

A Comparative Analysis of Single-Cell Transcriptome Identifies Reprogramming Driver Factors for Efficiency Improvement

Hanshuang Li,^{1,2} Mingmin Song,^{1,2} Wuru Yang,^{1,2} Pengbo Cao,¹ Lei Zheng,¹ and Yongchun Zuo¹

¹State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, College of Life Sciences, Inner Mongolia University, Hohhot 010070, China

Terminally differentiated somatic cells can be reprogrammed into a totipotent state through somatic cell nuclear transfer (SCNT). The incomplete reprogramming is the major reason for developmental arrest of SCNT embryos at early stages. In our studies, we found that pathways for autophagy, endocytosis, and apoptosis were incompletely activated in nuclear transfer (NT) 2-cell arrest embryos, whereas extensively inhibited pathways for stem cell pluripotency maintenance, DNA repair, cell cycle, and autophagy may result in NT 4-cell embryos arrest. As for NT normal embryos, a significant shift in expression of developmental transcription factors (TFs) *Id1*, *Pou6f1*, *Cited1*, and *Zscan4c* was observed. Compared with pluripotent gene *Ascl2* being activated only in NT 2-cell, *Nanog*, *Dppa2*, and *Sall4* had major expression waves in normal development of both NT 2-cell and 4-cell embryos. Additionally, *Kdm4b/4d* and *Kdm5b* had been confirmed as key markers in NT 2-cell and 4-cell embryos, respectively. Histone acetylases *Kat8*, *Elp6*, and *Eid1* were co-activated in NT 2-cell and 4-cell embryos to facilitate normal development. *Gadd45a* as a key driver functions with *Tet1* and *Tet2* to improve the efficiency of NT reprogramming. Taken together, our findings provided an important theoretical basis for elucidating the potential molecular mechanisms and identified reprogramming driver factor to improve the efficiency of SCNT reprogramming.

INTRODUCTION

Somatic cell nuclear transfer (SCNT) is a method to artificially reprogram the terminally differentiated somatic cells into totipotency embryos in addition to normal fertilization.¹ Dr. John Gurdon first demonstrated that animals can be cloned from the differentiated frog somatic cells by SCNT;² after 30 years of effort, the first cloned mammal Dolly was also born,³ and more than 20 mammals were successfully cloned using various types of somatic cells by SCNT.⁴ In addition to animal cloning, SCNT also has great potential in human therapy. For instance, nuclear transfer embryonic stem cells (ntESCs) derived from elderly and patient donor cells may serve as useful cell sources for disease modeling *in vitro* and cell or tissue replacement therapy.^{5,6} Overall, these studies indicated that the future development prospect for SCNT is attractive.

Despite the huge potential, several technical hurdles limit the practical application of SCNT. One of the limitations is that the survival rate of cloned animals is extremely low. In addition to cattle and sheep with relatively high reproductive cloning efficiency (5%–20%), the overall reproductive cloning efficiency of all other species is very low (1%–5%);⁷ even with the successful generation of SCNT blastocyst in rhesus monkeys, the blastocyst rate remained at about 16%.⁸ Moreover, in almost all of the cloned mammalian species, abnormalities in extra-embryonic tissue also have been frequently observed.⁶

Many studies have shown that SCNT embryo developmental arrest may be related to zygotic genome activation (ZGA) failure.^{5,9} Moreover, epigenetic abnormalities also are one of the main causes for abnormal development of SCNT embryos,^{10,11} in which H3K9me3 prevents the development of SCNT embryos at the pre-implantation stage in mice and humans by inhibiting the transcriptional activation of ZGA-related regions.^{1,12} Importantly, the heterotopic expression of H3K9me3 demethylase *Kdm4a/4b/4d* helps to eliminate the barrier of ZGA¹⁰ and to greatly promote the development of SCNT embryos.^{1,12,13} Furthermore, the loss of the H3K27me3 imprinting genes may cause abnormal activation of Xist, which leads to heterochromatin of the X chromosome,¹⁴ eventually hindering the development of SCNT embryos after implantation.^{5,15} Additionally, the correct regulation of some key genes in SCNT embryo development is also crucial for SCNT embryo development. Previous studies have identified a number of maladjusted genes in 2-cell SCNT embryos of mouse,¹⁶ and the numbers and roles of the these genes determine the fate of each cloned embryo.⁴ Therefore, comparative transcriptome analysis based on single-cell sequencing^{17–24} identifies reprogramming driver factors that may be a feasible method to improve the efficiency of cloning.

Received 14 November 2019; accepted 26 December 2019;
<https://doi.org/10.1016/j.omtn.2019.12.035>.

²These authors contributed equally to this work.

Correspondence: Yongchun Zuo, State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, College of Life Sciences, Inner Mongolia University, Hohhot 010070, China.

E-mail: yczuo@imu.edu.cn



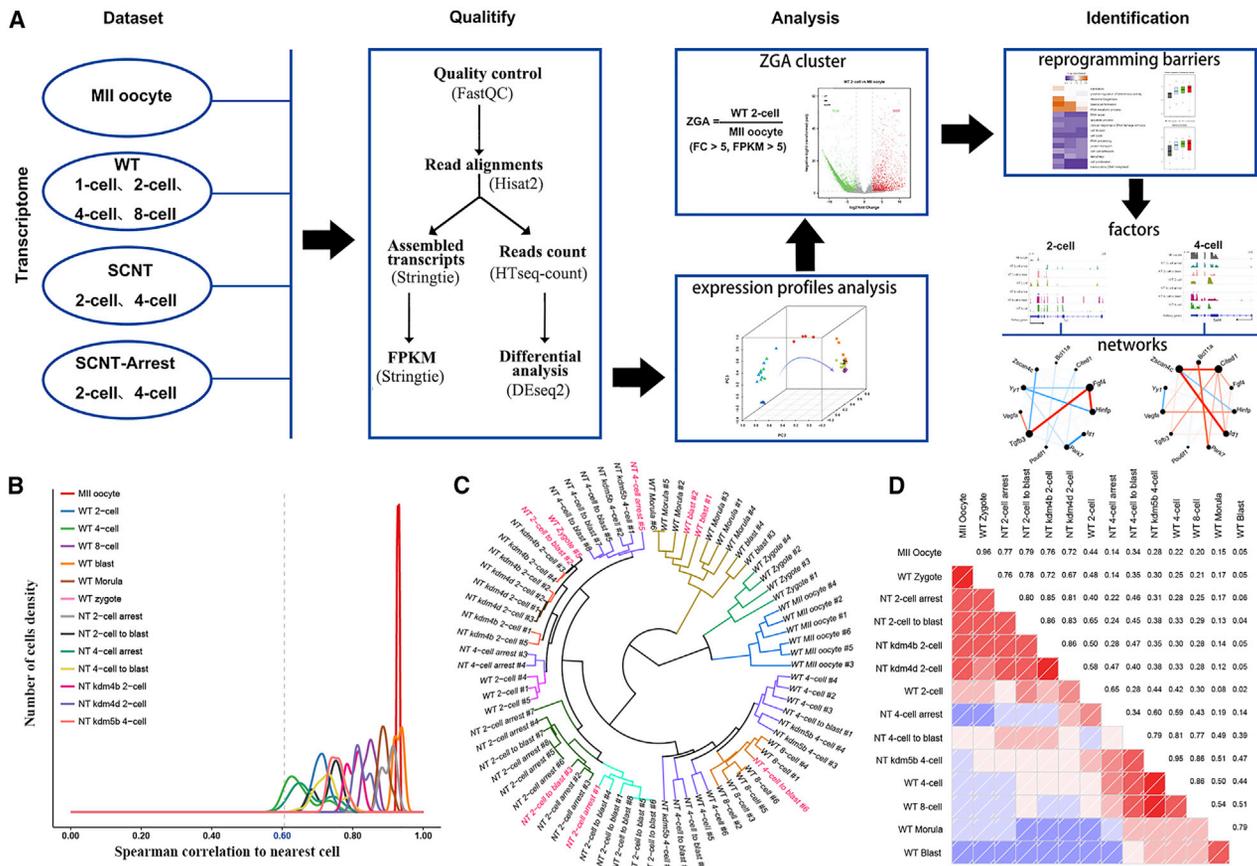


Figure 1. The Dynamic Changes of the Global Transcription Profile

(A) Flowchart of this study. (B) The intra-class correlation coefficients of expressed genes in normal and NT embryos. (C) The cluster analysis of the embryos samples with the different developmental fate. (D) The correlation of the embryo samples with a different developmental fate.

