in vivo* pig embryos were washed from the oviduct or uterus with phosphate-buffered saline (PBS) containing 5% fetal bovine serum (FBS) after estrus and natural mating. We use the Nong Da Xiang mini-pigs to obtain the oocytes and fibroblasts that respectively refer to the gametes and the donors in somatic cell nuclear transfer. Then through oocyte maturation, cell cycle synchronization of donor cells, enucleation, cell fusion, oocyte activation and embryo culture, we collected the *in vitro* embryos. The blastocyst segregated TE and ICM physically by an ultra-sharp splitting blade (Bioniche, Animal Health US, Inc) with a stereomicroscope.

After the collection of embryos, we mixed 4–10 embryos at the same stage together for each sample and transferred and centrifuged into a 0.5 mL tube with 50 µL extraction buffer from the PicoPure RNA Isolation Kit (Arcturus, KIT0204, Life Technologies, US) at 42 °C for 30 min. The extract was then held at –80 °C until RNA isolation. Total RNA was isolated from each sample according to the manufacturer's protocol and eluted into 10 µL elution buffer.

RNA sequencing. The total RNA was used as starting material to select poly-A RNA and used for constructing SOLiD libraries according to the protocols supplied by the Applied Biosystems SOLiD 4 System manufacturer. RNA samples were spiked in with NIST standards before libraries were constructed. RNA isolated from each embryonic sample was used for double-stranded cDNA synthesis, PCR amplification. The fragment library was prepared according to the Library Preparation Protocol for whole transcriptome analysis of a single cell and the Applied Biosystems SOLiD 4 System Library Preparation Guide (<http://www.appliedbiosystems.com>). The RNA-seq libraries were sequenced on ABI SOLiD sequencing platform as 50-base reads according to the manufacturer's recommendations.

Reads mapping and transcripts assembly. Short reads were mapped to the mm9 (mouse) and susScr2 (pig) reference genome with Tophat^{58,59}. The annotation files were obtained from the Ensembl database: mouse (NCBIM37.65) and pig (Sus scrofa9.65). Aligned reads of each sample were assembled with cufflinks (version 2.21)⁶⁰ and we performed Reference Annotation Based Transcript (RABT) assembly⁶¹ with the Ensembl annotation to compensate incompletely assembled transcripts caused by read coverage gaps in the regions of known genes in the assembly. We merged together all cufflinks assemblies and used cuffcompare to filter the transcripts with those annotations in Ensembl gene, Refseq gene and UCSC gene. Then we screened the transcripts with those transfrag class codes: 'i' (A transfrag falling entirely within a reference intron), 'j' (Potentially novel isoform or fragment that at least on splice junction is shared with a reference transcript), 'o' (generic exonic overlap with a reference transcript), 'u' (intergenic transcript) and 'x' (exonic overlap with reference on the opposite strand).

Long non-coding RNA Identification. We first filter the high confidence transcript with a cutoff defined in Supplementary Fig. 1. Then we extracted the transfrags class code marked with 'i' (a transfrag falling entirely within a reference intron), 'j' (potentially novel isoform), 'o' (generic exonic overlap with a reference transcript), 'u' (intergenic transcript), and 'x' (Exonic overlap with reference on the opposite strand) which may be lncRNA candidates and subsequently excluded the single exon, short exonic length (<200 nt) and short open reading frame (<300 nt) transcripts. We performed both the Coding Potential Calculator (CPC) and Coding-Potential Assessment Tool (CPAT) to calculate the coding potential, as well as PhyloCSF to analyze the conserved regions of the multi-species genome sequence alignment of all candidate transcripts to distinguish the coding and non-coding sequences. Furthermore, the transcripts were searched against the Pfam database of hidden Markov models (HMMs) to filter the sequences that contained protein domains.

Weighted Gene Co-expression analysis Network Analysis (WGCNA). Although method of WGCNA was designed for the analysis of microarray data, it has been already applied to RNA-Seq data^{62,63}. Gene expression FPKM values were $\log_2(x + 1)$ transformed before being processed through the WGCNA R package²⁹. We applied co-expression analysis to identify the specific modules of highly correlated genes and lncRNAs by the referenced methods⁶⁴ and an unsigned network was constructed via creating a matrix of pairwise correlation between all pairs of genes across the samples. Based on the scale free topology, we picked up the most proper soft-thresholding power⁶⁵. The max block size was set to 12000 and the minimum module size was set to 75. Dynamic hybrid tree cut algorithm was used to cut the hierarchical clustering tree, and the branches from the tree cutting were used to define modules. Eigengenes whose correlation value greater than 0.75 in each module were merged and the first principal component of each individual module was used to evaluate their relationship with cell type. Other parameters were used with the default to construct a block-wise network.

