Transcriptome analysis of in vitro fertilization and parthenogenesis activation during early embryonic development in pigs

Xin Li
College of Animal Science and Technology, Huazhong Agricultural University
Ziying Huang
Huazhong Agriculture University College of Animal and Science
Cheng Zou
Huazhong Agriculture University college of Animal and science
Mengxun Li
Huazhong Agriculture University College of Animal and Science
Xiangdong Liu
Huazhong Agriculture University college of animal and science
Shuhong Zhao
Huazhong Agriculture University College of Animal and Science
Changchun Li (lichangchun@mail.hzau.edu.cn)
Huazhong Agriculture University https://orcid.org/0000-0002-8030-8632

Research article

Keywords: Pig, parthenogenesis activation, single-cell RNA-seq, alternative splicing, lincRNAs

Posted Date: March 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17009/v1

Version of Record: A version of this preprint was published at Genes on September 22nd, 2021. See the published version at https://doi.org/10.3390/genes12101461.

Abstract

Background

Parthenogenesis-activated oocytes exhibit inferior embryonic development and lower total cell count per blastocyst than fertilized embryos, thus often leading to pre-implantation failure in porcine. The mechanisms underlying the deficiencies of embryos generated from parthenogenesis activation have not been completely understood. Alternative splicing events (AS) and long intergenic non-coding RNAs (lincRNAs) have emerged as key regulators in various biological processes. However, their regulatory mechanisms for parthenogenesis have not been fully elucidated, and their relation to parthenogenesis in pig is largely unknown.

Result

Here, we reconstructed 189,228 transcripts and identified 7,185 lincRNAs in a single mature oocyte and discovered three stages between in vitro fertilization (IVF) and parthenogenesis activation (PA). A large change in transcriptome occurred in early blastocyst and morula in IVF and PA, respectively, indicating tremendous alteration for the transcriptome in both processes. Compare to protein-coding genes, differential expression protein-coding genes have higher alternative splicing rate, and most transcripts possibly originated from transcription start site (TSS) and transcription terminal site (TTS) types of alternative splicing according to the average expression of 12 AS types. Most lincRNAs with remarkably stage specificity were found in both in procceses. Function prediction results suggest that the specifically expressed lincRNAs in distinct stages have different biological functions. Hub lincRNAs analysis showed that three lincRNAs may play important roles on early embryonic development, and five may function on abnormal parthenogenesis embryonic development. Comparison of the transcriptome between IVF and PA revealed that most DEGs enriched the regulation of apoptotic and late embryonic development-related process.

conclusion

We performed a comprehensive comparative analysis framework of RNA-seq data, including mRNA, alternative splicing and long intergenic noncoding RNA. This work identified functional protein-coding genes and lincRNAs, and showed differences in the expression pattern of alternative splicing in IVF and PA. Which detailed the early embryonic development-related procedure in porcine, some protein-coding genes, and lincRNAs that may affect pre-implantation failure in PA embryos. This study will help future research on these genes and molecular-assisted breeding for pig parthenotes.

Background

Pig (Sus scrofa) is one of the most important domesticated animals and is a well-suited biomedical model for human disease because of its similarities in anatomy and physiology with humans(Brevini et al., 2007; Hall, 2008; Lunney, 2007; Patterson et al., 2008). Therefore, the in vitro production of early

porcine embryos is of particular scientific and economic interest(Zhou et al., 2014). Parthenogenesis is a common model of asexual reproduction that does not involve males and is common in lower species but has been reported only in few vertebrate species (approximately 0.1%)(Dua et al., 2019). Embryos derived from parthenogenetic activation (PA) are valuable for studies on embryonic stem cells and gene imprinting with few ethical issues(Brevini and Gandolfi, 2008; Cibelli et al., 2006; Kono et al., 2004). These specimens can be utilized for various tissue engineering applications. However, embryos generated from PA contain exclusively maternal genomes, resulting in embryonic death in mammals. Mammalian PA embryonic death occurs at 12, 21, 28, and 30 days of gestation in rabbits, sheep, pig, and cattle, respectively(Zhu et al., 2003).

Alternative splicing (AS) is an important post-transcriptional process in which the mRNA precursor splices intron to form the mature mRNA in different manners and serves as a major source of cellular protein diversity(Yang et al., 2016b). Up to 100,000 distinct isoform transcripts could be produced from ~ 20,000 human protein-coding genes(Pan et al., 2008), and over a million distinct polypeptides can be obtained through the post-translational modification of products of all possible transcript isoforms(Smith et al., 2013; Wang et al., 2019b). Many studies identified AS events during the early embryonic development of mice(Matsuda et al., 2019), cattle(Laporta et al., 2011), chicken(Boije et al., 2013), and other animals. However, the global contribution of AS to proteomic complexity in pig early embryonic development remains unknown.

lincRNAs are a class of intergenic transcripts that are greater than 200 nt in length and have limited protein-coding potential. These molecules have been increasingly recognized as an important regulatory factor of gene expression and acts as molecular platform that associates different domains with DNA, RNA, or proteins(Blank-Giwojna et al., 2019; Li et al., 2019; Wang et al., 2019b) in various biological processes, such as gene regulation, maintenance of pluripotency, and transcriptional regulation(Guttman et al., 2011; Loewer et al., 2010). Owing to the development of sequencing technology, many lincRNAs, such as H19, p21, and Xist have been identified, and their functional relevance to embryogenesis has been observed in many species(Chen et al., 2019; Keniry et al., 2012; Suvorova et al., 2016). However, the functional characterization of most lincRNAs remains unclear, and their regulatory mechanisms on pre-implantation failure in pig parthenogenesis embryos needs further study.

Here, we sequenced the transcriptome of oocytes (meiosis II, MII) and early embryos at three developmental stages (zygote, morula, and blastocyst) between in vitro fertilization (IVF) and parthenogenesis activation (PA) in porcine. We analyzed 21 single cells to investigate the genes and lincRNAs that influence the biochemical and biological process during early embryonic development and abnormal parthenogenesis embryos in porcine. A total of 9,579 and 6,279 differentially expressed genes for IVF and PA, respectively, were detected to be related to early embryonic development. Moreover, approximately 10,000 AS events were discovered in annotated and un-annotated genes, indicating a significantly regular change with early embryo development.

lincRNAs in adjacent stages were also compared, yielding 7,185 putative lincRNAs and 3,852 novel lincRNAs that had not been previously identified in porcine. Basing on the expression information, we detected some differentially expressed lincRNAs in IVF and PA, respectively. Gene ontology and pathway analyses were conducted on the putative target genes (PTGs) of lincRNAs and revealed their significant participation in some cell cycle-related and transcription-related biological processes. By combining the differential expression analysis results of mRNAs, we found that most differentially expressed lincRNAs (DELs) can positively regulate the PTG expression. Hence, our study provides an in-depth understanding of the dynamics of transcriptional regulation during early embryonic development and some insights into the functional roles of specific lincRNAs that may improve PA in porcine embryogenesis.

Results

1. Overview of sequencing data

We obtained a total of 3,484.4 million raw reads via single-cell transcriptome sequencing with 166.0 million reads per sample. After filtering adapter and low-quantity reads, we detected 2,303.2 million clean reads with 110.0 million read per sample. Each RNA-seq dataset was separately aligned to the Sus scrofa (11.1) genome by using HiSat2. Aligned result showed that most of the samples had a comparative high align rate over 85%, except PA_6_3 with unique mapping rate range of 55%–75% (TABLE1). After reconstructing the transcriptome for each group and merging the 21 assembled transcripts into a non-redundant transcriptome, we counted the number of reads in each sample by using HTseq-count. We obtained these genes as expressed gene with FPKM $\geq$ 0.1 in at least three samples. Finally, we identified 15,322 known protein-coding genes accounting for 68.6% of annotated porcine protein-coding gene (22,342). We also performed Pearson correlation analysis on repeat samples in one group, and the results are shown in the Supplementary Figure S1.

2. Differential expression analysis of mRNAs in adjacent stages

First, we normalized the transcript expression levels to the FPKM values by using HTSeq count. We then performed cluster analysis with these 21 samples and found mature oocytes and zygote embryos clustered closely together but away from morula and early blastocyst, indicating that the one-cell stage exhibits a distinct transcriptome pattern (Figure 1A). Next, we conducted differential expression analysis for IVF and PA samples by using DeSeq2 to explore the early embryonic development-related genes. We identified 11,110 differential expression genes (DEGs) in IVF and PA. By comparing the number of DEGs in adjacent stages, we found the largest number of DEGs occurred in early blastocyst stage in IVF, and morula stage in PA. In pigs, ZGA mainly occurs between the four- and eight-cell stages (Figure 6 D). However, in this work, transcriptomic change in IVF occurred in the early blastocyst stage, indicating that zygotic activation may be sustained until the early blastocyst in IVF. Pathway enrichment analysis with the DEGs of these two stages is shown in Supplementary Figure S2 A-D. We found that most of DEGs in IVF were involved in transcriptome, cell division, cell proliferation and differentiation, energy metabolism, biosynthesis, embryo development related-biological pathways. Most DEGs in PA participated in such

pathways due to their similarity with those in IVF, but others were significantly involved in the regulation of apoptotic processes.

We performed cluster analysis of the DEGs in IVF and PA by using the short time-series expression miner (Stem) software(Luo and Liu, 2019). We performed functional annotation and enrichment analysis in each group for the nine clusters with p-value < 0.01 in IVF and PA, respectively (Figure 1B, Figure 2). We found that these nine clusters of IVF and PA were significantly related to early embryonic development processes, such as transcription, cell cycle, and cell proliferation and differentiation. However, compared with those in IVF, many pathways related to the regulation of apoptotic process occurred in the DEGs of parthenogenesis-activated embryos, such as AIMP2, BIRC5, BNIP1, and BOK(Choi et al., 2009; Luo et al., 2014; MacDonald et al., 2018; Wang et al., 2013).