In our study, we first systematically analyzed the molecular differences of embryos with diverse developmental fates; the results show that the developmental fate of the embryos can be directly reflected in the transcriptome level. Next, we found some abnormal functional pathways in NT 2- and 4-cell arrest embryos, respectively. Finally, we comprehensively identified the reprogramming driver factor of SCNT embryos from transcription factors (TFs), pluripotent genes, and epigenetic modification factors. The results of this study will provide new insights into the molecules' mechanisms of SCNT reprogramming.

RESULTS

Global Profile and Pattern Analysis

To verify the reliability of transcriptome data of *in vitro* fertilization (wild-type [WT]) and NT embryos with different developmental fates, we analyzed the correlation of expression levels among the repeated samples within the group after quality control. The correlation between each repeat sample reached 0.60 (Figure 1B), showing the biological replicates of each sample are highly reproducible.

Based on the cluster relationship (Figure 1C) and correlation (Figure 1D) of the samples with different developmental fates, we found that the cells with the same development fate and stages can be clustered together, among which WT Metaphase II (MII) Oocyte had the closest relationship with WT Zygote (Figure 1C), and the correlation between them was also highest ($r = 0.96$) (Figure 1D). Interestingly, NT 2-cell *Kdm4b/4d* was close to WT 2-cell but far away from NT 2-cell arrest. The same result was also observed in NT 4-cell, indicating that histone demethylase can alter the developmental fate of NT embryos. Moreover, WT Morula was greatly related to WT Blastocyst ($r = 0.79$) (Figure 1D), and both can form an independent branch, showing that a unique developmental fate exists in Morula and blastocyst. These results displayed that the developmental fate of the embryos can be directly reflected in the transcriptome level, which can be used for further screening of molecular markers.

To further explore the dynamics of gene expression for NT or WT embryos with different developmental fate at 2-cell and 4-cell stages, we performed 28 pairwise comparisons on a total of eight embryos to obtain the differentially expressed genes (DEGs; $q < 0.05$, \log_2 (fold

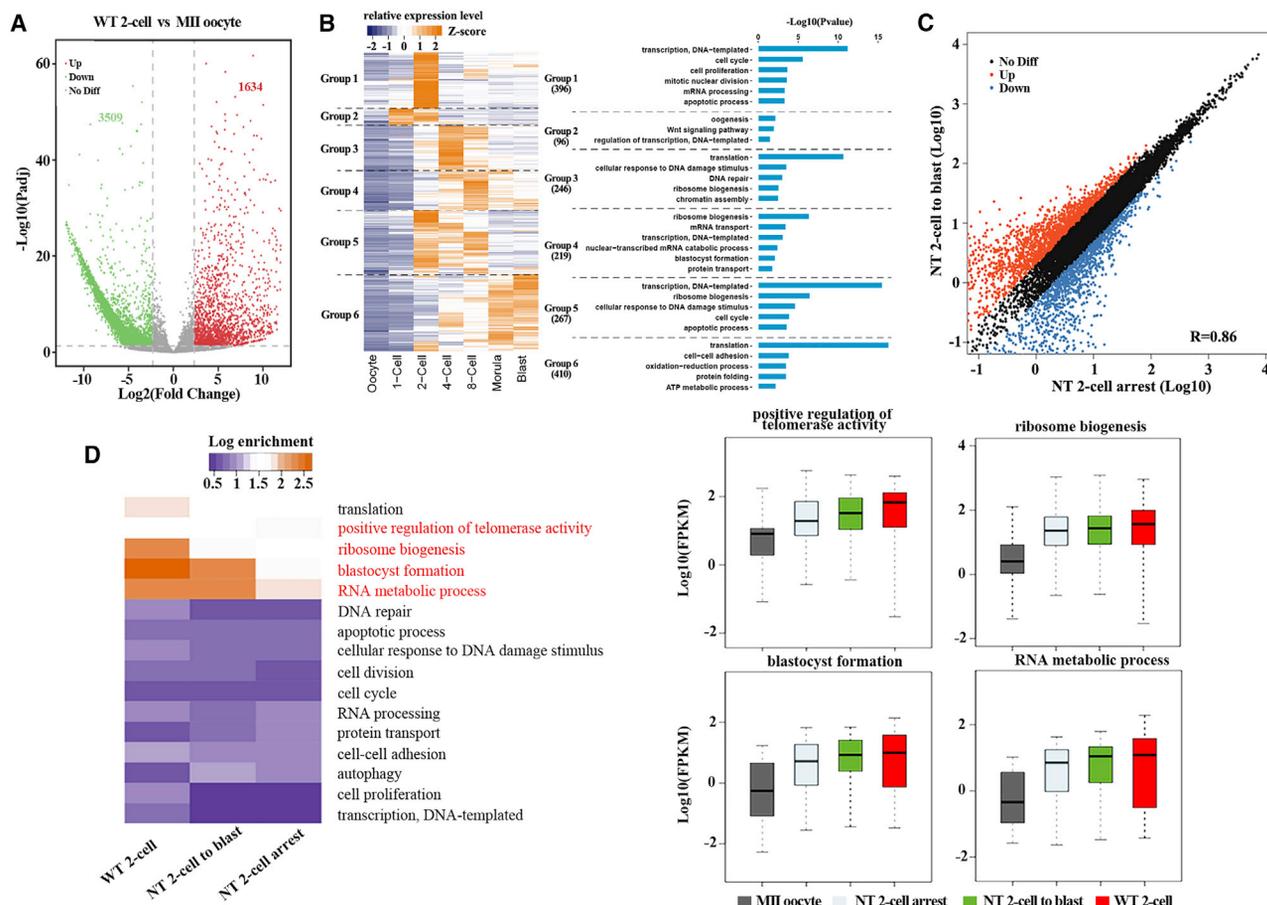


Figure 2. Enrichment Analysis of Key Functional Pathways in NT and WT Two-Cell Embryos

(A) Volcano plot comparing the gene expression profiles of WT 2-cell embryo and MII Oocyte. Upregulated and downregulated genes are colored red and green, respectively. (B) Analysis of pre-implantation embryos transcript profiles and pathway enrichment. (C) Analysis of expression profile correlation between NT 2-cell arrest and normal embryos. (D) GO enrichment analysis of ZGA genes in 2-cell embryo with different developmental fates. GO pathway enrichment degree using rich factor to measure: the higher the rich factor value, the greater the enrichment degree of this pathway. The boxplot shows the expression level of the enrichment genes of four representative pathways.

change) [\log_2FC] > 1; Figure S1A). Compared with MII Oocyte, the number of DEGs in 2-cell and 4-cell had an increased wave, which may be related to maternal-zygotic transition. There were fewer DEGs (290) between NT *Kdm4b* 2-cell and NT *Kdm4d* 2-cell, and the NT *Kdm4d* 2-cell was closer to WT 2-cell, which exhibited that both *Kdm4b* and *Kdm4d* functions have crucial roles to overcome the H3K9me3 barrier during SCNT embryo development.

ZGA Genes Identification

ZGAs in 2-cell embryos are believed to be important for embryonic development;¹⁰ we identified 1,634 upregulated genes ($p < 0.05$, FPKM [fragments per kilobase of exon model per million mapped reads] ≥ 5 ; Figure 2A) in WT 2-cell compared with the MII Oocyte.²⁵ These genes showed a stage-specific expression pattern and could be classified into six different groups (Figure 2B). There was a total of 492 genes in groups 1 and 2, and these genes were highly expressed in 2-cell samples. We found these genes significantly enriched in

transcription, cell cycle, and cell-division-related biological processes, implying that transcriptional mechanisms are gradually established with the ZGA.