Differential expression analysis and function enrichment analysis. For protein coding genes, we use HTseq⁶⁶ to obtain read count tables from binary sequence alignment/map (BAM) file. The DESeq 2⁶⁷ package was used to normalize count data and calculate the differential expression tests between each two groups associated cell type. A gene or lncRNA will be reported significant differential expression gene if the DESeq 2 results show that the p-value was less than 0.05 and the absolute \log_2 fold change was greater than 1. The Gene Ontology (GO) functional and KEGG pathway enrichment analyses of differential expression genes and WGCNA module genes were performed in R/Bioconductor using the package: GOSTats, GSEA⁶⁸ to compute the statistics and GOplot⁶⁹ to visualize the function analysis. All the statistical significance described above was p-value.

Cell culture. The porcine embryonic fibroblasts (pEFs) were isolated from embryos (approximately embryonic 26-30d) arising from Nong Da Xiang. pEFs were cultured in DMEM medium containing 10% fetal bovine serum (FBS), 1% non-essential amino acid, 1% GlutaMAX-L and 1% penicillin/streptomycin (Gibico). 293-GP2 cells were cultured in DMEM with 10% FBS that used for viral packaging. The generated piPSCs were cultured in the commercial medium (NutriStem XF/FF Culture Medium, Stemgent 01-0005), and were incubated at 38 °C at 5% CO₂. The piPSCs were passaged by 0.1% collagenase IV (Gibico).

The generation of piPSCs. To perform the reprogramming of pEFs, 4 retroviral vectors were constructed with porcine *Oct4*, *Sox2*, *c-Myc* and *Klf4* respectively. Retroviral vectors production was conducted according to the method reported previously⁷⁰. pEFs were infected with the indicated combinations of retroviruses for 12 h with 8 μ g/ml polybrene (Sigma, 107689). The infection was performed twice. Infected pEFs were passaged and seeded into the plated feeder layer, which was inactivated mouse embryonic fibroblast (MEF) with mitomycin C. Then changing the pEF medium to piPSC medium on the second day and the medium was changed every two days. The reprogramming efficiency was evaluated by AP staining at day 15. Colonies were picked up at day 30.

Quantitative real-time PCR analysis. Total RNA of piPSCs and porcine fibroblast were extracted with an RNeasy Mini Kit (QIAGEN, 74104). The reverse transcription was conducted with Oligo-dT primer and M-MLV

Reverse Transcriptase (Promega, M1701). Quantitative real-time PCR was conducted with LightCycler 480 SYBR Green I Master Kit (Roche, 4887352001) according to the manufacturer's instructions, Ct (fit point method) and Cp (second derivative method) values were detected with LightCycler 480II (Roche) and exported into an MS Excel data sheet (Microsoft) for analysis after background subtraction. The data was analyzed with the comparative Ct ($2^{-\Delta\Delta Ct}$) method. The ΔCt was calculated using EF1- α as a internal control. All experiments encompassed at least 3 biological replicates. The primers used in these analyses are listed in Supplementary Table S1.

Conservation analysis. PhastCons scores and PhyloP scores were used to estimate the conservation level of long non-coding transcripts. We download the pairwise alignments of susScr2 among human (hg19), mouse (mm9) and cow (bosTau7) that is best chained or with gaps in the best chains filled by the next-best chains where possible in the susScr2 genome. Then group blocks of chained alignments into longer stretches of synteny and recreate the multiple alignments. We conduct PhastCons⁷¹ to calculate the whole genome phyloP score and phastCon score with default parameters. The phyloP and phastCons score of long non-coding transcripts were visualized by RPHAST⁷².

ChIP-seq analysis. To identify the porcine lncRNAs whether involved in pluripotency in piPSCs, we use a series of histone modification and DNA-binding factors datasets in GSE36114 which was downloaded from Gene Expression Omnibus (GEO). After quality check (QC) we mapped the raw reads with bowtie against susScr2. MACS2⁷³ were used to process for the peak calling. We integrated the RNA-seq data and ChIP-seq data and visualized them in Integrative Genomic Viewer (IGV).

siRNA injection of porcine embryos by parthenogenetic activation. We microinjected 5–10 pl of 2 μ M siRNA that targeted lncRNA that regulated zygotic genome activation, together with GFP mRNA, into the parthenogenetically activated embryos. All the siRNAs that targeted ZGA associated lncRNA are listed in Supplementary Table S2.