3. Identification and nearest-neighbor analysis of lincRNAs

Some lincRNAs play important roles in various biological processes, such as epigenetic regulation, maintenance of pluripotency, and transcriptional regulation(Atianand et al., 2016; Guo et al., 2018; Nordin et al., 2014). However, lincRNAs are still poorly characterized in porcine early embryonic development both in IVF and PA. Here, we first filtered transcripts according to our recently updated pipeline to identify putative lincRNAs (Figure 3A). Finally, we obtained 15,070 putative lincRNA transcripts produced by 7,185 loci, which were distributed in all chromosomes, except the Y chromosome (Figure 3B), and 6,113 of these 15,070 lincRNA transcripts had no overlap with currently annotated transcripts. Moreover, based on the comparison of the basic characterization between lincRNAs and protein-coding genes, such as average transcript length, mean exon length, exon number, and average expression, lincRNAs have shorter transcript length, longer exon length, and smaller exon number (Supplementary Figure S3 A-D), which are consistent with previous reports(Zou et al., 2018). In addition, we perform differential expression analysis of linRNAs in adjacent stages. DELs had a similar trend with protein-coding genes in which a large number of DEGs and DELs both occurred at the same stage (Supplementary Figure S3 E-F), indicating that the function of DELs is similar to that of DEGs.

The prediction of the functions of lincRNAs remains challenging due to their lack of annotation and low expression levels. The function of lincRNAs can be predicted via cis-acting manner (proximity regulation) and trans-acting manner (correlation prediction). lincRNAs could act in cis to regulate their neighboring genes(Casero et al., 2015). Correlation-based approaches have been also used to infer the function of lincRNAs(Ahn et al., 2016). To explore the potential cis-acting correlation between pig lincRNAs and their neighboring protein-coding genes, we performed GO enrichment analysis on expressed protein-coding genes transcribed near lincRNA (<100kb). We found that these neighboring protein-coding genes of pig lincRNAs were mostly enriched in biological processes or pathways related to transcription, protein modification, and cell development, such as transcription from RNA polymerase II promoter, nature killer cell activation involved in immune response, positive regulation of peptidyl-seine phosphorylation of SAT protein, regulation of cell cycle arrest, and cell–cell adhesion (Figure 3C).

4. Identification and function analysis of hub lincRNAs

Here, we performed a transcriptome-wide weighted gene co-expression network analysis (WGCNA) to infer the potential roles of lincRNAs in IVF and PA. We identified 15 co-expression modules with size ranging from 74 to 3841 (mean: 647; median: 273). Then, we filtered some modules by development stages specific (correlation ≤ 0.6, p-value ≤ 0.05) (Figure 4A) and numbers of lincRNAs per each module ($n < 1$). In total, we selected 6 modules (blue, green, green-yellow, pink, purple, and turquoise) and then performed GO enrichment analysis of protein-coding genes in these six modules. Most of the modules were enriched for transcription regulation, cell cycle, cell proliferation, and metabolic processes. This finding suggested that the lincRNAs in these modules may play vital roles in early embryonic development (Figure 4B, Supplementary Table S7).

Next, we selected these six modules for further hub lincRNA analysis. Hub genes are centrally located in a scale-free network of each module and reflect the core functions of network. To identify hub lincRNAs in pig early embryonic development, we measured the intra-modular connectivity (also named weight) of each gene by WGCNA to select the top 5% lincRNAs as hub lincRNAs in each module. Totally, we identified 5 and 3 hub lincRNAs in green and green-yellow modules, respectively. We constructed two correlation networks between these hub lincRNAs and protein-coding genes, which co-expressed with them by Cytoscape, to further clarify the function of hub lincRNAs in two modules (Figure 5 A, B, Supplementary Table S7). Moreover, the enrichment pathways of protein-coding genes in module green mainly involve methylation-dependent chromatin silencing, transcription from polymerase II promoter, and mRNA. Some of these genes have a vital role on abnormal parthenogenesis embryonic development, such as *PPP2CA, NANOG, DMXL2, GATM, API5* and *WDR36*(Esposito et al., 2019; Gallenberger et al., 2011; Mayank et al., 2015; Pan et al., 2015; Veil et al., 2018; Zhou et al., 2007). Other genes such as *DDX5, POLG2, SPATA22, CKS2, ORC4,* and *SPC25* have an important role on transcriptome and cell development in early embryonic development(Humble et al., 2013; Ishishita et al., 2014; Legrand et al., 2019; Martinsson-Ahlzen et al., 2008; Nguyen et al., 2015; Wang et al., 2019a). In module green-yellow, the enrichment pathways of protein-coding gens mostly include multicellular organism development and regulation of transcription. The *TNFAIP6, DIS3L2, PAK5, DDB1*, and *KDM8* genes in this module had important roles on cell proliferation in early embryonic development(Cang et al., 2006; Ishimura et al., 2012; Luo et al., 2007; Towler et al., 2016; Wojtanowicz-Markiewicz et al., 2019). These related genes have a high correlation to hub lincRNAs over 0.9 and have a highly similar expression trend with hub lincRNAs. Hence, these five hub lincRNAs in module green on abnormal transcriptome and basal metabolism and the three hub lincRNAs in module green-yellow may participate in cellular development by regulating cell proliferation. Although knowledge about how these hub lincRNAs involved in embryonic development are quite limited, these lincRNAs would serve as ideal candidates for further functional studies.

5. Detection of AS events

Alternative splicing (AS) increases regulatory complexity, controls the developmental programs in many species, and is closely correlated to stem cell lineage differentiation, suggesting its key role for splicing in controlling gene regulatory networks(Tahmasebi et al., 2016; Yang et al., 2016a; Zhang et al., 2016). AS is also required for cell differentiation in embryonic development. According to the reconstructed

transcriptome, 100,538 protein-coding transcripts corresponded to 15,322 genes (68.6% of the 22,342 annotated genes in Ssc11.1), indicating an average of 7 isoforms per gene. The average transcript number of AS genes in all developmental stages ranged from 3 to 5 (Figure 6A). In general, the number of AS genes presented a significantly increasing trend from mature oocyte to zygote stage in IVF and PA and then decreased gradually during embryonic development in IVF. However, the number AS genes in PA increased steadily until the morula stage and then exhibited a sharp decrease in the early blastocyst stage (Figure 6B). This finding suggests that a large transcriptomic change occurred from the zygote to morula stage in PA. Furthermore, we performed a comparative analysis of alternative splicing rate between protein-coding genes, DE protein-coding genes, lincRNAs and DE lincRNAs. Alternative splicing rate of protein-coding genes shows significantly higher than that of lincRNAs (protein-coding genes vs lincRNAs: 85.0% vs 8.0%). While there is almost no difference between lincRNAs and DE lincRNAs (lincRNAs vs DE lincRNAs: 8% vs 9%, that of DE protein-coding genes have slightly different from protein-coding genes (protein-coding genes vs DE protein-coding genes: 83.0% vs 89.0%) (Figure 6 C, D). It indicated that alternative splicing events have a crucial role in regulation of early embryonic development in pigs, especially in protein-coding genes level.

By using the ASProfile program, we analyzed 12 types of basic AS events. We detected 14,069 and 14131 AS genes in IVF and PA, respectively, with 98.5% similarity (16995 common AS genes). The frequencies of all AS types were similar to each other among all development stages. Among the AS events identified, TSS and TTS were the most abundant types (TSS: 39%–40%; TTS: 37.2%–37.9%), followed by SKIP (nearly 9%), XSKIP and MSKIP (both nearly 2%), IR and XIR (both nearly 1%), and other AS types (all below 1%) (Supplementary Figure S4). This finding further suggests that mRNA processes related to TSS and TTS of alternative splicing events may occur in the transcription regulation of early embryonic development.

6. Comparative analysis of DEGs between IVF and PA

Inner cell mass (ICM) is often not obvious and only a few cells are present in porcine early embryos, which cause failure on pig parthenogenesis activation embryo and pose a huge challenge for parthenogenetic embryo stem cell (pESC) culture and breeding parthenogenetic pig(Hall and Hyttel, 2014; Park et al., 2009). Here, we performed differential expression analysis of protein-coding genes in the same stages between IVF and PA. We identified 92 (67 upregulated, 25 downregulated), 3,670 (2,731 upregulated, 939 downregulated), and 259 (212 upregulated, 47 downregulated) DEGs in the zygote, morula, and early blastocyst stages, respectively (Supplementary Table S7). Then, we performed GO enrichment analysis of these DEGs and found some pathways related to basic cell development and metabolic process in morula stage, such as regulation of transcription and translation, energy metabolism, and cell proliferation and differentiation. We also found other pathways related to apoptotic process in the morula stage, such as regulation of apoptotic process, extrinsic apoptotic signaling pathway, and regulation of cell death. In the early blastocyst stage, most of DEGs participated in transcription, cell proliferation and differentiation, cell cycle, and apoptotic-related process (Supplementary Table S7).

Furthermore, upregulated genes from the oocyte to morula stages shared some common pathways in IVF, such as mRNA splicing-related biological process, but these pathways were downregulated in the early blastocyst stage. By contrast, the upregulated genes from the oocyte to zygote stages show mRNA splicing-related pathway in PA, and downregulated genes showed mRNA splicing-related pathways from the zygote to early blastocyst stages. Next, we that found some of these mRNA splicing-related genes, such as *SRSF7, HNRNPA2B1*, and *IWS1(Despic et al., 2017; Kasowitz et al., 2018; Oqani et al., 2019)*, had vital effects on early embryonic development. Moreover, some biological process related to embryonic development have occurred in advance from morula to early blastocyst in PA, such as positive regulation of kidney development, retina development in camera-type eye, and inner ear receptor cell differentiation. The above results indicated that the pre-production of some biological process related to embryo development and the regulation of apoptotic process may mostly be the key points of parthenogenesis-activated embryo failure.

7. Quantitative Reverse-Transcription PCR validation of transcriptome sequencing results

8 genes (4 protein-coding genes and 4 lincRNAs) randomly selected from the comparative gene expression data were evaluated using SYBR Green I-based Real-time Quantitative PCR (QPCR). The primer sequences for all target genes are listed in Supplementary Table S8. All selected genes (*LDHB, HSPE, NUP35, C14orf166, MSTRG.12993, MSTRG.53456, MSTRG.58641, MSTRG.341*) showed a statistically consistency with our sequencing data (Figure 7), thereby further improving our research reliability.