We further explored whether ZGA occurred correctly in mouse SCNT embryos; the result showed that the 2-cell embryos and MII Oocyte were obviously distinguished, as expected. Surprisingly, the samples of NT 2-cell arrest and NT 2-cell to blast had similar expression patterns (Figure 2C). Therefore, we identified upregulated genes in NT 2-cell arrest and NT 2-cell to blast embryos to dissect the potential molecular differences among different embryos. These genes activated in WT or NT 2-cell embryos were evidently enriched in the same five pathways (Figure 2D), that is, translation, positive regulation of telomerase activity, ribosome biogenesis, blastocyst formation, and RNA metabolic process, which was consistent in other studies.^{26–29} The expression levels of genes related to four of these pathways (except translation) (Figure 2D) in WT or NT 2-cell

embryos were significantly higher than MII Oocytes. Among which, the gene expression level in WT 2-cell embryos was the highest and in NT 2-cell arrested embryos was relatively low, suggesting that the incomplete activation of these genes may lead to the abnormal development of NT 2-cell embryos.

Molecular Markers Screening in 2- and 4-Cell Arrest Embryos

To identify key molecular markers of SCNT embryos with distinct developmental potentials, we screened a total of 3,708 significantly upregulated genes (fold change [FC] > 5, $q < 0.05$) in 2-cell samples (WT 2-cell, NT 2-cell arrest, and NT 2-cell to blast). These genes showed stage-specific expression patterns and could be clustered into seven different groups (Figure 3A). We identified 595 (group 1) and 421 genes (group 3) that were activated in only NT 2-cell to blast and WT 2-cell embryos, respectively. Gene Ontology (GO) analysis showed that these genes were related in apoptotic signaling pathways, positive regulation of transcription metabolic processes, positive regulation of telomerase activity, etc., which indicated that the cells have undergone a great transition, in which a large number of maternally stored RNAs are degraded rapidly and replaced by newly synthesized ZGA RNAs.⁵

Based on the analysis above, we identified four key pathways of apoptosis, autophagy, endocytosis, and DNA repair that were significantly activated in NT 2-cell to blast compared with NT 2-cell arrest embryos (Figure 3A). The overall expression levels of these pathway-related genes in NT normal embryos were significantly higher than that in NT arrest embryos. Notably, the genes enriched in autophagy,³⁰ endocytosis, and apoptosis pathways were mainly activated in NT 2-cell embryos, but the genes involved in DNA repair were primarily activated in NT 4-cell embryos, indicating that the temporal activation of these pathways is important in adapting the transition of cell identity during reprogramming.

In addition to 2-cell arrest, 4-cell arrest is also observed in SCNT embryos. We analyzed DEGs in 4-cell embryos with different developmental fate (WT 4-cell, NT 4-cell arrest, and NT 4-cell to blast). Moreover, 246 upregulated genes, 7 KEGG pathways, and 34 GO biological processes were shared in WT or NT 4-cell to blast (Figure 3C), in which some significant GO terms and representative genes for these pathways were shown in Figures 3D and 3E. We further observed the overall expression levels of genes enriched in the four key pathways, including stem cell pluripotency maintenance, DNA repair, cell cycle, and autophagy (Figure 3F). The expression levels of NT 4-cell to blast and WT 4-cell embryos were significantly higher ($p < 0.001$) than those of NT 4-cell arrest embryos, which showed that the DEGs in these key pathways in NT embryos might be vital for the normal development of NT 4-cell embryos. Besides, the seven most prominent pathways contain ubiquitin-mediated proteolysis and nucleotide excision repair and are more prominent in NT 4-cell to blast than in WT 4-cell embryos (Figure 3B), which may mean that NT embryo reprogramming requires excessive activation of genes involved in protein degradation and nucleotide repair.

Identification and Functional Annotation of TFs and Pluripotent Genes in NT Embryos

In the process of SCNT, embryo-specific TFs play important roles in activating embryo-specific expressed genes and maintaining the expression of genes that normally develop embryos.³¹ Therefore, we screened 97 and 34 TFs with differential expression in NT 2-cell and 4-cell normal embryos, respectively, and found these TFs to be involved in four pathways both in 2-cell and 4-cell (Figure 4A). Next, we observed the specific expression levels of these TFs and found that *Id1*, *Pou6f1*, *Cited1*, and *Zscan4c* as the representative TFs were highly expressed in NT normal embryos compared with the NT arrest embryos (Figure 4A), which was compatible with previous studies in which these TFs performed an essential function in cell growth, differentiation, and embryonic development.^{32–36} Moreover, the peaks of these three key TFs (except *Cited1*) had significant differences between NT arrest and normal embryos (Figure 4E), indicating that they may be partially responsible for the developmental arrest of NT embryos.

We next screened 17 pluripotency maintenance genes highly expressed in NT 2-cell or 4-cell embryos (Figure 4B); the expression level of these pluripotent genes in NT normal embryos was higher than arrest embryos, indicating that pluripotency maintenance genes may function as positive factors in facilitating the developmental progress of NT embryos. Furthermore, we analyzed the relative expression levels of these genes in NT embryos and found that the key pluripotent gene *Ascl2* was activated only in NT 2-cell, but some key genes, including *Nanog*, *Dppa2*, and *Sall4*, were first activated in NT 2-cell and activated again in NT 4-cell, which was in agreement with the peak distribution of these genes at the transcriptional level (Figure 4C). Surprisingly, *Dppa2* as a positive regulator of transcription of ZGA genes has been proved.³⁷ These results suggest that abnormal expression of pluripotency maintenance genes in NT embryos is also likely to be one of the important reasons for their developmental arrest. The co-expression network analysis of these genes (Figure 4D) showed that the different cooperative regulatory relationships of pluripotent genes existed in different embryos, in which *Ascl2*, *Bmp7*, and *Ctnnb1* had high expression levels but weak correlation in NT 2-cell to blast; but in NT 4-cell to blast, both the expression levels and correlation of *Nanog* and *Sall4* were high, and *Bmp7* and *Ctr9* had high expression but weak correlation.

Comprehensive Analysis of the Key Epigenetic Modification Factors

To evaluate the effects of epigenetic modification factors on NT embryo development, including DNA/RNA demethylase, histone demethylase, and histone acetylase, we found that *Tet1* and *Tet2* were significantly activated in NT 4-cell to blast and NT 2-cell to blast embryos, respectively (Figure 5A), suggesting that the activation of both enzymes is a positive factor for the development of NT embryos.^{38–40} Intriguingly, the differential expression of these genes was closely related to the silence of individual exons at the transcription (Figure 5A), in which the peaks on the last exon differed greatly between NT 2- and 4-cell arrested and normal embryos. Remarkably, *Gadd45a* as a key player

