Stem Cell Reports, Volume *18*

Supplemental Information

CFIm-mediated alternative polyadenylation safeguards the development of mammalian pre-implantation embryos

Na Li, Ying Cai, Min Zou, Jian Zhou, Ling Zhang, Liquan Zhou, Wenpei Xiang, Yan Cui, and Huaibiao Li

Supplemental Figures and Legends

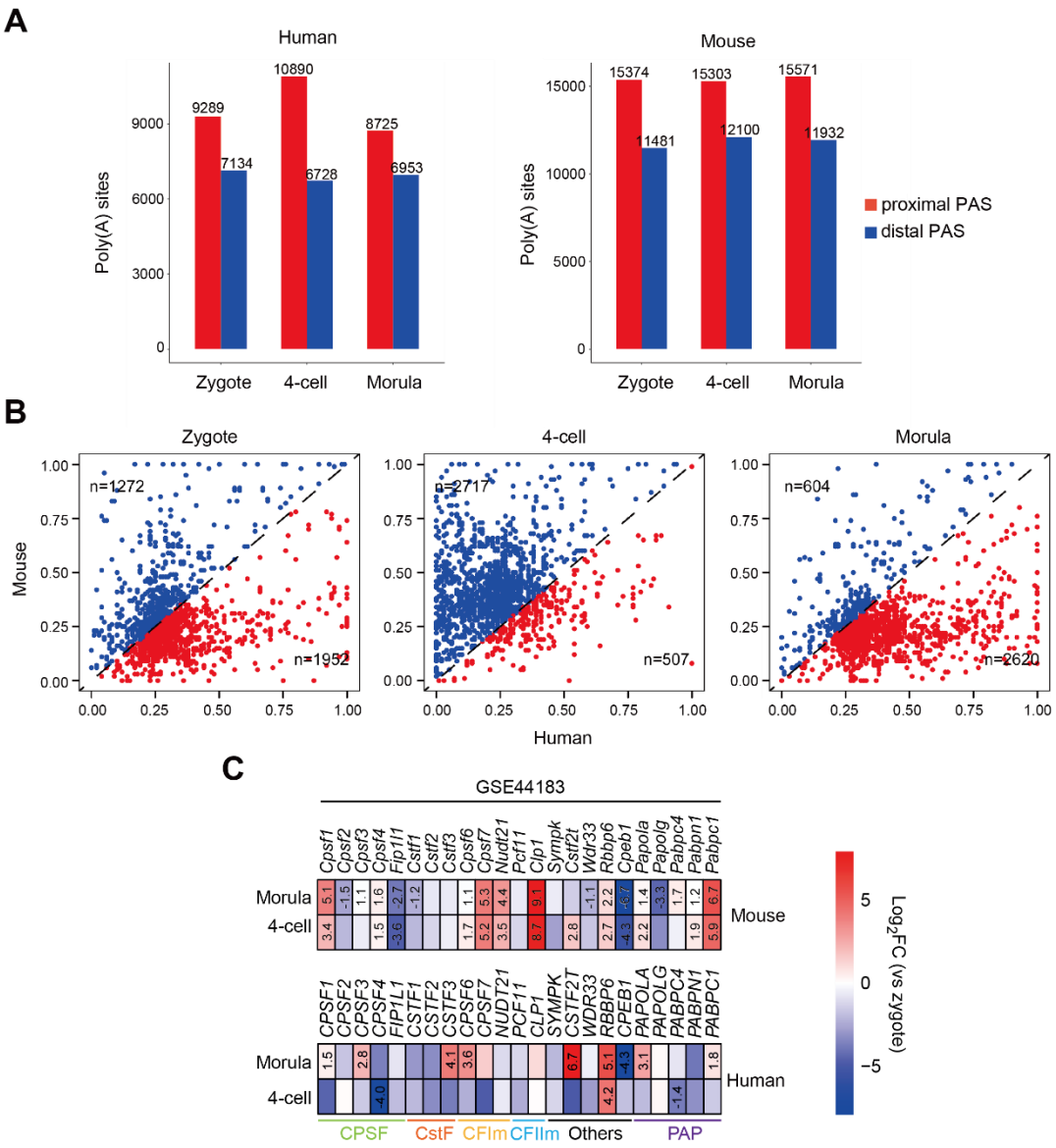


Figure S1. Analysis of PAS usage in human and mouse early embryos, related to Figure 1

(**A**) The total number of proximal and distal PASs within 3′UTRs at the three developmental stages of human and mouse early embryos. The PASs were detected by APAtrap. (**B**) Scatter plot of proportion coverage of conserved distal PASs normalized to 3′UTRs. The coverage of conserved distal PASs in mouse embryos was plotted against that of human embryos at the same developmental stage. n indicates the number of genes with a conserved distal PAS. (**C**) The expression of polyadenylation core components in human and mouse early embryos, generated by re-analysis of the scRNA-seq dataset GSE44183 using DESeq2. Fold changes of APA genes differentially expressed between 4-cell and morula stages (adjusted *p* value < 0.05) are shown in the heatmap.