Discussion

In this study, we developed a platform for mRNA sequencing during pig early embryonic development in IVF and PA. These results provide a comprehensive framework of transcriptome landscape for 21 single cells of mature oocyte and three typical stages (zygote, morula, and early blastocyst). Our major findings include the following:

According to the result of differential expression analysis, we found early embryonic development of IVF and PA had a significant difference on the most active stage of transcription. The largest number of DEGs and linRNAs was recorded in the early blastocyst stage (IVF) and morula stage (PA). Based on the enrichment pathways analysis of these two stages, most biological processes related to transcription, cell proliferation and differentiation, and energy metabolism both occurred in IVF and PA. However, some DEGs in PA were significantly involved in apoptotic processes, such as AIMP2, BIRC5, BNIP1, and BOK. Simultaneously, we detected some biological processes related to late-embryo development have occurred in advance from morula to early blastocyst in PA, and the expression of some corresponding genes decreased, such as positive regulation of kidney development, retina development in camera-type eye, and inner ear receptor cell differentiation.

The regulation of transcription, as well as mRNA processing are crucial for genome reprogramming and zygotic genome activation during early embryonic development. In our analysis of the enrichment

pathway of DEGs, we found that some DEGs shared a common mRNA splicing-related pathway. Furthermore, we performed AS analysis. AS is a mechanism of post-transcriptional RNA processing whereby a single gene can encode multiple distinct transcripts, and it is a major source of cellular protein diversity. Currently, reports on AS during early embryonic development are limited to a few species. However, no systematic survival analysis of alternative splicing on embryonic development in porcine has been reported. Here, we found that the number of AS genes presented a significantly increasing trend from MII to 1-cell stage both in IVF and PA, and the number of AS genes decreased gradually during embryonic development in IVF. However, the number of AS genes in PA continuously increased until the morula stage, and then sharply decreased in the early blastocyst, which have a highly correlation with the change of DEGs both in IVF and PA. Next, we investigated the alternative splicing rate of protein-coding genes, DE protein-coding genes, lincRNAs and DE lincRNAs and the frequency of 12 AS types. It shows DESs have slightly higher AS rate than protein-coding genes, and also shows the higher AS rate of protein-coding genes than lincRNAs. We found TSS and TTS are the most abundant types. Which suggested that AS has a role on regulating the process of early embryonic development in pigs.

A large number of lincRNAs can be found in mammalian genomes, and their exact number may be equal to or even surpass the number of protein-coding genes. We performed a comprehensive identification and analysis of putative lincRNAs by co-expression networks analysis. Here, we constructed and improved our new lincRNA identification pipeline through integration with previously published lincRNA identification procedures. Finally, we identified 7,185 novel putative lincRNAs, which have broadened the pig lincRNA annotation. Moreover, transcriptome analysis revealed that major bursts of transcription occurred at the morula stage in IVF and zygote stage in PA, which are largely similar to DEGs. GO and KEGG analysis of DEGs in these two periods provided a more global view of changes, yielding obvious enrichment in transcriptome, cell division, mRNA splicing and metabolic pathway in IVF, negative regulation of cell differentiation, gene silencing by miRNA, natural killer cell activation involved in immune response, and cysteine-type endopeptidase activity involved in apoptotic process in PA, which may correspond to subsequent parthenogenesis failure. Subsequently, we constructed a co-expression network by using WGCNA to clearly understand the functional roles of lincRNAs. Finally, according to the enrichment analysis of protein-coding genes in each module, we found that green and green-yellow modules had similar roles on transcriptome and cell cycle. The green module also participated on mRNA splicing via spliceosome and methylation-dependent chromatin silencing. Therefore, the hub-lincRNAs in the green module participated in transcriptome, cell cycle-related biological process, as well as in the abnormal development of parthenogenesis embryos with a large possibility.

Notably, we identified the most DEGs in the morula stage in PA, indicating that a huge transcriptome change in PA occurs in the morula stage. In addition to some basic metabolic processes, most DEGs are enriched in some pathways, such as the regulation of apoptotic process, extrinsic apoptotic signaling pathway, and regulation of cell death. Hence, failure of parthenogenesis activation embryo was highly corresponded with apoptotic process and abnormal metabolism that some biological processes happened sooner rather than later. Furthermore, CDK1 and Programmed cell death 5 (PDCD5) significantly increased in the morula and early blastocyst stages in PA, which present a complete

difference in IVF. The monoallelic expression of CDK1(AF) is lethal in the early embryonic stage in mice and induces S phase arrest accompanied by gammaH2AX and DNA damage checkpoint activation in mouse embryonic fibroblasts (MEFs)(Szmyd et al., 2019). PDCD5 is an apoptosis promoter molecule that displays multiple biological activities. Knockout of the PDCD5 gene in mouse resulted in placental defects and embryonic lethality(Li et al., 2017). The expression of the genes related to apoptotic process increased in the morula stage, which could be the key point for parthenogenesis embryo failure.

Conclusion

In this study, we performed the transcriptome-wide gene expression dynamics in pig oocytes and early embryos to reveal protein-coding genes and lincRNAs involved in early embryonic development and parthenogenesis activation embryos failure. In summary, we explore the gene expression features of protein-coding genes, AS events, and lincRNAs in the early embryonic development. The comparison of in vitro fertilization (IVF) and parthenogenesis activation (PA) embryos revealed similarity and differences with or without the participation of paternal genome. Our work together uncovered candidate genes and lincRNAs to further verify its molecular mechanism, the early embryonic development laws of in vitro fertilization and parthenogenesis activation group, and speculates on mechanism of parthenogenesis activation embryo failure. This study will help future research on these genes and molecular-assisted breeding for pig parthenotes.

Methods

1. Animal ethics statement

All animal studies were performed according with the guidelines of the Key Lab of Agriculture Animal Genetics, Breeding, and Reproduction of Ministry of Education, Animal Care and Use Committee, Wuhan, 430070, China.

2. Collection and culture of pig oocytes and embryos

- Oocyte collection

In this study, porcine ovaries were obtained from a local slaughterhouse and transported to the laboratory within 3 hours while maintained at 38℃. After washing ovaries three times in warm physiological saline solution supplemented with streptomycin and penicillin, an 18-gauge needle attached to a 10-mL disposable syringe was used to aspirate the follicular fluid from 3 to 6-mm follicles. Cumulus oocyte complexes(COCs) with uniform cytoplasm and several layer of cumulus cells were selected and 3-5 layers of cumulus cells were selected, and then washed twice in wash buffer. In vitro maturation, conducted in 4-cell dishes in TCM-199 culture medium, included incubation at 38.5℃ with 5% CO2 in air, at maximum humidity, for 42-44 hr(Tao et al., 2017).

- In vitro fertilization

Groups of 30 denuded oocytes were washed three times in vitro fertilization (IVF) medium, and then placed in 100 ul drops of in vitro fertilization (IVF) medium covered with paraffin oil. Which were held at 38.5℃ in an atmosphere of 5% CO2 in air for approximately 30 min until the addition of spermatozoa. Fresh semen was washed three times by centrifugation 1900 × g for 3 min in Dulbecco's PBS. The resulting pellet was re-suspended in fertilization medium. The spermatozoa suspension (50ul) was added to a drop of IVF medium containing the oocytes. The oocytes were incubated with the sperm for 10 hr at 38.5°C with 5% CO2 in air(Tao et al., 2017).

- Parthenogenetic activation

The denuded oocytes having the first polar body (MII) were placed between 0.2-mm-diameter platinum electrodes 1 mm apart in activation medium. Activation was induced with two direct-current (DC) pulses of 1.2 kV/cm for 40 μs on a BTX Elector-Cell Manipulator 200 (BTX, San Diego, CA) according to the experimental design. The medium used for activation was 0.3M mannitol, supplemented with 1.0 mM CaCl2, 0.1 mM MgCl2, and 0.5 mM Hepes. The orientation of oocytes and polar bodies was not vertical to platinum wire electrodes during electrical activation. The parthenogenesis activation (PA) embryos were washed and exposed to NCSU medium with 5 μg/ml cytochalasin B for 4 hr to inhibit second polar body extrusion, then cultured in 150 μl NCSU medium covered with mineral oil in a 96-well culture plate. The embryos were incubated at 38.5℃ with 5% CO2 in air(Wang et al., 2017).

Single cells of porcine MII, 1-cell, morula and blastocyst stage derived from in vitro fertilization and parthenogenesis activation were collected as described previously and stored individually. The day of in vitro fertilization is set to day 0. All embryo samples were placed on dry ice immediately after collection and stored at -80℃.

3. Preparation of cDNA and transcription profiling

Total RNA was extracted from 21 single cell using a RNA extraction kit (SMART-Seq® v4 Ultra™ Low Input RNA Kit（Takara, CA, USA）) according to the manufacturer instruction. RNA purity was qualified using NanoDrop ND 2000 spectrophotometer at 260 and 280 nm (Thermo Fisher Scientific, Wilmington, MA, USA), and RNA integrity was vertified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The OD260/280 ratios of all the samples were greater than 1.8, and the RIN values were greater than 7. The sequencing process was performed on an Illumina HiSeq 2000 instruments, Novogene, China.

4. Quality control (QC) and transcriptome assembly

The raw reads were cleaned by filtering the adapter and low-quality reads by Trimmomatic (version 0.3.2) (Bolger et al., 2014). First, the adaptors were removed; then, low quality reads (the number of mismatch > 2 in a read) were removed, and the reads in which the average quality of four continuous bases were < 15 were discarded. Then, the clean reads were mapped to the pig reference genome (sus scrofa 11.1) by HISAT2(version 2.0.1). According to different database construction methods, we selected the

corresponding parameters “known-splicesite-infile” and got the sam file containing all mapping readings for each sample. Using samtools (version 1.3.1), the sam files were processed and converted into sorted bam files with parameters “view –H” and “sort –o”. finally, to facilitate transcript assembly and quantification, all bam files were assembled into one complete GTF files with the default parameters and merge function of StringTie (version 1.3.5)(Contreras-Lopez et al., 2018).