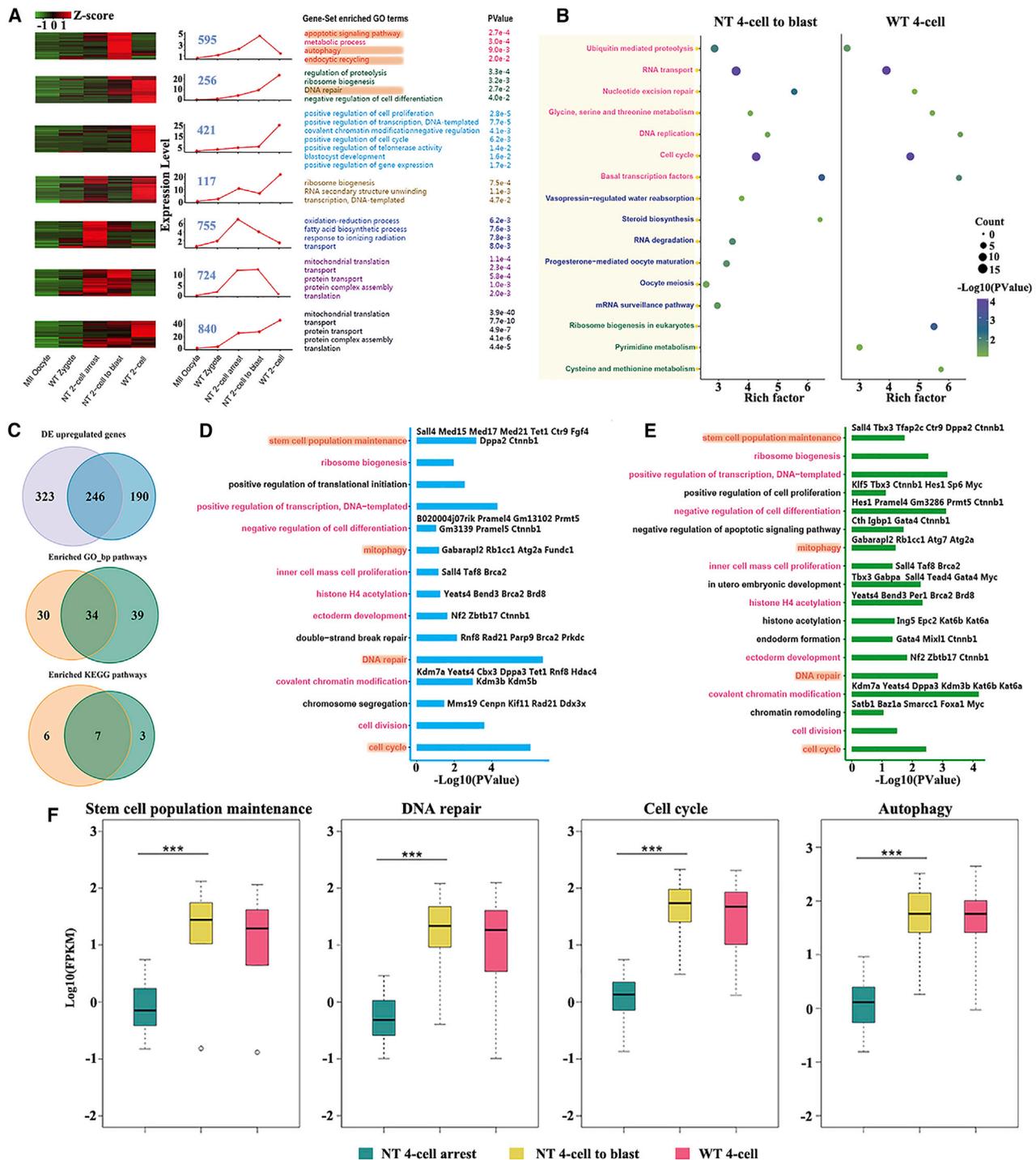


Figure 3. Cluster Analysis and Molecular Markers Screening

(A) Cluster and GO annotation analysis of ZGA genes. (B) KEGG analysis of NT 4-cell normal and WT 4-cell embryo upregulated genes; red markers for both shared functional pathways. (C) NT 4-cell to blast embryos and WT 4-cell embryos compared with NT 4-cell arrest embryo differentially upregulated genes and the pathways were shown in the Venn diagram. (D) GO analysis of upregulated genes in NT 4-cell to blast embryos; both shared pathways in NT 4-cell to blast and WT 4-cell embryos were marked in red. (E) GO analysis of upregulated genes in WT 4-cell embryos; both shared pathways in NT 4-cell to blast and WT 4-cell embryos were marked in red. (F) Expression level of genes enriched in NT 4-cell embryo representative signaling pathway.

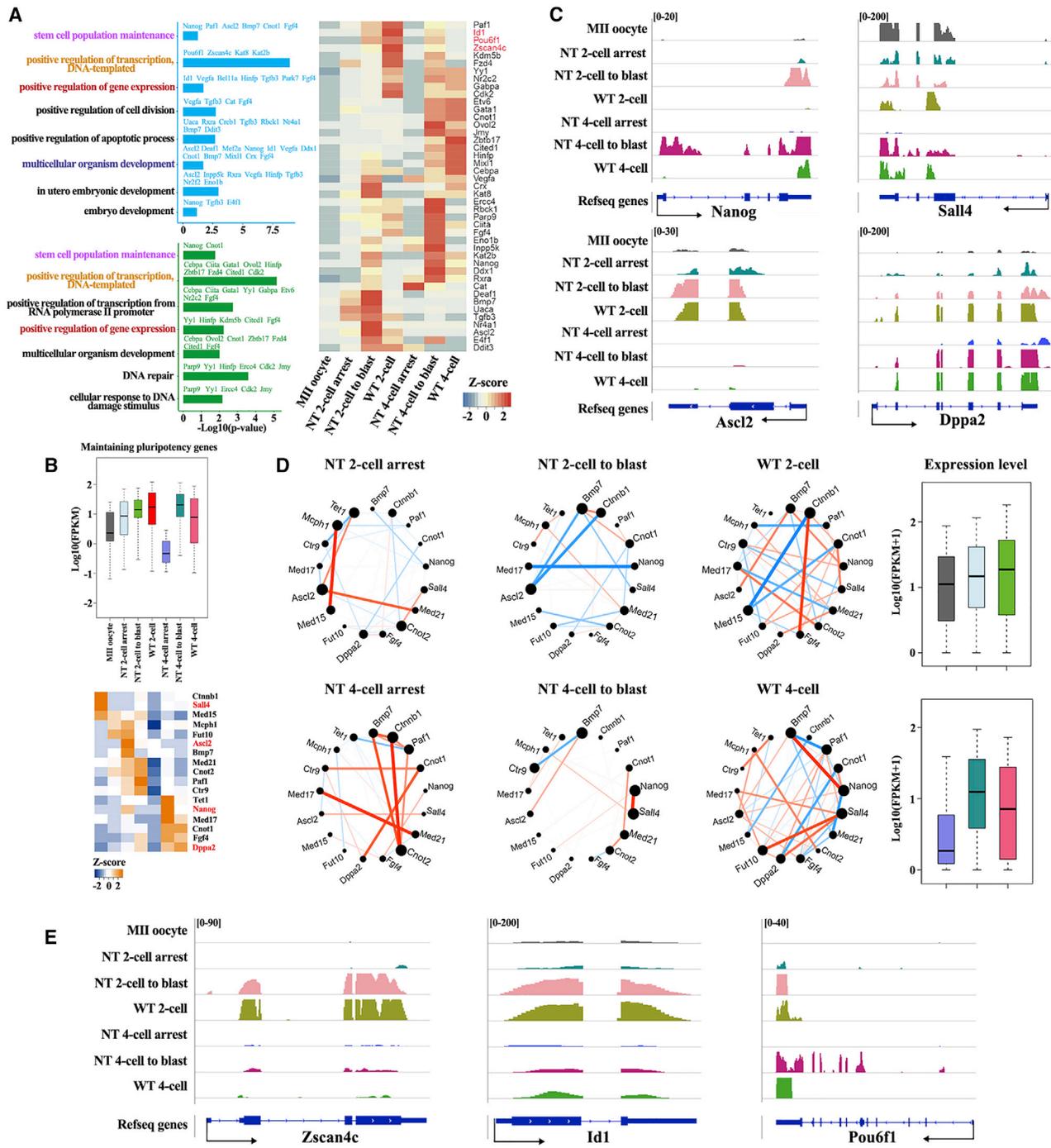


Figure 4. The Dynamic Changes of Representative TFs

(A) The bar graph is the GO annotation pathway for upregulated TFs of NT 2-cell normal embryos (top) and NT 4-cell normal embryos, respectively (bottom). The same GO pathways of both were marked with the same color. (B) The expression levels of pluripotency maintenance genes in different fate embryos. (C) Genomic views of the transcript level of representative pluripotency maintenance genes. (D) Co-expression network analysis of pluripotent genes. The red line indicates a strong correlation of gene expression levels, the blue line indicates a weak correlation, and the size of the point represents the normalized expression levels of genes. The boxplot on the right shows the overall expression level of each group of genes. (E) Genome browser view of three representative TFs, *Zscan4c*, *Id1*, and *Pou6f1*.