References

- Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775–1789, <https://doi.org/10.1101/gr.132159.111> (2012).
- Jia, H. *et al.* Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *Rna* **16**, 1478–1487, <https://doi.org/10.1261/rna.1951310> (2010).
- Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11667–11672, <https://doi.org/10.1073/pnas.0904715106> (2009).
- Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development* **25**, 1915–1927, <https://doi.org/10.1101/gad.174466.11> (2011).
- Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419, <https://doi.org/10.1016/j.cell.2010.06.040> (2010).
- Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227, <https://doi.org/10.1038/nature07672> (2009).
- Nam, J. W. & Bartel, D. P. Long noncoding RNAs in *C. elegans*. *Genome research* **22**, 2529–2540, <https://doi.org/10.1101/gr.140475.112> (2012).
- Brown, J. B. *et al.* Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399, <https://doi.org/10.1038/nature12962> (2014).
- Kaushik, K. *et al.* Dynamic expression of long non-coding RNAs (lncRNAs) in adult zebrafish. *PLoS one* **8**, e83616, <https://doi.org/10.1371/journal.pone.0083616> (2013).
- Pauli, A. *et al.* Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome research* **22**, 577–591, <https://doi.org/10.1101/gr.133009.111> (2012).
- Li, T. *et al.* Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* **99**, 292–298, <https://doi.org/10.1016/j.ygeno.2012.02.003> (2012).
- Weikard, R., Hadlich, F. & Kuehn, C. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC genomics* **14**, 789, <https://doi.org/10.1186/1471-2164-14-789> (2013).
- Caballero, J. *et al.* Exploring the function of long non-coding RNA in the development of bovine early embryos. *Reproduction, fertility, and development* **27**, 40–52, <https://doi.org/10.1071/RD14338> (2014).
- Billerey, C. *et al.* Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC genomics* **15**, 499, <https://doi.org/10.1186/1471-2164-15-499> (2014).
- Esteve-Codina, A. *et al.* Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. *BMC genomics* **12**, 552, <https://doi.org/10.1186/1471-2164-12-552> (2011).
- Zhou, Z. Y. *et al.* Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome biology and evolution* **6**, 1387–1392, <https://doi.org/10.1093/gbe/evu113> (2014).
- Du, Z. Q. *et al.* Identification of species-specific novel transcripts in pig reproductive tissues using RNA-seq. *Animal genetics* **45**, 198–204, <https://doi.org/10.1111/age.12124> (2014).
- Zhao, W. *et al.* Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development. *Sci Rep* **5**, 8957, <https://doi.org/10.1038/srep08957> (2015).
- Quek, X. C. *et al.* lncRNAdbv2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic acids research* **43**, D168–173, <https://doi.org/10.1093/nar/gku988> (2015).
- Volders, P. J. *et al.* LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic acids research* **41**, D246–251, <https://doi.org/10.1093/nar/gks915> (2013).
- Dinger, M. E. *et al.* NRED: a database of long noncoding RNA expression. *Nucleic acids research* **37**, D122–126, <https://doi.org/10.1093/nar/gkn617> (2009).
- Li, A. *et al.* ALDB: a domestic-animal long noncoding RNA database. *PLoS one* **10**, e0124003, <https://doi.org/10.1371/journal.pone.0124003> (2015).
- Xie, C. *et al.* NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic acids research* **42**, D98–103, <https://doi.org/10.1093/nar/gkt1222> (2014).