5. Expression calculation and differentially expressed analysis

Read count tables were generated from binary sequence alignment/map (BAM) files using HTseq software. The value for fragments per kilobase of exon per million fragments mapped (FPKM)(Trapnell et al., 2010) was calculated to estimate gene expression under each sample following the rules below:

$$FPKM = \frac{Read\ count \times 10^6}{Total\ reads\ mapped\ to\ genome \times gene\ length\ (kb)}$$

The DSeq2 software package was used to calculate differences in gene expression of IVF and PA, and the same stage between IVF and PA. The differentially expressed transcripts with FoldChange greater (less) than 1 (-1) and an adjusted p-value less than 0.05, are considered to be significant(Contreras-Lopez et al., 2018).

6. Identification of alternative splicing (AS) and lincRNA

To determine alternative splicing events, Asprofile (v1.0.4) software(Florea et al., 2013) was used to classify in all samples to determine exons included in one transcript and skipped in the other, and then count the alternative splicing events in each sample. In addition, TBtools was used for the statistics of the number of genes with alternative splicing in each sample(Chen et al., 2018).

Moreover, by using the assembled GTF files, StringTie software is able to estimate the expression levels of genes and transcripts in all samples for subdequent studies with the parameters “-e” and “-B”(Contreras-Lopez et al., 2018). And the assembled transcriptome was then filtered to get the putative lincRNAs. Our pipeline for lincRNA identification as shown in Figure 4A was based on the way described in our previous study.

7. Function analysis of lincRNAs

- Neighboring gene analysis

For each DEL locus the nearest protein-coding neighbor within < 100,000 bp was identified by BEDtools (version 2.17.0), and these neighboring genes were expressed in at least one sample. This resulted in a list of lncRNA protein-coding loci pairs. Pearson correlation was used to explore the expression-based relationship between these pairs. Ensemble was used to convert the protein-coding gene id into human gene id for annotation, visualization and integrated discovery (DAVID). Then, we performed GO analysis

and KEGG pathway analysis on all expressed protein-coding gene of DELs. Terms with a p-value < 0.05 were considered statistically significant.

- Weighted gene co-expression network analysis

Co-expression networks were constructed by WGCNA (version 3.5.2) (Langfelder and Horvath, 2008) package in RStudio environment using two parts of genes (differentially expressed putative lincRNAs, differentially expressed protein-coding genes). A signed weighted correlation network was constructed by creating a matrix of pairwise Pearson correlation coefficients, and module detection was used by the step=by-step signed network construction with a soft threshold of power = 18. Then, the topological overlap distance calculated from the adjacency matrix is then clustered with the average linkage hierarchical clustering. We retrieved the protein-coding genes that co-expressed with lincRNAs in each module, then GO enrichment and pathway analysis were performed on them. The minimum module size war set to 50 to ensure a qualified number of genes for further analysis. For each module, we defined the first principle component as the eigengene according to WGCNA terminology. To detect the relationship between modules and seven development stages, we defined a vector to encode seven development stages. Then, we correlated this vector with the eigengenes of each module, and a higher correlation indicated that the module was related to corresponding development stage.

- Identification of hub lincRNAs

Four modules were finally chosen for identifing of hub lincRNAs, we calculated the connectivity of each gene based on their intra-module connectivity. LincRNAs with top 5% intra-module connectivity were defined as hub lincRNAs, networks of which were displayed graphically using Cytoscape (version 3.7.0) software. The correlation between hub lincRNAs and co-expression protein-coding genes must be larger than 0.9.

8. Gene ontology and pathway analysis

We performed DAVID analysis by running queries for each protein-coding gene against the DAVID database. Because the annotation for the genes in sus scrofa 11.1 is relatively limited, we used the BIOMART system to abstract orthologous relationships between pig and human gene pairs and convert mapping information regarding gene identities between human ensemble and reference sequences.

9. Validation of RNA-seq data by Quantitative Reverse-Transcription PCR

To validate the RNA-Seq results, we performed qRT-PCR with SYBR Green I-based Real-time Quantitative PCR (QPCR) (CWBIO, Beijing, China, CW0957). Ten pairs of primers for qRT-PCR were designed using Primer 5 program (Supplementary Table S8). The 18s rRNA served as the endogenous control gene in the Roche LightCyler 480 system (Roche, Mannheim, Germany). The qRT-PCR data were analyzed using the delta delta Ct ([Delta][Delta]Ct) method. We used the trend of gene expression during total stages to judge whether qRT-PCR results were in accordance with the RNA-Seq results.

Declarations

List of abbreviations:

AS	Alternative splicing
lincRNA	Long intergenic non-coding RNAs
IVF	In vitro fertilization
PA	Parthenogenesis activation
TSS	Alternative 5' first exon (transcription start site)
TTS	Alternative 3' last exon (transcription terminal site)
SKIP	Skipped exon
XSKIP	Approximate SKIP
MSKIP	Multi-exon SKIP
XMSKIP	Approximate MSKIP
IR	Intron retention
XIR	Approximate IR
MIR	Multi-IR
XMIR	Approximate MIR
AE	Alternative exon ends
XAE	Approximate AE
MII	Meiosis II
PTG	Putative targeted gene
DEL	Differential expression lincRNA
DEG	Differential expression gene
Stem	The short time-series expression miner
WGCNA	Weighted gene co-expression network analysis

ICM	Inner cell mass
pESC	Parthenogenesis embryo stem cell
QPCR	Real-time Quantitative PCR
COC	Cumulus oocyte complexes
DC	Direct-current
QC	Quality control
DE protein-coding genes	Differential expression protein-coding genes
DE lincRNA	Differential expression lincRNA

Ethics approval and consent to participate

Animal slaughtering and sample collection were carried out according with the pre-approval guidelines of the Standing Committee of the Hubei Provincial People's Congress No. 5. All animal studies were performed according with the guidelines of the Key Lab of Agriculture Animal Genetics, Breeding, and Reproduction of Ministry of Education, Animal Care and Use Committee, Wuhan, 430070, China.

Author Contributions

CCL conceived and supervised the study. CCL, XDL and SHZ made manuscript revisions. XL performed experiments and analyzed the RNA-seq data and wrote the manuscript with ZYH, CZ and MXL's helps. All authors reviewed the results and approved the final version of the manuscript.

Funding

The work was supported by National Natural Science Foundation of China under Grant No.31872322; the Fundamental Research Funds for the Central Universities under Grant No.2662017PY030; National Key Project of Transgenic Animal and Crops of Ministry of Science and Technology of China under Grant No. 2016ZX08006-003.

Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential competing interests.

Availability of data and meterials

All the RNA-seq data generated in this study have been deposited in the GEO database under the accession number PRJNA595438.

Acknowledgements

Not applicable

References

1. Ahn, R., Gupta, R., Lai, K., Chopra, N., Arron, S. T., and Liao, W. (2016). Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs. *BMC Genomics* **17**, 841.
2. Atianand, M. K., Hu, W., Satpathy, A. T., Shen, Y., Ricci, E. P., Alvarez-Dominguez, J. R., Bhatta, A., Schattgen, S. A., McGowan, J. D., Blin, J., Braun, J. E., Gandhi, P., Moore, M. J., Chang, H. Y., Lodish, H. F., Caffrey, D. R., and Fitzgerald, K. A. (2016). A Long Noncoding RNA lincRNA-EPS Acts as a Transcriptional Brake to Restrain Inflammation. *Cell* **165**, 1672-1685.
3. Blank-Giwojna, A., Postepska-Igielska, A., and Grummt, I. (2019). lncRNA KHPS1 Activates a Poised Enhancer by Triplex-Dependent Recruitment of Epigenomic Regulators. *Cell Rep* **26**, 2904-2915.e4.
4. Boije, H., Ring, H., Shirazi Fard, S., Grundberg, I., Nilsson, M., and Hallbook, F. (2013). Alternative splicing of the chromodomain protein Morf4l1 pre-mRNA has implications on cell differentiation in the developing chicken retina. *J Mol Neurosci* **51**, 615-28.
5. Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20.
6. Brevini, T. A., Antonini, S., Cillo, F., Crestan, M., and Gandolfi, F. (2007). Porcine embryonic stem cells: Facts, challenges and hopes. *Theriogenology* **68 Suppl 1**, S206-13.
7. Brevini, T. A., and Gandolfi, F. (2008). Parthenotes as a source of embryonic stem cells. *Cell Prolif* **41 Suppl 1**, 20-30.
8. Cang, Y., Zhang, J., Nicholas, S. A., Bastien, J., Li, B., Zhou, P., and Goff, S. P. (2006). Deletion of DDB1 in mouse brain and lens leads to p53-dependent elimination of proliferating cells. *Cell* **127**, 929-40.
9. Casero, D., Sandoval, S., Seet, C. S., Scholes, J., Zhu, Y., Ha, V. L., Luong, A., Parekh, C., and Crooks, G. M. (2015). Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol* **16**, 1282-91.
10. Chen, C., Xia, R., Chen, H., and He, Y. (2018). TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv* 289660.
11. Chen, X., Zhu, Z., Yu, F., Huang, J., Jia, R., and Pan, J. (2019). Effect of shRNA-mediated Xist knockdown on the quality of porcine parthenogenetic embryos. *Dev Dyn* **248**, 140-148.
12. Choi, J. W., Kim, D. G., Park, M. C., Um, J. Y., Han, J. M., Park, S. G., Choi, E. C., and Kim, S. (2009). AIMP2 promotes TNFalpha-dependent apoptosis via ubiquitin-mediated degradation of TRAF2. *J Cell Sci* **122**, 2710-5.