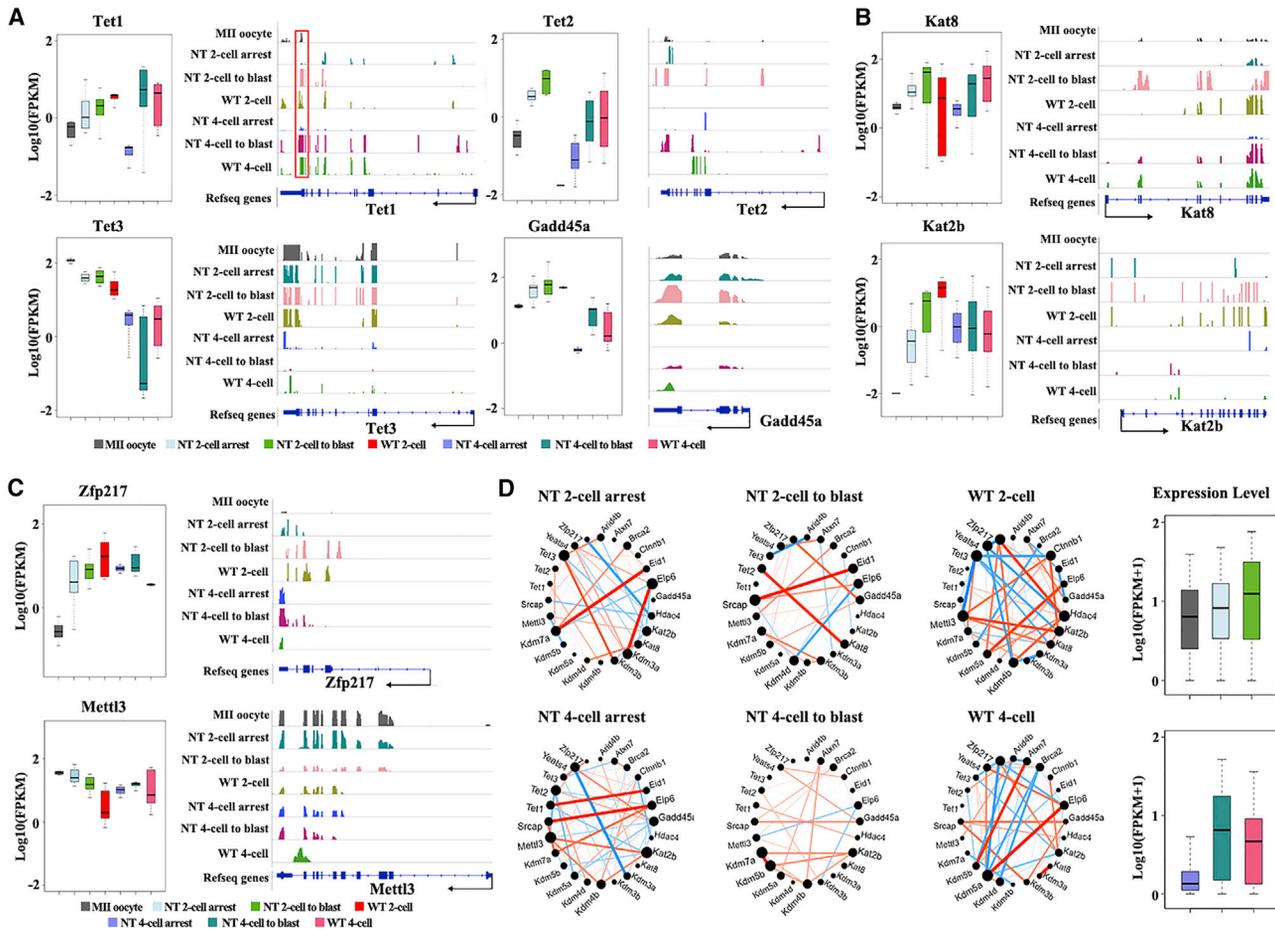


Figure 5. Identification and Excavation of Epigenetic Markers of Different Types of Embryos

(A) The expression levels and genome browser view of genes related to DNA demethylation. (B) The expression levels and genome browser view of histone acetylases *Kat8* and *Kat2b*. (C) The expression levels and genome browser view of genes related to RNA methylation. (D) Co-expression network analysis of representative epigenetic markers.

in DNA demethylase activation was found to have high expression in NT 2-cell and 4-cell, further suggesting that it may function with *Tet1* and *Tet2* to promote NT embryo development.⁴¹

Histone methylation is a very important epigenetic modification in the development of SCNT embryos.⁴² We screened a total of eight histone demethylases, in which *Kdm3a*, *Kdm7a*, *Kdm4b*, and *Kdm4d* had prominent high expression levels in normal NT 2-cell embryos. Previous experiments had confirmed that SCNT embryos injected with *Kdm4b/4d* mRNA developed to the blastocyst stage with high efficiency,²⁵ which was consistent with our results that SCNT 2-cell embryos injected with *Kdm4b/4d* are closer to WT 2-cell embryos (Figure 6A). Interestingly, *Kdm4d* mRNA injection can greatly improve the blastocyst rate of SCNT embryos to above 80% when cumulus cells, Sertoli cells, or MEF cells were used as donor cells¹⁰ (Figure 6F), so we speculate that *Kdm3a* and *Kdm7a* may be candidate genes of NT 2-cell development embryos. On the contrary, *Kdm3b*, *Kdm5a*, *Kdm5b*, and *Ctnnb1* were signifi-

cantly higher in NT 4-cell to blast, in which NT *Kdm5b* 4-cell embryos were closer to WT 4-cell embryos (Figure 6A). The result indicated that *Kdm3b*, *Kdm5a*, *Kdm5b*, and *Ctnnb1* may be key factors for NT 4-cell embryonic development. Next, we classed some genes activated in NT 2-cell embryos into five groups (Figure 6B) and performed GO analysis of shared genes between different embryos (NT 2-cell and WT 2-cell, WT 4-cell, and NT *Kdm5b* 4-cell) (Figures 6C and 6E); the results further showed that *Kdm4b/4d* and *Kdm5b* as key genes can properly rescue 2-cell and 4-cell arrest, respectively.

Histone acetylation is also a crucial epigenetic modification during development, in which co-knockdown of *Hdac1* and *Hdac2* (but not individually) can result in the failure of preimplantation development.⁴³ We found 10 genes (Figures 5B and 5D) with the relevance of histone modification, in which *Srcap* and *Kat2b* specifically activated in NT 2-cell to blast; *Hdac4*, *Yeats4*, *Atxn7*, *Arid4b*, and *Brca2* had a significant wave in NT 4-cell to blast; and *Elp6*,