24. Kobayashi, E., Hishikawa, S., Teratani, T. & Lefor, A. T. The pig as a model for translational research: overview of porcine animal models at Jichi Medical University. *Transplantation Research* **1**, 8–8, <https://doi.org/10.1186/2047-1440-1-8> (2012).
25. Brevini, T. A., Antonini, S., Cillo, F., Crestan, M. & Gandolfi, F. Porcine embryonic stem cells: Facts, challenges and hopes. *Theriogenology* **68**(Suppl 1), S206–213, <https://doi.org/10.1016/j.theriogenology.2007.05.043> (2007).
26. Hall, V. Porcine embryonic stem cells: a possible source for cell replacement therapy. *Stem cell reviews* **4**, 275–282, <https://doi.org/10.1007/s12015-008-9040-2> (2008).
27. Cao, S. *et al.* Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. *BMC genomics* **15**, 4, <https://doi.org/10.1186/1471-2164-15-4> (2014).
28. Luo, D. J., Hu, W., Chen, S. P. & Zhu, Z. Y. Critical developmental stages for the efficiency of somatic cell nuclear transfer in zebrafish. *International journal of biological sciences* **7**, 476–486 (2011).
29. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, <https://doi.org/10.1186/1471-2105-9-559> (2008).
30. Notarianni, E., Laurie, S., Moor, R. M. & Evans, M. J. Maintenance and differentiation in culture of pluripotent embryonic cell lines from pig blastocysts. *J Reprod Fertil Suppl* **41**, 51–56 (1990).
31. Wu, S. C., Kallin, E. M. & Zhang, Y. Role of H3K27 methylation in the regulation of lincRNA expression. *Cell research* **20**, 1109–1116, <https://doi.org/10.1038/cr.2010.114> (2010).
32. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550, <https://doi.org/10.1016/j.cell.2011.11.055> (2011).
33. Wang, Y. *et al.* Analyses of Long Non-Coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Sci Rep* **6**, 20238, <https://doi.org/10.1038/srep20238> (2016).
34. Wright, K. J. & Tjian, R. Wnt signaling targets ETO coactivation domain of TAF4/TFIID *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 55–60, <https://doi.org/10.1073/pnas.0811914106> (2009).
35. Fukuda, A. *et al.* The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. *Nature communications* **5**, 5464, <https://doi.org/10.1038/ncomms6464> (2014).
36. Ancelin, K. *et al.* Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. *eLife* **5**, e08851, <https://doi.org/10.7554/eLife.08851> (2016).
37. Shi, J. *et al.* Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell RNA-seq. *Development* **142**, 3468–3477, <https://doi.org/10.1242/dev.123950> (2015).
38. Ducibella, T. *et al.* Egg-to-embryo transition is driven by differential responses to Ca(2+) oscillation number. *Dev Biol* **250**, 280–291 (2002).
39. Aoki, F., Hara, K. T. & Schultz, R. M. Acquisition of transcriptional competence in the 1-cell mouse embryo: requirement for recruitment of maternal mRNAs. *Mol Reprod Dev* **64**, 270–274, <https://doi.org/10.1002/mrd.10227> (2003).
40. Ralston, A. *et al.* Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137**, 395–403, <https://doi.org/10.1242/dev.038828> (2010).
41. Hosseini, S. M. *et al.* Transcriptome profiling of bovine inner cell mass and trophectoderm derived from *in vivo* generated blastocysts. *BMC developmental biology* **15**, 49, <https://doi.org/10.1186/s12861-015-0096-3> (2015).
42. Hardy, K. *et al.* Non-invasive measurement of glucose and pyruvate uptake by individual human oocytes and preimplantation embryos. *Human reproduction* **4**, 188–191 (1989).
43. Houghton, F. D., Thompson, J. G., Kennedy, C. J. & Leese, H. J. Oxygen consumption and energy metabolism of the early mouse embryo. *Mol Reprod Dev* **44**, 476–485, [https://doi.org/10.1002/\(SICI\)1098-2795\(199608\)44:4<476::AID-MRD7.3.0.CO;2-I](https://doi.org/10.1002/(SICI)1098-2795(199608)44:4<476::AID-MRD7.3.0.CO;2-I) (1996).
44. Narsinh, K. H., Plews, J. & Wu, J. C. Comparison of Human Induced Pluripotent and Embryonic Stem Cells: Fraternal or Identical Twins? *Molecular therapy: the journal of the American Society of Gene Therapy* **19**, 635–638 (2011).
45. Biechele, S., Cockburn, K., Lanner, F., Cox, B. J. & Rossant, J. Porcn-dependent Wnt signaling is not required prior to mouse gastrulation. *Development* **140**, 2961–2971, <https://doi.org/10.1242/dev.094458> (2013).
46. Sokol, S. Y. Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development* **138**, 4341–4350, <https://doi.org/10.1242/dev.066209> (2011).
47. Lu, C. W. *et al.* Ras-MAPK signaling promotes trophectoderm formation from embryonic stem cells and mouse embryos. *Nature genetics* **40**, 921–926, <https://doi.org/10.1038/ng.173> (2008).
48. Qi, X. *et al.* BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6027–6032, <https://doi.org/10.1073/pnas.0401367101> (2004).
49. Hirai, H., Karian, P. & Kikyo, N. Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. *The Biochemical journal* **438**, 11–23, <https://doi.org/10.1042/BJ20102152> (2011).
50. Nakasato, M. *et al.* Involvement of the STAT5 signaling pathway in the regulation of mouse preimplantation development. *Biology of reproduction* **75**, 508–517, <https://doi.org/10.1095/biolreprod.105.047860> (2006).
51. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460**, 1132–1135, <https://doi.org/10.1038/nature08235> (2009).
52. Chen, G. *et al.* Blocking autocrine VEGF signaling by sunitinib, an anti-cancer drug, promotes embryonic stem cell self-renewal and somatic cell reprogramming. *Cell research* **24**, 1121–1136, <https://doi.org/10.1038/cr.2014.112> (2014).
53. Zhang, W. *et al.* Pluripotent and Metabolic Features of Two Types of Porcine iPSCs Derived from Defined Mouse and Human ES Cell Culture Conditions. *PLoS one* **10**, <https://doi.org/10.1371/journal.pone.0124562> (2015).
54. Jacinto, F. V., Benner, C. & Hetzer, M. W. The nucleoporin Nup153 regulates embryonic stem cell pluripotency through gene silencing. *Genes & development* **29**, 1224–1238, <https://doi.org/10.1101/gad.260919.115> (2015).
55. Salomonis, N. *et al.* Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. *PLoS computational biology* **5**, e1000553, <https://doi.org/10.1371/journal.pcbi.1000553> (2009).
56. Zhang, K., Huang, K., Luo, Y. & Li, S. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC genomics* **15**, 845, <https://doi.org/10.1186/1471-2164-15-845> (2014).
57. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology* **20**, 1131–1139, <https://doi.org/10.1038/nsmb.2660> (2013).
58. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352> (2009).
59. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36, <https://doi.org/10.1186/gb-2013-14-4-r36> (2013).
60. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562–578, <https://doi.org/10.1038/nprot.2012.016> (2012).
61. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329, <https://doi.org/10.1093/bioinformatics/btr355> (2011).
62. Giorgi, F. M., Del Fabbro, C. & Licausi, F. Comparative study of RNA-seq- and microarray-derived coexpression networks in Arabidopsis thaliana. *Bioinformatics* **29**, 717–724, <https://doi.org/10.1093/bioinformatics/btt053> (2013).