13. Cibelli, J. B., Cunniff, K., and Vrana, K. E. (2006). Embryonic stem cells from parthenotes. *Methods Enzymol* **418**, 117-35.
14. Contreras-Lopez, O., Moyano, T. C., Soto, D. C., and Gutierrez, R. A. (2018). Step-by-Step Construction of Gene Co-expression Networks from High-Throughput Arabidopsis RNA Sequencing Data. *Methods Mol Biol* **1761**, 275-301.
15. Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzel, L., Straube, K., Gerstein, M. B., Butter, F., and Neugebauer, K. M. (2017). Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Res* **27**, 1184-1194.
16. Dua, D., Nagoorvali, D., Chauhan, M. S., Palta, P., Mathur, P., and Singh, M. K. (2019). Calcium ionophore enhanced developmental competence and apoptotic dynamics of goat parthenogenetic embryos produced in vitro. *In Vitro Cellular & Developmental Biology - Animal* **55**, 159-168.
17. Esposito, A., Falace, A., Wagner, M., Gal, M., Mei, D., Conti, V., Pisano, T., Aprile, D., Cerullo, M. S., De Fusco, A., Giovedi, S., Seibt, A., Magen, D., Polster, T., Eran, A., Stenton, S. L., Fiorillo, C., Ravid, S., Mayatepek, E., Hafner, H., Wortmann, S., Levanon, E. Y., Marini, C., Mandel, H., Benfenati, F., Distelmaier, F., Fassio, A., and Guerrini, R. (2019). Biallelic DMXL2 mutations impair autophagy and cause Ohtahara syndrome with progressive course. *Brain* **142**, 3876-3891.
18. Florea, L., Song, L., and Salzberg, S. L. (2013). Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res* **2**, 188.
19. Gallenberger, M., Meinel, D. M., Kroeber, M., Wegner, M., Milkereit, P., Bosl, M. R., and Tamm, E. R. (2011). Lack of WDR36 leads to preimplantation embryonic lethality in mice and delays the formation of small subunit ribosomal RNA in human cells in vitro. *Hum Mol Genet* **20**, 422-35.
20. Guo, X., Wang, Z., Lu, C., Hong, W., Wang, G., Xu, Y., Liu, Z., and Kang, J. (2018). LincRNA-1614 coordinates Sox2/PRC2-mediated repression of developmental genes in pluripotency maintenance. *J Mol Cell Biol* **10**, 118-129.
21. Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., Young, G., Lucas, A. B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J. L., Root, D. E., and Lander, E. S. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300.
22. Hall, V. (2008). Porcine embryonic stem cells: a possible source for cell replacement therapy. *Stem Cell Rev* **4**, 275-82.
23. Hall, V. J., and Hyttel, P. (2014). Breaking down pluripotency in the porcine embryo reveals both a premature and reticent stem cell state in the inner cell mass and unique expression profiles of the naive and primed stem cell states. *Stem Cells Dev* **23**, 2030-45.
24. Humble, M. M., Young, M. J., Foley, J. F., Pandiri, A. R., Travlos, G. S., and Copeland, W. C. (2013). Polg2 is essential for mammalian embryogenesis and is required for mtDNA maintenance. *Hum Mol Genet* **22**, 1017-25.
25. Ishimura, A., Minehata, K., Terashima, M., Kondoh, G., Hara, T., and Suzuki, T. (2012). Jmjd5, an H3K36me2 histone demethylase, modulates embryonic cell proliferation through the regulation of Cdkn1a expression. *Development* **139**, 749-59.

26. Ishishita, S., Matsuda, Y., and Kitada, K. (2014). Genetic evidence suggests that Spata22 is required for the maintenance of Rad51 foci in mammalian meiosis. *Sci Rep* **4**, 6148.
27. Kasowitz, S. D., Ma, J., Anderson, S. J., Leu, N. A., Xu, Y., Gregory, B. D., Schultz, R. M., and Wang, P. J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet* **14**, e1007412.
28. Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol* **14**, 659-65.
29. Kono, T., Obata, Y., Wu, Q., Niwa, K., Ono, Y., Yamamoto, Y., Park, E. S., Seo, J. S., and Ogawa, H. (2004). Birth of parthenogenetic mice that can develop to adulthood. *Nature* **428**, 860-4.
30. Laporta, J., Driver, A., and Khatib, H. (2011). Short communication: expression and alternative splicing of POU1F1 pathway genes in preimplantation bovine embryos. *J Dairy Sci* **94**, 4220-3.
31. Legrand, J. M. D., Chan, A. L., La, H. M., Rossello, F. J., Anko, M. L., Fuller-Pace, F. V., and Hobbs, R. M. (2019). DDX5 plays essential transcriptional and post-transcriptional roles in the maintenance and function of spermatogonia. *Nat Commun* **10**, 2278.
32. Li, G., Xu, C., Lin, X., Qu, L., Xia, D., Hongdu, B., Xia, Y., Wang, X., Lou, Y., He, Q., Ma, D., and Chen, Y. (2017). Deletion of Pdcd5 in mice led to the deficiency of placenta development and embryonic lethality. *Cell Death Dis* **8**, e2811.
33. Li, Y.-P., Duan, F.-F., Zhao, Y.-T., Gu, K.-L., Liao, L.-Q., Su, H.-B., Hao, J., Zhang, K., Yang, N., and Wang, Y. (2019). A TRIM71 binding long noncoding RNA Trincr1 represses FGF/ERK signaling in embryonic stem cells. *Nature Communications* **10**, 1368.
34. Loewer, S., Cabili, M. N., Guttman, M., Loh, Y. H., Thomas, K., Park, I. H., Garber, M., Curran, M., Onder, T., Agarwal, S., Manos, P. D., Datta, S., Lander, E. S., Schlaeger, T. M., Daley, G. Q., and Rinn, J. L. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42**, 1113-7.
35. Lunney, J. K. (2007). Advances in swine biomedical model genomics. *Int J Biol Sci* **3**, 179-84.
36. Luo, D., Caniggia, I., and Post, M. (2014). Hypoxia-inducible regulation of placental BOK expression. *Biochem J* **461**, 391-402.
37. Luo, T., Xu, Y., Hoffman, T. L., Zhang, T., Schilling, T., and Sargent, T. D. (2007). Inca: a novel p21-activated kinase-associated protein required for cranial neural crest development. *Development* **134**, 1279-89.
38. Luo, X., and Liu, J. (2019). Transcriptome Analysis of Acid-Responsive Genes and Pathways Involved in Polyamine Regulation in Iron Walnut. *Genes (Basel)* **10**.
39. MacDonald, J. A., Kura, N., Sussman, C., and Woods, D. C. (2018). Mitochondrial membrane depolarization enhances TRAIL-induced cell death in adult human granulosa tumor cells, KGN, through inhibition of BIRC5. *J Ovarian Res* **11**, 89.
40. Martinsson-Ahlzen, H. S., Liberal, V., Grunenfelder, B., Chaves, S. R., Spruck, C. H., and Reed, S. I. (2008). Cyclin-dependent kinase-associated proteins Cks1 and Cks2 are essential during early embryogenesis and for cell cycle progression in somatic cells. *Mol Cell Biol* **28**, 5698-709.

41. Matsuda, T., Namura, A., and Oinuma, I. (2019). Dynamic spatiotemporal patterns of alternative splicing of an F-actin scaffold protein, afadin, during murine development. *Gene* **689**, 56-68.
42. Mayank, A. K., Sharma, S., Nailwal, H., and Lal, S. K. (2015). Nucleoprotein of influenza A virus negatively impacts antiapoptotic protein API5 to enhance E2F1-dependent apoptosis and virus replication. *Cell Death Dis* **6**, e2018.
43. Nguyen, H., Ortega, M. A., Ko, M., Marh, J., and Ward, W. S. (2015). ORC4 surrounds extruded chromatin in female meiosis. *J Cell Biochem* **116**, 778-86.
44. Nordin, M., Bergman, D., Halje, M., Engstrom, W., and Ward, A. (2014). Epigenetic regulation of the Igf2/H19 gene cluster. *Cell Prolif* **47**, 189-99.
45. Oqani, R. K., Lin, T., Lee, J. E., Kang, J. W., Shin, H. Y., and Il Jin, D. (2019). Iws1 and Spt6 Regulate Trimethylation of Histone H3 on Lysine 36 through Akt Signaling and are Essential for Mouse Embryonic Genome Activation. *Sci Rep* **9**, 3831.
46. Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**, 1413-5.
47. Pan, X., Chen, X., Tong, X., Tang, C., and Li, J. (2015). Ppp2ca knockout in mice spermatogenesis. *Reproduction* **149**, 385-91.
48. Park, S. K., Won, C., Choi, Y. J., Kang, H., and Roh, S. (2009). The leading blastomere of the 2-cell stage parthenogenetic porcine embryo contributes to the abembryonic part first. *J Vet Med Sci* **71**, 569-76.
49. Patterson, J. K., Lei, X. G., and Miller, D. D. (2008). The pig as an experimental model for elucidating the mechanisms governing dietary influence on mineral absorption. *Exp Biol Med (Maywood)* **233**, 651-64.
50. Smith, L. M., Kelleher, N. L., and Consortium for Top Down, P. (2013). Proteoform: a single term describing protein complexity. *Nat Methods* **10**, 186-7.
51. Suvorova, Il, Grigorash, B. B., Chuykin, I. A., Pospelova, T. V., and Pospelov, V. A. (2016). G1 checkpoint is compromised in mouse ESCs due to functional uncoupling of p53-p21Waf1 signaling. *Cell Cycle* **15**, 52-63.
52. Szmyd, R., Niska-Blakie, J., Diril, M. K., Renck Nunes, P., Tzelepis, K., Lacroix, A., van Hul, N., Deng, L. W., Matos, J., Dreesen, O., Bisteau, X., and Kaldis, P. (2019). Premature activation of Cdk1 leads to mitotic events in S phase and embryonic lethality. *Oncogene* **38**, 998-1018.
53. Tahmasebi, S., Jafarnejad, S. M., Tam, I. S., Gonatopoulos-Pournatzis, T., Matta-Camacho, E., Tsukumo, Y., Yanagiya, A., Li, W., Atlasi, Y., Caron, M., Braunschweig, U., Pearl, D., Khoutorsky, A., Gkogkas, C. G., Nadon, R., Bourque, G., Yang, X. J., Tian, B., Stunnenberg, H. G., Yamanaka, Y., Blencowe, B. J., Giguere, V., and Sonenberg, N. (2016). Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc Natl Acad Sci U S A* **113**, 12360-12367.