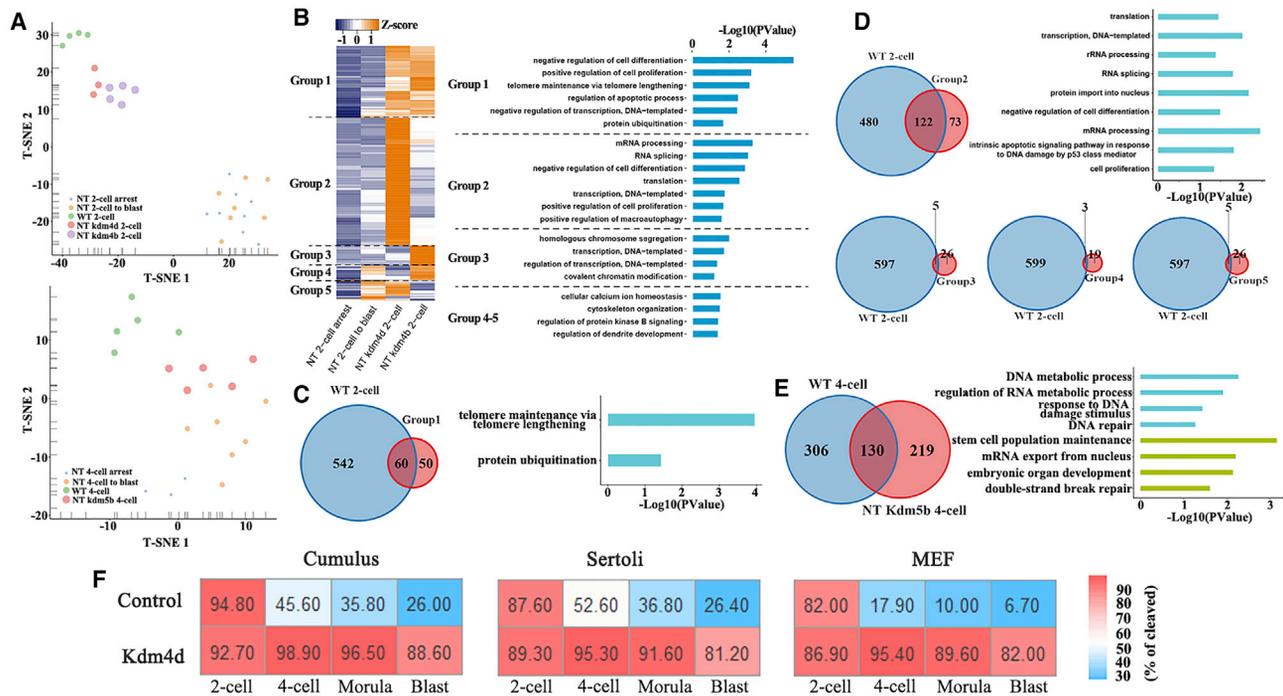


Figure 6. Histone Methylation Is Important to the Development of SCNT Embryos

(A) The T-SNE clusters of 2-cell and 4-cell embryos with different developmental potentials. Data are from Liu et al.²⁵ (B) Clustering and GO annotation analysis of NT *Kdm4b/4d* 2-cell embryos upregulated genes. (C) Shared genes of group 1 in (B) and WT 2-cell and GO analysis. (D) Shared gene of group 2 and groups 3–5 in (B) and WT 2-cell and GO analysis. (E) NT *Kdm5b* 4-cell and WT 4-cell shared genes and GO analysis. (F) *Kdm4d* mRNA injection improves the development of SCNT embryos when cumulus cells, Sertoli cells, and MEF are donor cells. The percentage of embryos that reaches the indicated stages was shown in figures (data are reanalyzed from Matoba et al.¹⁶).

Eid1, and *Kat8* in both NT 2-cell and 4-cell normal embryos had significantly high expression. The result indicated that the abnormal expression of genes related to histone modification is one of the important factors for the developmental failure of NT embryos.

Chemical modification of RNA is also an important epigenetic modification that directly and rapidly manipulates the transcriptome to regulate gene expression.⁴⁴ In Figure 5C, *Zfp217* was highly expressed in NT 2-cell normal embryos, but *Mettl3* showed significantly low expression, which was consistent with a previous study in which *Zfp217* can inactivate the N6-methyladenosine [m(6)A] methyltransferase *Mettl3* by arresting the methylation of these mRNAs to increase the efficiency of reprogramming.⁴⁵ The result showed that the low expression of *Zfp217* in NT 2-cell arrested embryos may be one of the reasons for developmental arrest.

We further performed co-expression networks analysis of the above identified key analytical markers (Figure 5D) and found that histone acetylation genes *Elp6* and histone demethylation enzymes *Kdm5a* both have a high expression level and strong correlation in WT embryos, but in NT 2-cell to blast, *Srcap* and *Eid1* related to histone acetylation play the main roles in the normal development of NT 2-cell. Moreover, the expression levels of *Kdm5b* and *Kdm7a* in NT 4-cell to blast were significantly higher than in others, showing that *Kdm7a* may function with *Kdm5b* to promote NT 4-cell embryo

development.²⁵ In short, the results revealed that coordinate regulation of different epigenetic modification factors is necessary for the normal development of embryos.

DISCUSSION

To elucidate the potential molecular mechanism of embryo development, we systematically characterized the global transcriptomes profiles of NT embryos and WT embryos with different developmental fates based on the SCNT embryo biopsy system of a previous study,²⁵ which collectively suggests that the developmental fate of the embryos can be directly reflected in the differences of transcriptome level. Furthermore, compared with NT arrest embryos, some key functional pathways activated in NT 2- and 4-cell normal embryos were comprehensively investigated. As shown in Figure 7, the genes enriched in autophagy, endocytosis, and apoptosis pathways were dramatically activated in NT 2-cell, but stem cell pluripotency maintenance, DNA repair, cell cycle, and autophagy key pathways may play important roles in the normal development of NT 4-cell embryos.

In previous studies, *Kdm4b/4d* and *Kdm5b* have been confirmed to be key genes in NT 2-cell and 4-cell embryos, respectively.^{10,25} Strikingly, *Kdm4d* mRNA injection can significantly improve the developmental efficiency of SCNT embryos, regardless of the donor cell types.¹⁰ Numerous previous studies also have shown that the SCNT reprogramming process requires coordinated regulation among

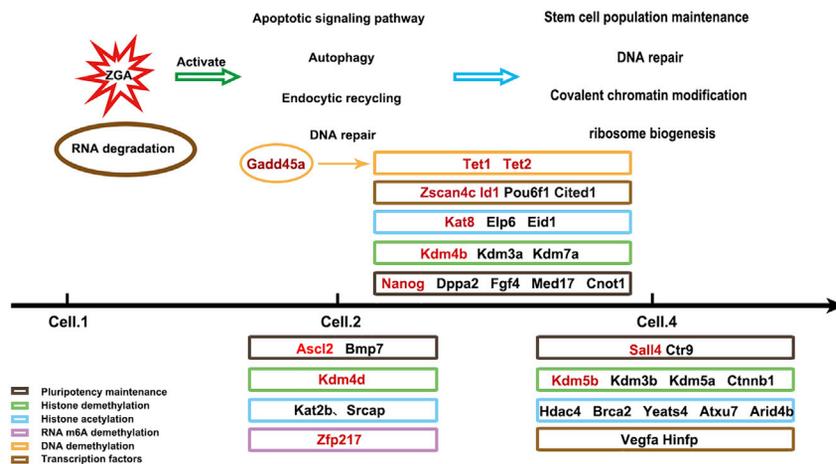


Figure 7. The Probable Timing Activation Pathways and Reprogramming Signatures in SCNT Embryos Identified by Comparative Single-Cell Transcriptomes

Genes marked red have been proved experimentally in previous studies.

multiple factors.^{5,46} Consistently, we identified some potential molecular markers in SCNT embryos, in which *Id1*, *Pou6f1*, *Cited1*, and *Zscan4c* as the representative TFs were highly expressed in NT normal embryos compared with the NT arrested embryos, showing that these factors act as important drivers in NT normal embryos. Histone acetylases *Kat8*, *Elp6*, and *Eid1* were co-activated in both NT 2-cell and NT 4-cell to facilitate embryo normal development. As in the previous study, *Zfp217* can inactivate the m(6)A methyltransferase *Mettl3* to increase the efficiency of reprogramming.⁴⁵ Moreover, *Gadd45a* as a key player functions in DNA demethylase activation and coordinately regulates with *Tet1* and *Tet2* to improve the efficiency of NT reprogramming.

Taken together, our studies identified some reprogramming driver factors by comparative analysis of single-cell transcriptome, hoping to provide a new insight into deciphering the potential molecular mechanisms and improving the efficiency of SCNT reprogramming.⁴⁷ In future work, we will further analyze the effects of epigenetic disorders on reprogramming by integrating other multi-omics data, such as DNA methylation and histone modification data.^{36,48–50}

MATERIALS AND METHODS

Dataset Collection

The single-cell RNA sequencing (RNA-seq) data of this study were downloaded from Gene Expression Omnibus (GEO) database under accession number GEO: GSE70605.²⁵ There are two embryonic types, including SCNT embryos and *in vitro* fertilization (WT) embryos as a control. WT samples include cumulus cell, MII Oocyte, Zygote, 2-cell, 4-cell, 8-cell, Morula, Blastocyst, Inner cell mass (ICM), and Trophectoderm (TE), and each stage has three to six replicates. For SCNT or mRNA injection SCNT samples, three to eight replicates were performed for each stage, these stages can be divided into NT arrest and NT develop to blastocyst (NT to blast).