63. Iancu, O. D. *et al.* Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinformatics* **28**, 1592–1597, <https://doi.org/10.1093/bioinformatics/bts245> (2012).
64. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**, Article17, <https://doi.org/10.2202/1544-6115.1128> (2005).
65. Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC bioinformatics* **8**, 22, <https://doi.org/10.1186/1471-2105-8-22> (2007).
66. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, <https://doi.org/10.1093/bioinformatics/btu638> (2015).
67. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome biology* **15**, 550, <https://doi.org/10.1186/s13059-014-0550-8> (2014).
68. Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464–1465, <https://doi.org/10.1093/bioinformatics/bth088> (2004).
69. Walter, W., Sanchez-Cabo, F. & Ricote, M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914, <https://doi.org/10.1093/bioinformatics/btv300> (2015).
70. Ruan, W. *et al.* A novel strategy to derive iPS cells from porcine fibroblasts. *Science China. Life sciences* **54**, 553–559, <https://doi.org/10.1007/s11427-011-4179-5> (2011).
71. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* **15**, 1034–1050, <https://doi.org/10.1101/gr.3715005> (2005).
72. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in bioinformatics* **12**, 41–51, <https://doi.org/10.1093/bib/bbq072> (2011).
73. Liu, T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods in molecular biology* **1150**, 81–95, https://doi.org/10.1007/978-1-4939-0512-6_4 (2014).

Acknowledgements

This project was supported by The National Key Research and Development Program of China (2016YFA0100202), National and Beijing Natural Science Foundation of China (31601941, 6152004), General project of the Beijing Municipal Education Committee (PMX2016_014207_000008) and High-level Scientific Research Cultivation Project of BUA, The National Thousand Talents Program of China and Research Programs from the State Key Laboratories for Agrobiotechnology, China Agricultural University (2018SKLAB6-20).

Author Contributions

J.H. and S.C. conceived and designed the experiments. L.Z. analyzed the data. H.M. performed the qPCR experiments. B.W., W.Z. and Q.W. performed si-RNA injection. G.G. and S.C. helped interpret the results. L.Z. wrote the paper. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-24863-5>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018