54. Tao, C., Li, J., Zhang, X., Chen, B., Chi, D., Zeng, Y., Niu, Y., Wang, C., Cheng, W., Wu, W., Pan, Z., Lian, J., Liu, H., and Miao, Y. L. (2017). Dynamic Reorganization of Nucleosome Positioning in Somatic Cells after Transfer into Porcine Enucleated Oocytes. *Stem Cell Reports* **9**, 642-653.

55. Towler, B. P., Jones, C. I., Harper, K. L., Waldron, J. A., and Newbury, S. F. (2016). A novel role for the 3'-5' exoribonuclease Dis3L2 in controlling cell proliferation and tissue growth. *RNA Biol* **13**, 1286-1299.

56. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5.

57. Veil, M., Schaechtle, M. A., Gao, M., Kirner, V., Buryanova, L., Grethen, R., and Onichtchouk, D. (2018). Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. *Development* **145**.

58. Wang, P., Wu, Y., Li, Y., Zheng, J., and Tang, J. (2013). A novel RING finger E3 ligase RNF186 regulate ER stress-mediated apoptosis through interaction with BNip1. *Cell Signal* **25**, 2320-33.

59. Wang, Q., Zhu, Y., Li, Z., Bu, Q., Sun, T., Wang, H., Sun, H., and Cao, X. (2019a). Up-regulation of SPC25 promotes breast cancer. *Aging (Albany NY)* **11**, 5689-5704.

60. Wang, T., Gao, Y. Y., Chen, L., Nie, Z. W., Cheng, W., Liu, X., Schatten, H., Zhang, X., and Miao, Y. L. (2017). Melatonin prevents postovulatory oocyte aging and promotes subsequent embryonic development in the pig. *Aging (Albany NY)* **9**, 1552-1564.

61. Wang, Z., Xu, H., Li, T., Wu, J., An, L., Zhao, Z., Xiao, M., Adu-Asiamah, P., Zhang, X., and Zhang, L. (2019b). Chicken GHR antisense transcript regulates its sense transcript in hepatocytes. *Gene* **682**, 101-110.

62. Wojtanowicz-Markiewicz, K., Kocherova, I., Jeseta, M., Piotrowska-Kempisty, H., Brussow, K. P., Skowronski, M. T., Bruska, M., Bukowska, D., Nowicki, M., Kempisty, B., and Antosik, P. (2019). Expression of PTX3, HAS2 AND TNFAIP6 genes in relation to real-time proliferation of porcine endometrial luminal epithelial cells in primary cultivation model. *J Biol Regul Homeost Agents* **33**, 675-685.

63. Yang, J., Margariti, A., and Zeng, L. (2016a). Analysis of Histone Deacetylase 7 (HDAC7) Alternative Splicing and Its Role in Embryonic Stem Cell Differentiation Toward Smooth Muscle Lineage. *Methods Mol Biol* **1436**, 95-108.

64. Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y. A., Murray, R. R., Spirohn, K., Begg, B. E., Duran-Frigola, M., MacWilliams, A., Pevzner, S. J., Zhong, Q., Trigg, S. A., Tam, S., Ghamsari, L., Sahni, N., Yi, S., Rodriguez, M. D., Balcha, D., Tan, G., Costanzo, M., Andrews, B., Boone, C., Zhou, X. J., Salehi-Ashtiani, K., Charloteaux, B., Chen, A. A., Calderwood, M. A., Aloy, P., Roth, F. P., Hill, D. E., Iakoucheva, L. M., Xia, Y., and Vidal, M. (2016b). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**, 805-17.

65. Zhang, T., Lin, Y., Liu, J., Zhang, Z. G., Fu, W., Guo, L. Y., Pan, L., Kong, X., Zhang, M. K., Lu, Y. H., Huang, Z. R., Xie, Q., Li, W. H., and Xu, X. Q. (2016). Rbm24 Regulates Alternative Splicing Switch in

Embryonic Stem Cell Cardiac Lineage Differentiation. *Stem Cells* **34**, 1776-89.

66. Zhou, C., Dobrinsky, J., Tsoi, S., Foxcroft, G. R., Dixon, W. T., Stothard, P., Verstegen, J., and Dyck, M. K. (2014). Characterization of the altered gene expression profile in early porcine embryos generated from parthenogenesis and somatic cell chromatin transfer. *PLoS One* **9**, e91728.
67. Zhou, Q. Y., Huang, J. N., Xiong, Y. Z., and Zhao, S. H. (2007). Imprinting analyses of the porcine GATM and PEG10 genes in placentas on days 75 and 90 of gestation. *Genes Genet Syst* **82**, 265-9.
68. Zhu, J., King, T., Dobrinsky, J., Harkness, L., Ferrier, T., Bosma, W., Schreier, L. L., Guthrie, H. D., DeSousa, P., and Wilmut, I. (2003). In vitro and in vivo developmental competence of ovulated and in vitro matured porcine oocytes activated by electrical activation. *Cloning Stem Cells* **5**, 355-65.
69. Zou, C., Li, L., Cheng, X., Li, C., Fu, Y., Fang, C., and Li, C. (2018). Identification and Functional Analysis of Long Intergenic Non-coding RNAs Underlying Intramuscular Fat Content in Pigs. *Front Genet* **9**, 102.

Supplementary Files Legend

Supplementary Figure S1 Pearson correlation between samples

Supplementary Figure S2 Differential expression analysis of group IVF and group PA. (A): Heat map of differential expressed genes among morula and early blastocyst stage in IVF, and functional enrichment map was shown B; (C): Heat map of differential expressed genes among zygote and morula stage in PA, and functional enrichment map was shown D.

Supplementary Figure S3 Characteristics and differential expression analysis of lincRNAs. (A): transcript length distribution of lincRNAs and protein coding gene; (B): Distribution of exon length in lincRNAs and protein coding gene; (C): Distribution of exon number in lincRNAs and protein coding gene. (D): Comparison of average expression level between lincRNAs and protein-coding genes. Histogram of differentially expressed known and novel lincRNAs between adjacent stages of group IVF (E) and group PA (F).

Supplementary Figure S4 Classification and quantitative statistics of alternative splicing (AS) events. Bar represents number of AS events; color represents different type of alternative splicing events.

Supplementary Table S1 Information of differential expression genes in this study

Supplementary Table S2 Information of putative lincRNAs (known) in this study

Supplementary Table S3 Information of putative lincRNAs (novel) in this study

Supplementary Table S4 Information of lincRNAs neighboring protein-coding genes

Supplementary Table S5 Gene ontology and pathway analysis of protein-coding neighbors

Supplementary Table S6 Information of six modules in WGCNA

Supplementary Table S7 Information of differential expression genes in same stage between IVF and PA

Supplementary Table S8 Information of primers for qRT-PCR

Table

Due to technical limitations, Table 1 is only available for download from the Supplementary Files section.

Figures

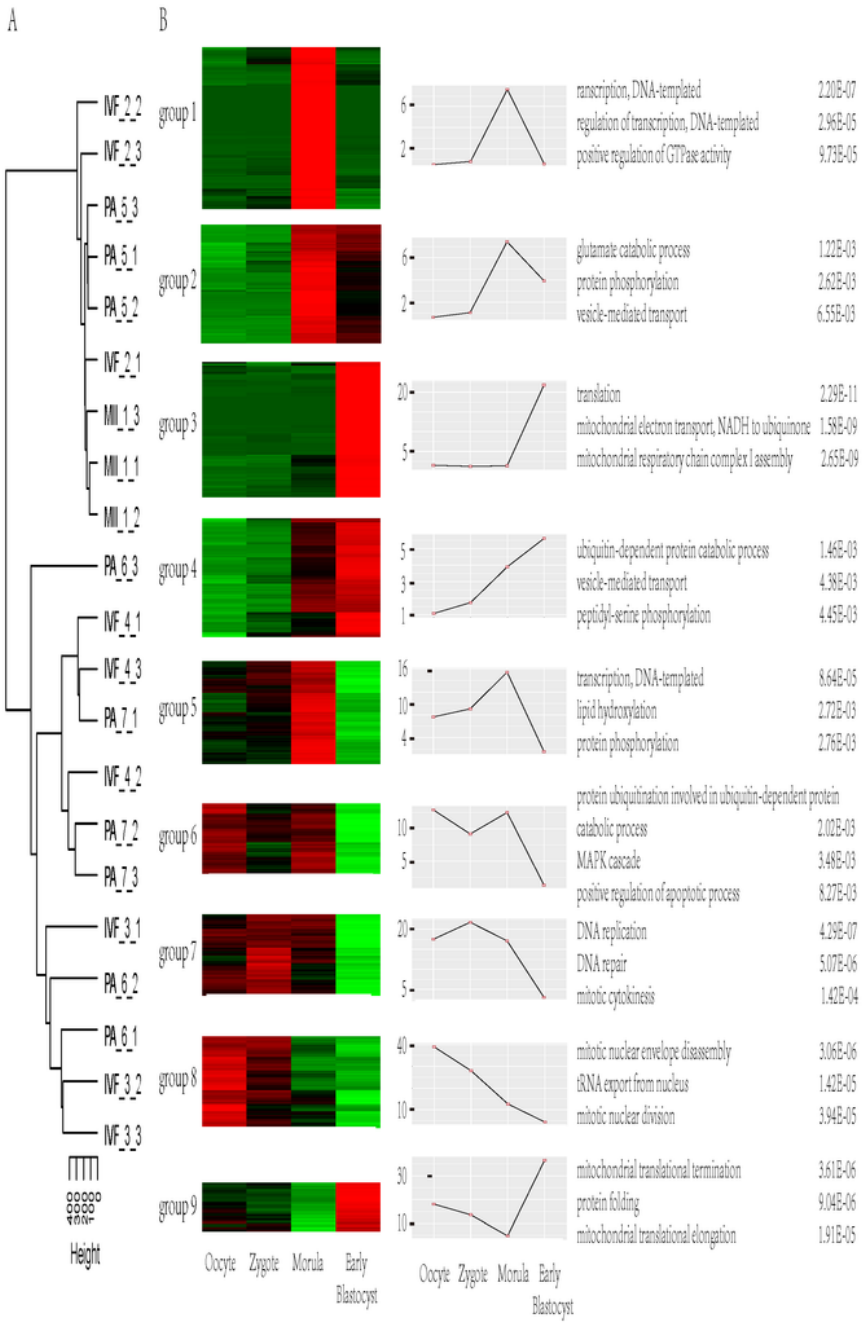

Figure 1

Unsupervised clustering of the expression profiles and DEG between adjacent stages of pig embryos in vitro fertilization (IVF). (A): Unsupervised hierarchical clustering of the expression profiles; (B): Clusters of DEGs between adjacent stages of pig embryos in vitro fertilization (IVF). Gene average log transformed expression values, top GO terms, and corresponding enrichment P-values were listed.