Data Processing

All RNA-seq data used FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and FASTX-Toolkit (version

0.0.13) software evaluation to remove low-quality samples, and the remaining samples in each class were statistically enough. Filtered clean reads were mapped to the mouse genome (mm9)¹⁰ with HISAT2⁵¹ (version 2.1.0). Read counts of gene expression were performed using HTseq-count (Python package). Transcriptome assembly was performed using Stringtie (version 1.3.3)^{51,52} and Ballgown⁵¹ (R package), and expression levels of each gene were quantified with FPKM¹⁰ to eliminate the effects of sequencing depth and transcript length.

Analysis of DEGs

The genes with FPKM ≥ 0.1 in each embryo sample were regarded as expressed genes.⁵³ R package DESeq2⁵⁴ was used to identify DEGs. The p value adjusted for multiple tests and the absolute value of log₂FC were obtained to identify DEGs, in which the genes with log₂FC > 1 and adjusted p < 0.05 were considered upregulated genes. Otherwise, they were (log₂FC < -1, adjusted p < 0.05) regarded as downregulated genes.⁵⁵

Gene Clustering and Gene Set Enrichment Analysis

After normalization of DEGs, we used the hierarchical clustering algorithm^{23,56–58} hclust(d, method = “complete,” members = NULL) to identify cluster-specific genes based on their expression profile similarity. GO and Kyoto Encyclopedia of Genes and Genomes (KEGG)⁵⁹ pathway enrichment analysis of DEGs for each group were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resource (<https://david.ncifcrf.gov/home.jsp>).⁶⁰ GO terms for each functional cluster were a summary of representative terms, and p values (p < 0.05) were plotted to show the significance. The KEGG systematically analyzed the gene function and genomic information database, and studied the gene and expression information as a whole network. In our studies, GO and KEGG analysis were visualized by bar plot and bubble graph, respectively.

Statistical Analysis

In this study, the normalization of expression values was conducted by Z score standardization. The formula is as follows:

$$Z = (x - \mu) / \sigma.$$

The Pearson's correlation coefficients^{61,62} (denoted as ρ) were applied in statistics to estimate the relationship between two samples. The formula is as follows:

$$\rho = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

where X and \bar{X} indicate the true expression value and mean in group 1, respectively, and expression value Y and mean \bar{Y} are in group 2.

The t test^{61–63} was used as a common statistical test to compare the difference between the two means relative to the variation.

Data Visualization

In this study, R/Bioconductor (<https://www.bioconductor.org/>) software packages were used for data visualization. For example, the t-Distributed Stochastic Neighbor Embedding (t-SNE) plot and Venn plot were produced using R packet Scater and VennDiagram, respectively, and the density graph, (Principal Component Analysis) PCA, bubble graph, and so on were generated with the R packet ggplot2 <http://ggplot2.org/>. The genome browser view was obtained by using the Integrative Genomics Viewer (IGV).⁶⁴ The R package for weighted gene co-expression network analysis (WGCNA)⁶⁵ was used to construct weighted gene co-expression networks. The flow-chart of this study is shown in Figure 1A.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omtn.2019.12.035>.

AUTHOR CONTRIBUTIONS

Y.Z. conceived and designed the study. M.S. did most of the bioinformatics analysis. P.C. and H.L. were helpful for materials collection. L.Z. participated in some statistical analyses. Y.Z. and H.L. wrote the paper. All authors read and approved the manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Shaorong Gao (Tongji University) and Prof. Yi Zhang (Harvard Medical School) for sharing their single-cell RNA-seq data of SCNT embryos in Gene Expression Omnibus (GEO) database. This work was supported by the National Nature Scientific Foundation of China (grants 61561036, 61702290, and 61861036); Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region (grant NJYT-18-B01); and the Fund for Excellent Young Scholars of Inner Mongolia (grant 2017JQ04). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Chung, Y.G., Matoba, S., Liu, Y., Eum, J.H., Lu, F., Jiang, W., Lee, J.E., Sepilian, V., Cha, K.Y., Lee, D.R., and Zhang, Y. (2015). Histone Demethylase Expression Enhances Human Somatic Cell Nuclear Transfer Efficiency and Promotes Derivation of Pluripotent Stem Cells. *Cell Stem Cell* 17, 758–766.
- Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* 10, 622–640.
- Wilmot, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* 385, 810–813.
- Rodríguez-Osorio, N., Urrego, R., Cibelli, J.B., Eilertsen, K., and Memili, E. (2012). Reprogramming mammalian somatic cells. *Theriogenology* 78, 1869–1886.
- Matoba, S., and Zhang, Y. (2018). Somatic Cell Nuclear Transfer Reprogramming: Mechanisms and Applications. *Cell Stem Cell* 23, 471–485.
- Ogura, A., Inoue, K., and Wakayama, T. (2013). Recent advancements in cloning by somatic cell nuclear transfer. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20110329.
- Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.S., Sritanaudomchai, H., et al. (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* 153, 1228–1238.
- Byrne, J.A., Pedersen, D.A., Clepper, L.L., Nelson, M., Sanger, W.G., Gokhale, S., Wolf, D.P., and Mitalipov, S.M. (2007). Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* 450, 497–502.
- Zuo, Y., Gao, Y., Su, G., Bai, C., Wei, Z., Liu, K., Li, Q., Bou, S., and Li, G. (2014). Irregular transcriptome reprogramming probably causes the developmental failure of embryos produced by interspecies somatic cell nuclear transfer between the Przewalski's gazelle and the bovine. *BMC Genomics* 15, 1113.
- Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell* 159, 884–895.
- Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., et al. (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Mol. Cell* 63, 1066–1079.
- Keniry, A., Gearing, L.J., Jansz, N., Liu, J., Holik, A.Z., Hickey, P.F., Kinkel, S.A., Moore, D.L., Breslin, K., Chen, K., et al. (2016). Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. *Epigenetics Chromatin* 9, 16.
- Liu, Z., Cai, Y., Wang, Y., Nie, Y., Zhang, C., Xu, Y., Zhang, X., Lu, Y., Wang, Z., Poo, M., and Sun, Q. (2018). Cloning of Macaque Monkeys by Somatic Cell Nuclear Transfer. *Cell* 174, 245.
- Inoue, K., Kohda, T., Sugimoto, M., Sado, T., Ogonuki, N., Matoba, S., Shiura, H., Ikeda, R., Mochida, K., Fujii, T., et al. (2010). Impeding Xist expression from the active X chromosome improves mouse somatic cell nuclear transfer. *Science* 330, 496–499.
- Matoba, S., Inoue, K., Kohda, T., Sugimoto, M., Mizutani, E., Ogonuki, N., Nakamura, T., Abe, K., Nakano, T., Ishino, F., and Ogura, A. (2011). RNAi-mediated knockdown of Xist can rescue the impaired postimplantation development of cloned mouse embryos. *Proc. Natl. Acad. Sci. USA* 108, 20621–20626.
- Vassena, R., Han, Z., Gao, S., Baldwin, D.A., Schultz, R.M., and Latham, K.E. (2007). Tough beginnings: alterations in the transcriptome of cloned embryos during the first two cell cycles. *Dev. Biol.* 304, 75–89.
- Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15, e8746.
- Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 9, 75.
- Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2019). Systematic comparative analysis of single cell RNA-sequencing methods. *bioRxiv*. <https://doi.org/10.1101/632216>.
- Mallick, S., and Zhao, Z. (2018). Identification of gene signatures from RNA-seq data using Pareto-optimal cluster algorithm. *BMC Syst. Biol.* 12 (Suppl 8), 126.

21. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* 8, e43803.
22. Olsen, T.K., and Baryawno, N. (2018). Introduction to Single-Cell RNA Sequencing. *Curr. Protoc. Mol. Biol.* 122, e57.
23. Mallik, S., and Zhao, Z. (2019). Multi-Objective Optimized Fuzzy Clustering for Detecting Cell Clusters from Single-Cell Expression Profiles. *Genes (Basel)* 10, 611.
24. Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., Yong, J., Sun, L., Wang, X., Zhao, Y., et al. (2018). Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* 28, 730–745.
25. Liu, W., Liu, X., Wang, C., Gao, Y., Gao, R., Kou, X., Zhao, Y., Li, J., Wu, Y., Xiu, W., et al. (2016). Identification of key factors conquering developmental arrest of somatic cell cloned embryos by combining embryo biopsy and single-cell sequencing. *Cell Discov.* 2, 16010.
26. Long, C., Li, W., Liang, P., Liu, S., and Zuo, Y. (2019). Transcriptome Comparisons of Multi-Species Identify Differential Genome Activation of Mammals Embryogenesis. *IEEE Access* 7, 7794–7802.
27. Hu, B., Zheng, L., Long, C., Song, M., Li, T., Yang, L., and Zuo, Y. (2019). EmExplorer: a database for exploring time activation of gene expression in mammalian embryos. *Open Biol.* 9, 190054.
28. Zuo, Y., Su, G., Wang, S., Yang, L., Liao, M., Wei, Z., Bai, C., and Li, G. (2016). Exploring timing activation of functional pathway based on differential co-expression analysis in preimplantation embryogenesis. *Oncotarget* 7, 74120–74131.
29. Chen, W., Feng, P., Liu, T., and Jin, D. (2019). Recent Advances in Machine Learning Methods for Predicting Heat Shock Proteins. *Curr. Drug Metab.* 20, 224–228.
30. Nakashima, A., Aoki, A., Kusabiraki, T., Shima, T., Yoshino, O., Cheng, S.-B., Sharma, S., and Saito, S. (2017). Role of autophagy in oocytogenesis, embryogenesis, implantation, and pathophysiology of pre-eclampsia. *J. Obstet. Gynaecol. Res.* 43, 633–643.
31. Godini, R., and Fallahi, H. (2019). Dynamics changes in the transcription factors during early human embryonic development. *J. Cell. Physiol.* 234, 6489–6502.
32. Ruzinova, M.B., and Benezra, R. (2003). Id proteins in development, cell cycle and cancer. *Trends Cell Biol.* 13, 410–418.
33. Falco, G., Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T., and Ko, M.S. (2007). Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.* 307, 539–550.
34. Suzuki, N., Yoshioka, N., Uekawa, A., Matsumura, N., Tozawa, A., Koike, J., Konishi, I., Kiguchi, K., and Ishizuka, B. (2010). Transcription factor POU6F1 is important for proliferation of clear cell adenocarcinoma of the ovary and is a potential new molecular target. *Int. J. Gynecol. Cancer* 20, 212–219.
35. Rodriguez, T.A., Sparrow, D.B., Scott, A.N., Withington, S.L., Preis, J.I., Michalick, J., Clements, M., Tsang, T.E., Shioda, T., Beddington, R.S., and Dunwoodie, S.L. (2004). Cited1 is required in trophoblasts for placental development and for embryo growth and survival. *Mol. Cell. Biol.* 24, 228–244.
36. Chen, W., Feng, P., Song, X., Lv, H., and Lin, H. (2019). iRNA-m7G: Identifying N⁷-methylguanosine Sites by Fusing Multiple Features. *Mol. Ther. Nucleic Acids* 18, 269–274.
37. Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., and Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* 33, 194–208.
38. Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.L., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* 24, 310–323.
39. Li, G.X., Ding, Z.Y., Wang, Y.W., Liu, T.T., Chen, W.X., Wu, J.J., Xu, W.Q., Zhu, P., and Zhang, B.X. (2019). Integrative analysis of DNA methylation and gene expression identify a six epigenetic driver signature for predicting prognosis in hepatocellular carcinoma. *J. Cell. Physiol.* 234, 11942–11950.
40. Liu, D., Li, G., and Zuo, Y. (2019). Function determinants of TET proteins: the arrangements of sequence motifs with specific codes. *Brief. Bioinform.* 20, 1826–1835.
41. Li, Z., Gu, T.P., Weber, A.R., Shen, J.Z., Li, B.Z., Xie, Z.G., Yin, R., Guo, F., Liu, X., Tang, F., et al. (2015). Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res.* 43, 3986–3997.
42. Yang, L., Song, L., Liu, X., Bai, L., and Li, G. (2018). KDM6A and KDM6B play contrasting roles in nuclear transfer embryos revealed by MERVL reporter system. *EMBO Rep.* 19, e46240.
43. Zhao, P., Wang, H., Wang, H., Dang, Y., Luo, L., Li, S., Shi, Y., Wang, L., Wang, S., Mager, J., and Zhang, K. (2019). Essential roles of HDAC1 and 2 in lineage development and genome-wide DNA methylation during mouse preimplantation development. *Epigenetics* 2019, 1–17.
44. Cao, G., Li, H.-B., Yin, Z., and Flavell, R.A. (2016). Recent advances in dynamic m6A RNA modification. *Open Biol.* 6, 160003.
45. Aguilo, F., Zhang, F., Sancho, A., Fidalgo, M., Di Cecilia, S., Vashisht, A., Lee, D.F., Chen, C.H., Rengasamy, M., Andino, B., et al. (2015). Coordination of m(6)A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming. *Cell Stem Cell* 17, 689–704.
46. Gao, R., Wang, C., Gao, Y., Xiu, W., Chen, J., Kou, X., Zhao, Y., Liao, Y., Bai, D., Qiao, Z., et al. (2018). Inhibition of Aberrant DNA Re-methylation Improves Post-implantation Development of Somatic Cell Nuclear Transfer Embryos. *Cell Stem Cell* 23, 426–435.e5.
47. Zuo, Y., Su, G., Cheng, L., Liu, K., Feng, Y., Wei, Z., Bai, C., Cao, G., and Li, G. (2017). Coexpression analysis identifies nuclear reprogramming barriers of somatic cell nuclear transfer embryos. *Oncotarget* 8, 65847–65859.
48. Li, H., Ta, N., Long, C., Zhang, Q., Li, S., Liu, S., Yang, L., and Zuo, Y. (2019). The spatial binding model of the pioneer factor Oct4 with its target genes during cell reprogramming. *Comput. Struct. Biotechnol. J.* 17, 1226–1233.
49. Lv, H., Zhang, Z.M., Li, S.H., Tan, J.X., Chen, W., and Lin, H. (2019). Evaluation of different computational methods on 5-methylcytosine sites identification. *Brief. Bioinform.* 2019, bbz048.
50. Xu, Z.C., Feng, P.M., Yang, H., Qiu, W.R., Chen, W., and Lin, H. (2019). iRNAD: a computational tool for identifying D modification sites in RNA sequence. *Bioinformatics* 35, 4922–4929.
51. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667.
52. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295.
53. Cao, S., Han, J., Wu, J., Li, Q., Liu, S., Zhang, W., Pei, Y., Ruan, X., Liu, Z., Wang, X., et al. (2014). Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. *BMC Genomics* 15, 4.
54. Slattery, M.L., Herrick, J.S., Mullany, L.E., Samowitz, W.S., Sevens, J.R., Sakoda, L., and Wolff, R.K. (2017). The co-regulatory networks of tumor suppressor genes, oncogenes, and miRNAs in colorectal cancer. *Genes Chromosomes Cancer* 56, 769–787.
55. Cheng, S., Pei, Y., He, L., Peng, G., Reinius, B., Tam, P.P.L., Jing, N., and Deng, Q. (2019). Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development. *Cell Rep.* 26, 2593–2607.e3.
56. Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* 3, 346–360.e4.
57. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401.
58. Hu, M.-W., Kim, D.W., Liu, S., Zack, D.J., Blackshaw, S., and Qian, J. (2019). PanoView: An iterative clustering method for single-cell RNA sequencing data. *PLoS Comput. Biol.* 15, e1007040.
59. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44 (D1), D457–D462.

60. Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* *4*, 44–57.
61. Ahlgren, P., Jarneving, B., and Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *J. Am. Soc. Inf. Sci. Technol.* *54*, 550–560.
62. Bandyopadhyay, S., Mallik, S., and Mukhopadhyay, A. (2014). A Survey and Comparative Study of Statistical Tests for Identifying Differential Expression from Microarray Data. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* *11*, 95–115.
63. Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., and Altman, D.G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* *31*, 337–350.
64. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* *14*, 178–192.
65. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* *9*, 559.