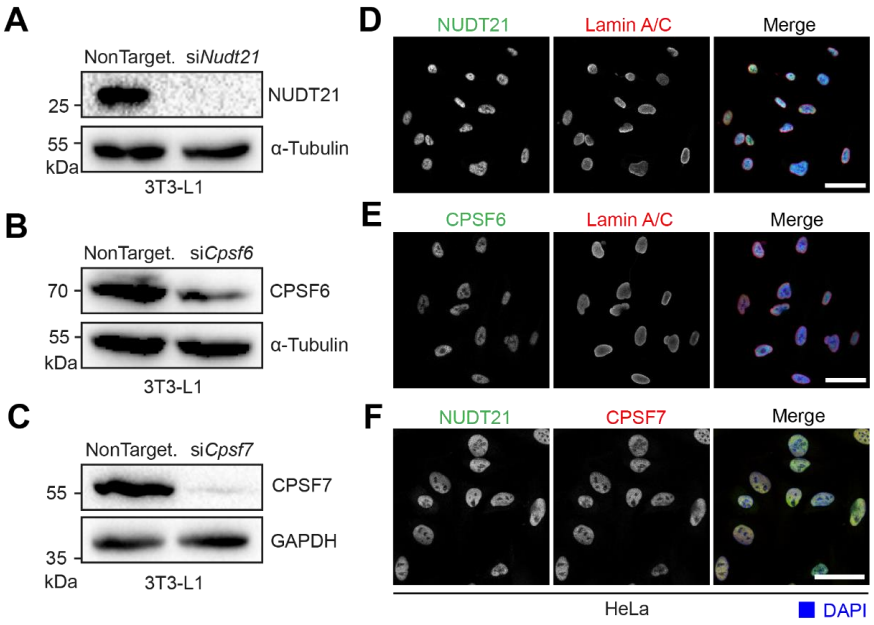


Figure S2. The specificity of antibodies against CFIm components tested by immunoblotting and immunostaining, related to Figure 4

(**A**–**C**) Immunoblotting analysis of CFIm components in 3T3-L1 cells after siRNA knockdown. α-Tubulin or GAPDH was used as a loading control. (**D**–**F**) Immunostaining of CFIm components in HeLa cells. NUDT21, CPSF6 and CPSF7 were exclusively localized in the nucleus. Scale bar, 50 μm.