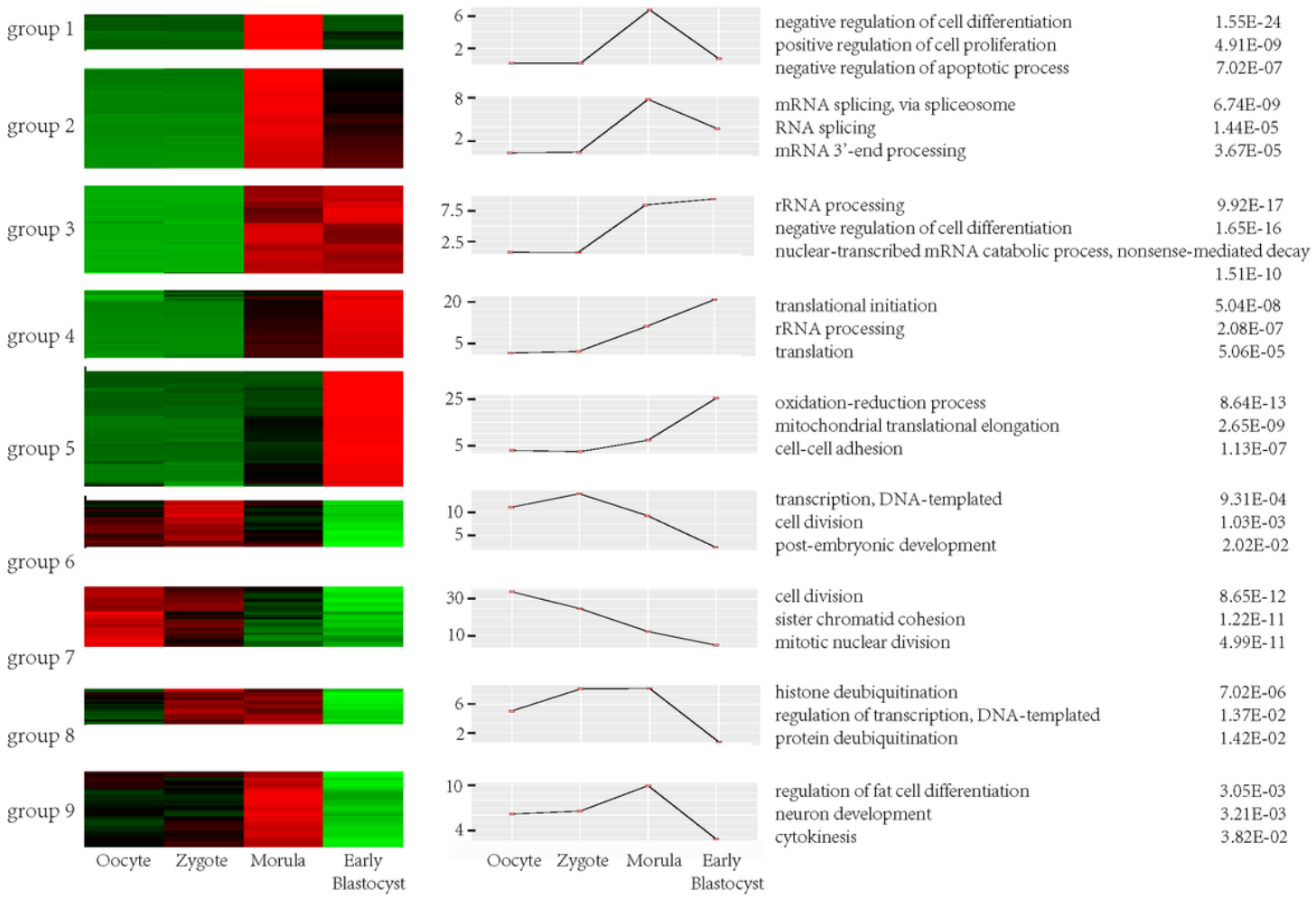

Figure 2

Clusters of DEGs between two adjacent stages of pig parthenogenesis activation (PA) embryos. Gene average log transformed expression values, top GO terms, and corresponding enrichment P-values were listed.

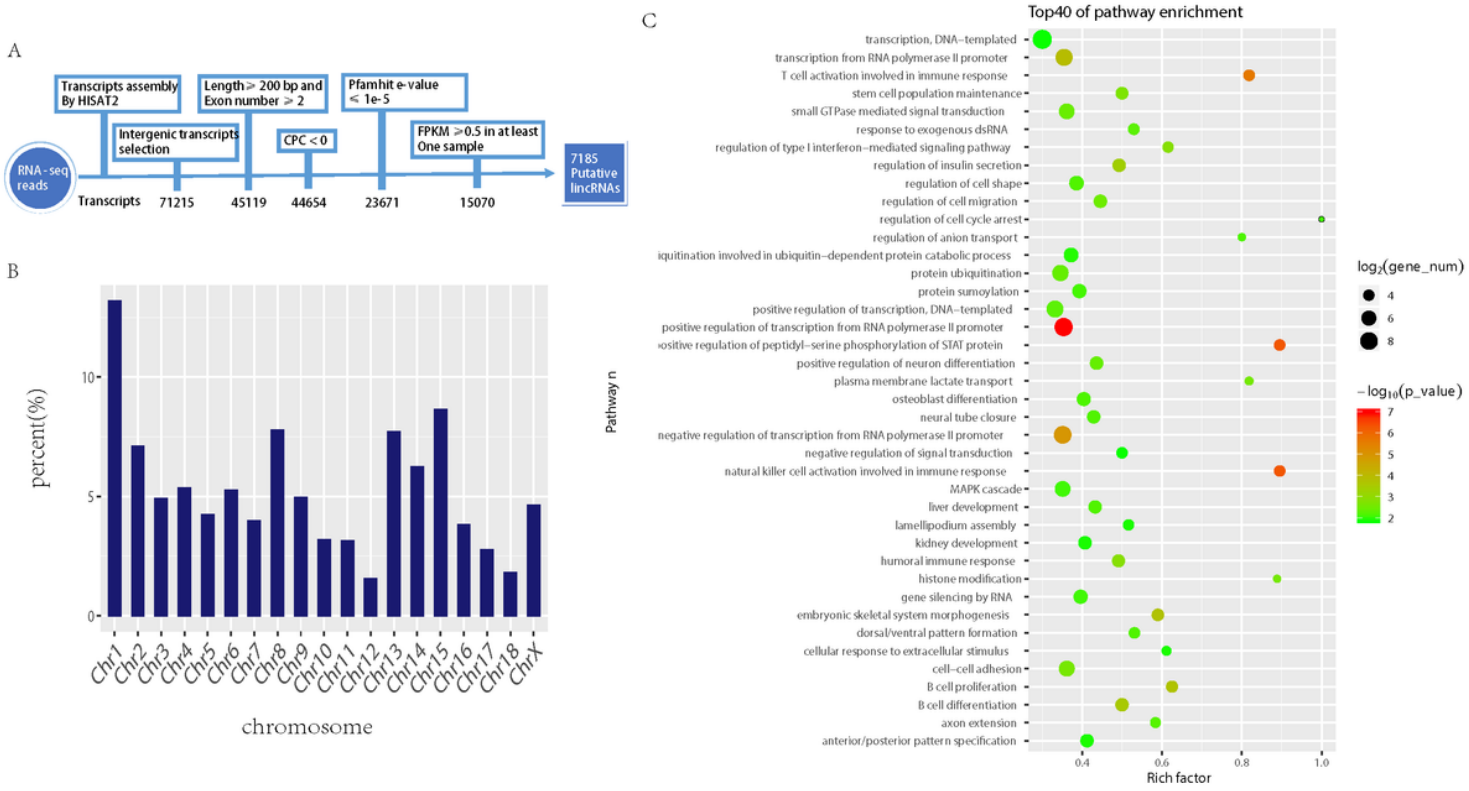

Figure 3

LincRNA identification process and distribution, and top Go terms of its neighboring protein-coding genes. (A): Workflow for lincRNA identification; (B): Distribution of LincRNA among different chromosome; (C): Top Go terms of its neighboring protein-coding genes.

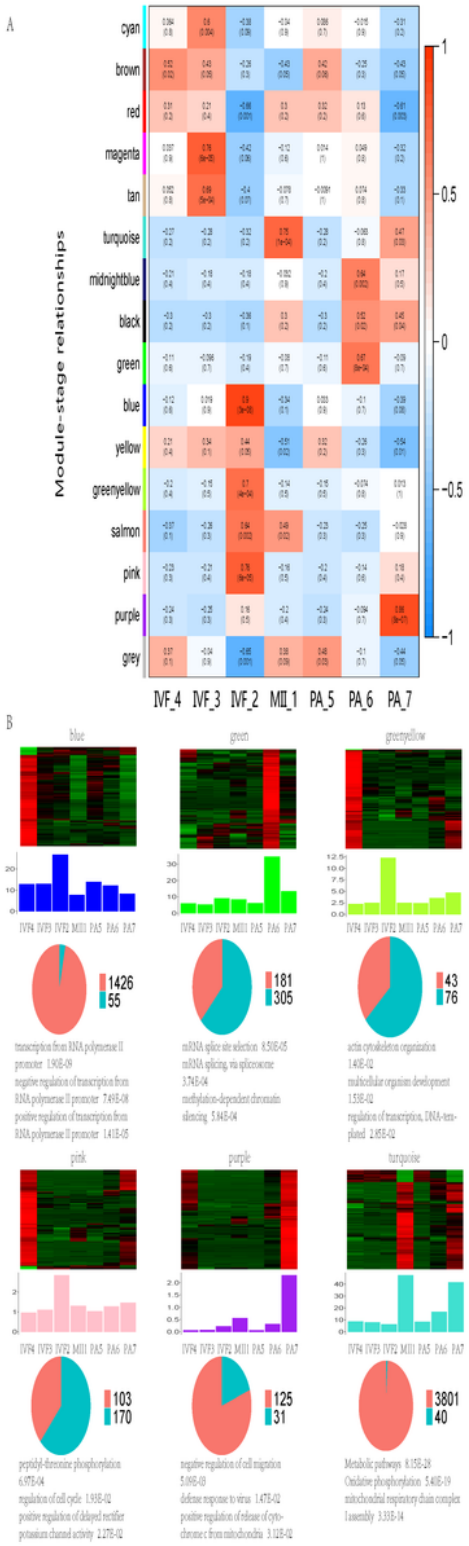

Figure 4

Module-stage correlation analysis and co-expression network analysis of differentially expressed novel lincRNAs and protein-coding genes. (A): Module-stage correlations and corresponding P values. On the left, different colors represent different modules; on the right, red indicates positive correlation, white indicates none correlation, blue indicates negative correlation; each cell contains the correlation and P value given in parentheses. (B): Co-expression networks of differentially expressed novel lincRNAs and

protein-coding genes in 6 modules, Top of each panel: heat maps for expression level of co-expressed genes in six modules. Red, increased expression; green, decreased expression. Middle of each panel: bar plots of the average expression of corresponding module eigengenes. Bottom of each panel: pie charts showing the abundance of lincRNAs and protein-coding genes and top Go terms of the later within each module.

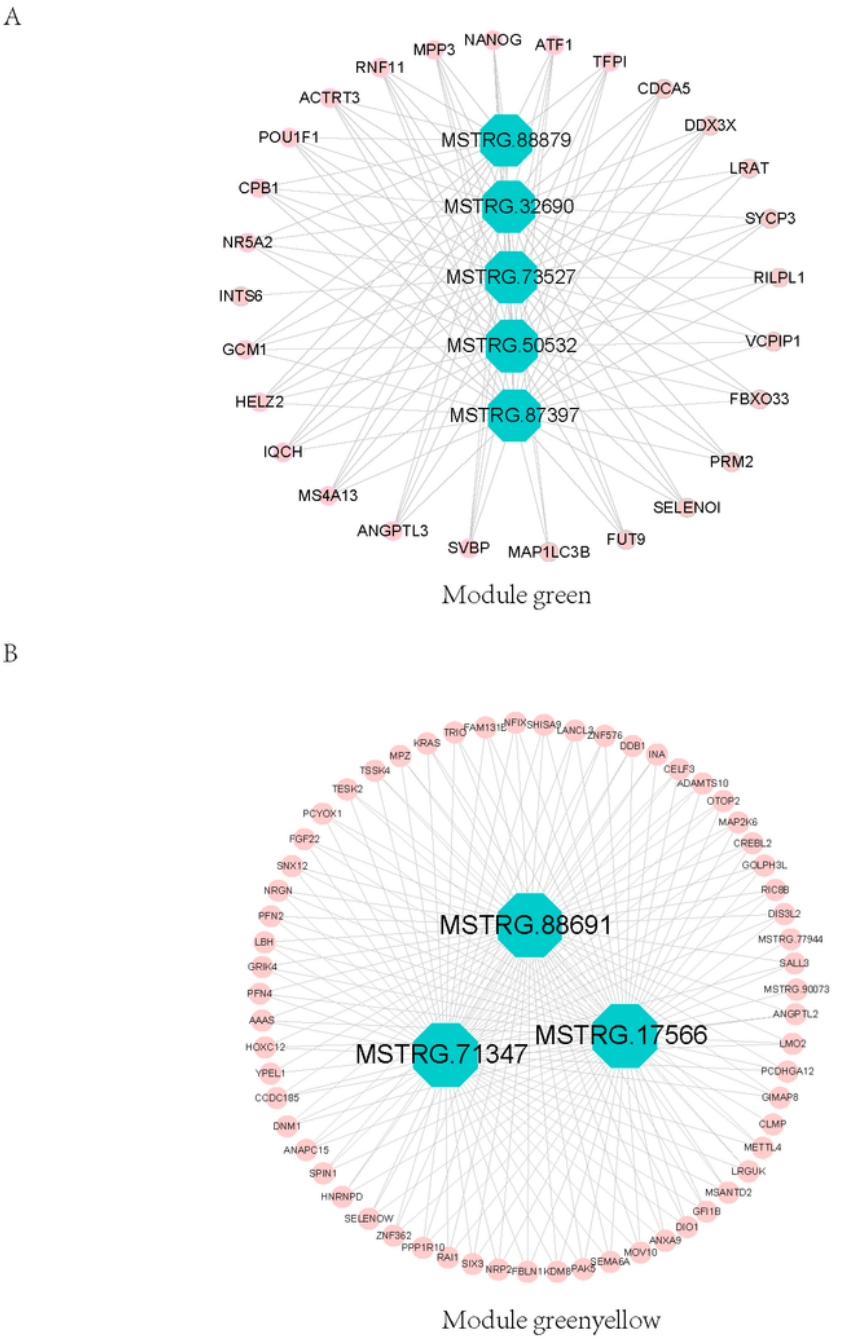

Figure 5

Co-expressed network visualization of hub lincRNAs and protein-coding genes in module green (A) and module greenyellow (B).

Figure 6

Identification of alternative splicing (AS) events in pig early embryonic development. (A): The average transcript number of alternative splicing genes; (B): Transcripts in each stage of AS genes (blue bar) and total transcripts of AS genes (red line); (C): Alternative splicing rate of lincRNA, DE lincRNA, protein-coding

gene and DEG; (D): Distribution of alternative splicing gene in DEGs and DE lincRNAs. Left histograms represent the count of DEGs (top) and DE lincRNAs (bottom) in different stage of group IVF; Middle histograms represent the count of DEGs (top) and DE lincRNAs (bottom) in different stage of group PA; Right histograms represent the count of DEGs (top) and DE lincRNAs (bottom) in the same stage between IVF and PA. Blue, alternative splicing gene; Orange, non-alternative splicing gene.

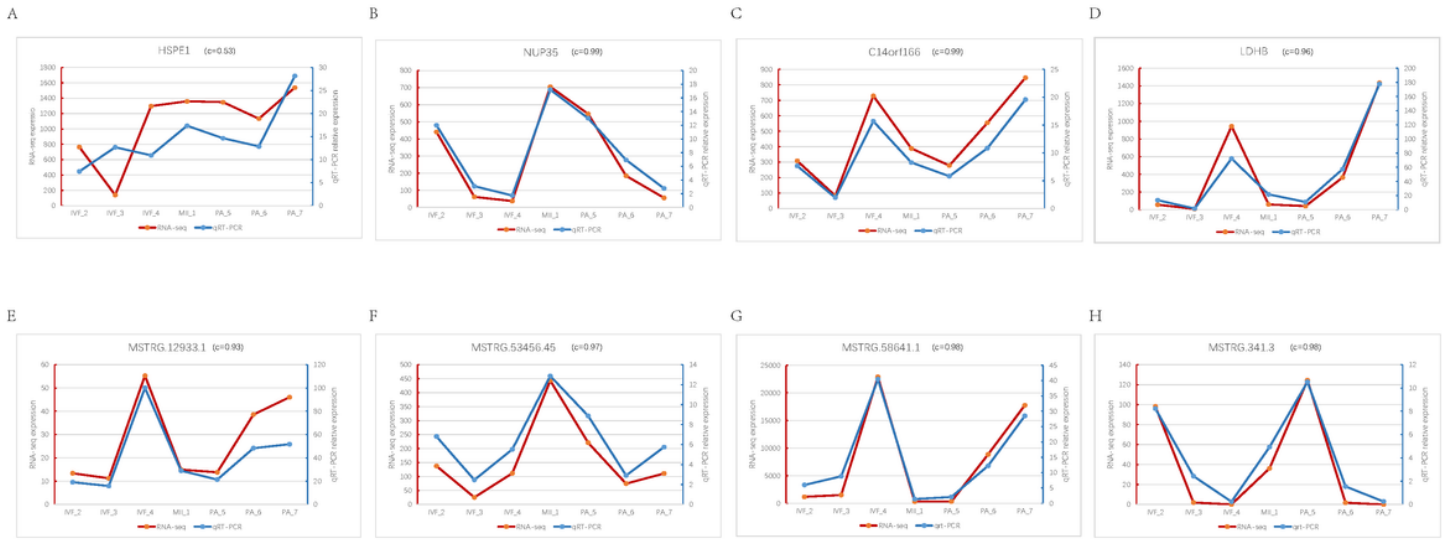

Figure 7

qRT-PCR validation of RNA-seq. qRT-PCR analysis of randomly selected protein-coding genes and lincRNAs were shown A-H. Red line represents average expression of RNA-seq; Blue line represents average expression of qRT-PCR, and R value is the correlation between RNA-seq and qRT-PCR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfigS4AScategory.pdf
- SupplementaryTableS1S2S3S4S5.csv
- supplementaryTableS6.csv
- supplementaryTableS7.csv

- supplementaryfigS2DEGIVFPA.pdf
- supplementaryTableS8.csv
- supplementaryfigS3AFlinccharacter.pdf
- TABLE1.pdf
- supplementaryfigS1pearsoncorrelationanalysis.pdf