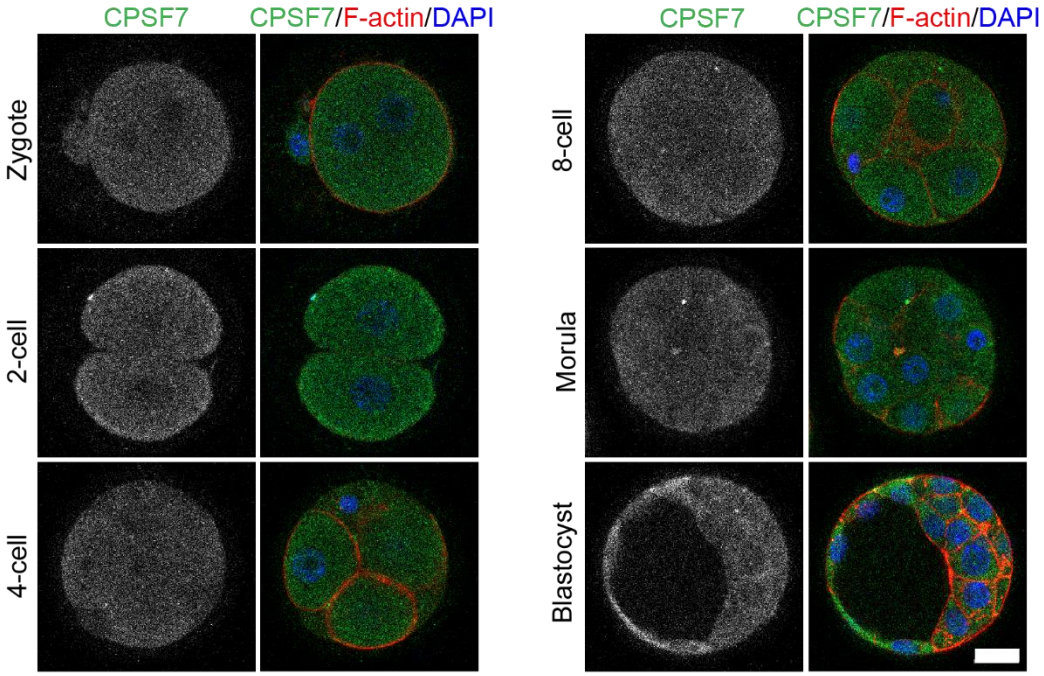


Figure S3. Immunostaining of CPSF7 in mouse early embryos, related to Figure 4

CPSF7 was mainly detected within the cytoplasm. Scale bar, 20 μm

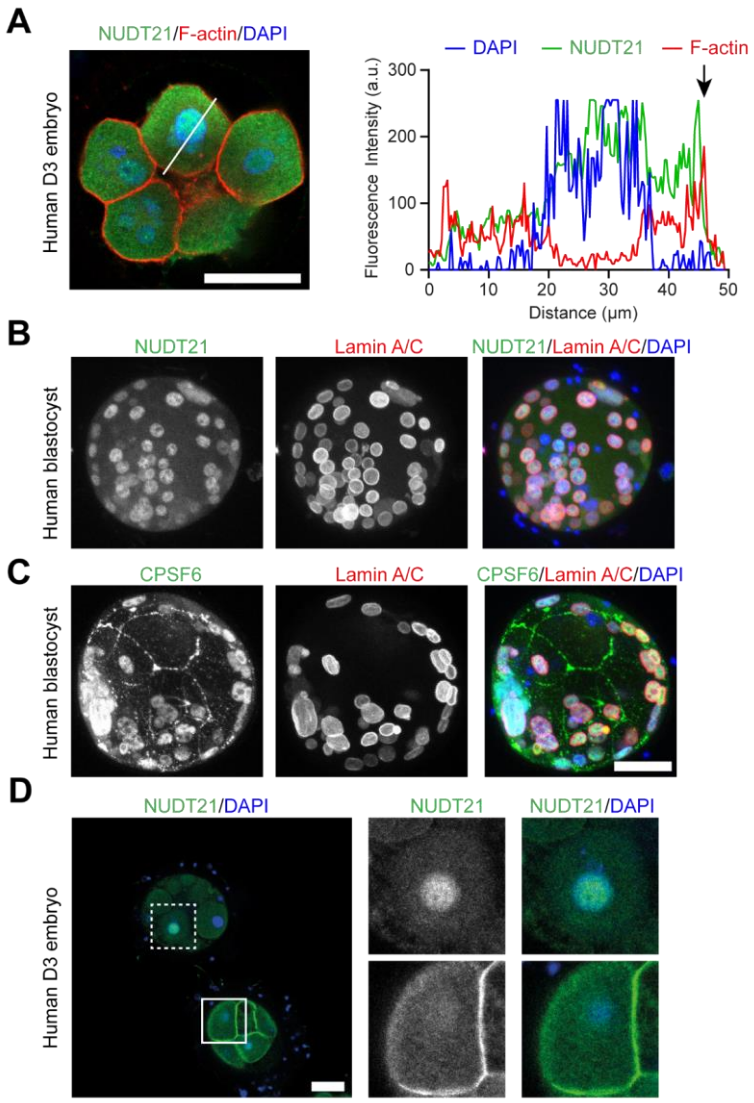
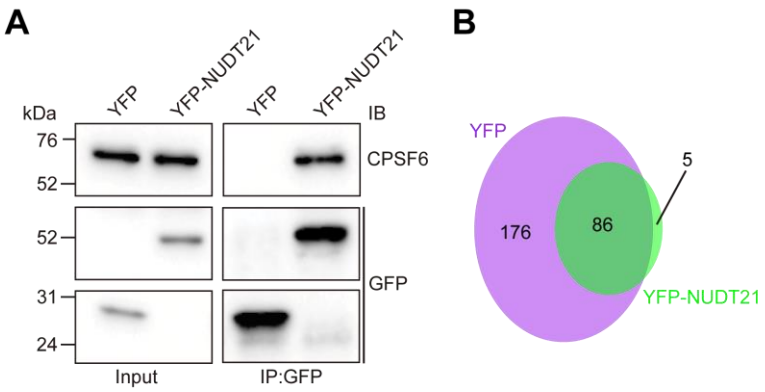


Figure S4. Immunostaining of NUDT21 or CPSF6 in human early embryos, related to Figure 5

(A) Co-staining of NUDT21 with F-actin in human D3 embryos. F-actin was stained with Phalloidin. Arrow indicates the cortex of the blastomere. (**B**, **C**) Co-staining of NUDT21 (**B**) or CPSF6 (**C**) with Lamin A/C in blastocysts showed that both NUDT21 and CPSF6 were localized in the nucleus of the cells. Maximum projected Z-stack images were shown. (**D**) Immunostaining of NUDT21 in human D3 embryos. Two embryos in the same image showed distinct localization of NUDT21. The upper one was corNUDT21−ve, while the lower one was corNUDT21+ve. Scale bar, 50 μm.



C

Protein	Uniq.Pep.(#)	Coverage(%)
NUDT21	8	39.65
hnRNPA3	4	13.76
CPSF7	3	6.79
RPS7	1	6.19
CPSF6	2	4.9

Figure S5. NUDT21-interacting proteins identified by immunoprecipitation followed by mass spectrometry, related to Figure 5

(**A**) Immunoblotting analysis of YFP immunoprecipitation. Endogenous CPSF6 was detected from YFP-NUDT21 immunoprecipitation. (**B, C**) Precipitated proteins were analyzed by mass spectrometry. Venn diagram shows the numbers of the shared and the unique proteins identified in the control (YFP) and YFP-NUDT21 immunoprecipitation. The unique proteins in the YFP-NUDT21 immunoprecipitation are listed in panel C. CPSF6 and CPSF7 are known interaction partners of NUDT21, whereas hnRNPA3 or RPS7 has not been reported as an interaction partner of NUDT21 before.

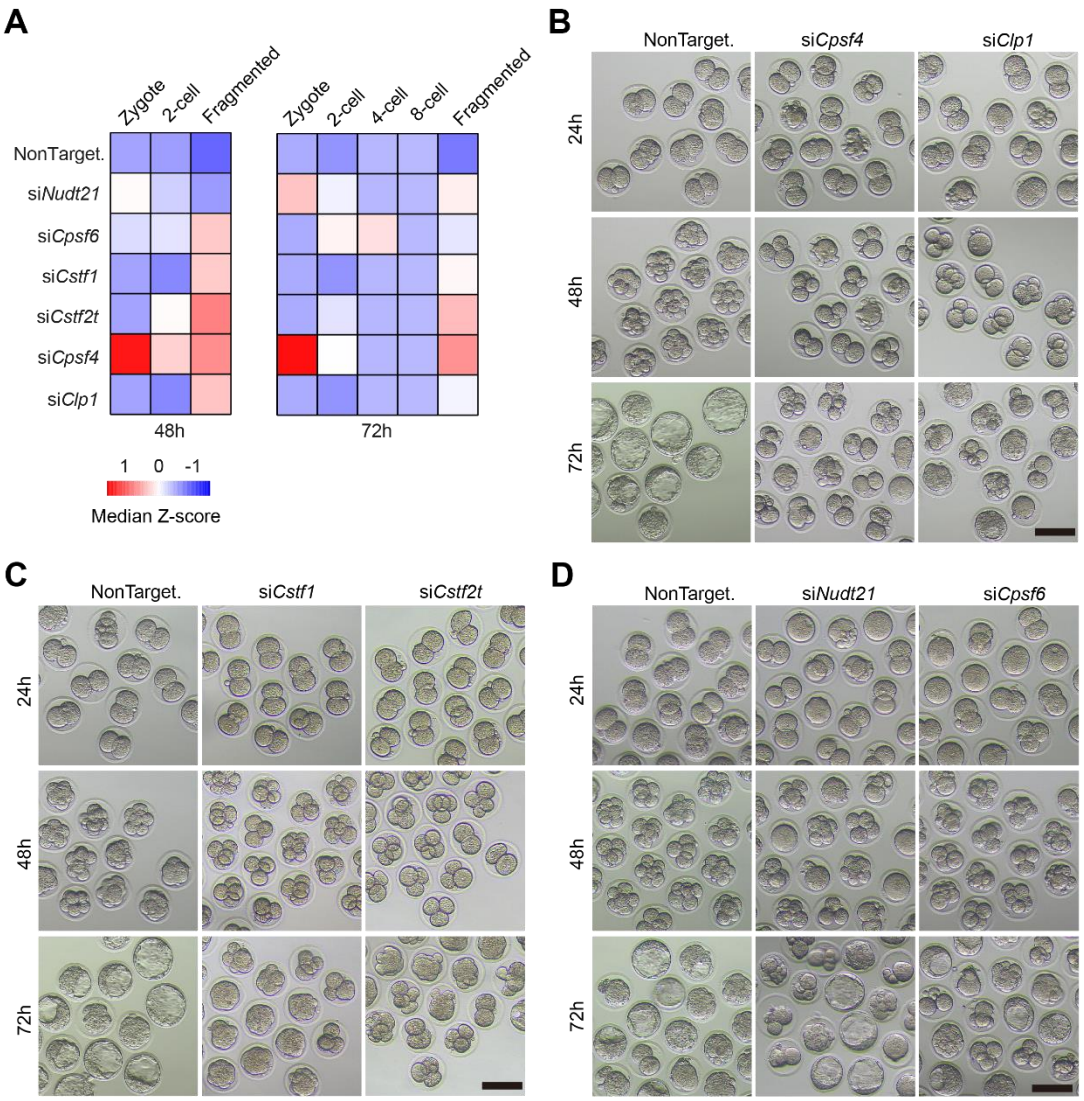


Figure S6. Analysis of mouse early embryo development *in vitro* after siRNA injection, related to Figure 6

(**A**) The distribution of abnormal embryos (arrested or fragmented) at 48 and 72 h after siRNA microinjection (the number of independent experiments is indicated in Figure 6B). (**B**–**D**) Bright field images of early embryos at 24, 48 and 72 h after microinjection of the siRNAs into mouse zygotes. Scale bar, 100 μm.

Table S1. Primary antibodies and oligonucleotides, related to experimental procedures

Table S2. Functional clustering of genes with 3'UTR extension in human morulae and mouse 4-cell embryos, related to Figure 3

Movie S1. Immunostaining of NUDT21 (green) and Lamin A/C (red) in a D3 human embryo, related to Figure 5A

Z-stack images were compiled into a movie to visualize the whole embryo that is shown in Figure 5A.

Supplemental Experimental Procedures

Cell culture

HEK-293 and HeLa cells (Procell) were cultured in MEM (Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% Pen/Strep (Gibco). 3T3-L1 cells (Procell) were cultured in DMEM (Gibco) with supplements as above. Cells were incubated at 37°C with 5% CO_2 in a humidified incubator. Trypsin (Gibco) was used for cell passaging.

Mice

Adult mice of the ICR strain were housed in the pathogen-free, temperature and humidity-controlled facility of Huazhong University of Science and Technology. The condition was 22 ± 2°C, 50% humidity and 12-h light/dark cycles. Mice were free to access water and food.

siRNAs

siGENOME SMARTpool siRNAs (Dharmacon), each containing 4 independent oligonucleotides targeting the same transcript, were dissolved at 100 μM in 1 x RNAi buffer, and stored at -80°C before use. NonTargeting siRNAs were used as a negative control. The nucleotide sequences of the siRNAs are listed in Table S1.

Plasmids

An *NUDT21* cDNA clone was ordered from Dharmacon (Cat. # MMM1013-202799195). The coding region was cloned into a destination vector containing a YFP tag by Gateway cloning according to a previous protocol (Li et al., 2016).

Antibodies

Primary antibodies used in this work are listed in Table S1.

Transfection

Plasmids were transfected into HEK-293 cells using Effectene Transfection Reagent (QIAGEN) according to the instruction. Briefly, cells were seeded into 60-mm culture dishes the day before transfection. 0.6 μg of the plasmid was mixed with Enhancer at the ratio of 1 μg of DNA to 8 μl of Enhancer, incubated at room temperature (RT) for 5 min, and then Effectene was added at the ratio of 1 μg of DNA to 12.5 μl of Effectene. The transfection complex was incubated at RT for 5 min, and then added into the culture medium with gentle mixing. Cells were harvested for immunoprecipitation 24 h after transfection.

siRNAs were transfected into 3T3-L1 cells using INTERFERin transfection reagent (Polyplus), according to a previous protocol (Li et al., 2016). Cells were harvested for immunoblotting 72 h post transfection.

Microinjection of siRNAs into mouse zygotes

Microinjection was performed using an Eppendorf TransferMan micromanipulator (Eppendorf). Mouse zygotes collected from the oviduct were immediately transferred into M2 medium covered with mineral oil. 5–10 pl of an siRNA solution was microinjected into each zygote. Zygotes were then cultured in EmbryoMax KSOM medium (Sigma-Aldrich, cat. # MR-106) at 37°C with 5% CO_2 in a humidified incubator, with the surface covered with mineral oil. Bright

field images of the embryos were taken with a Nikon Eclipse Ti2-U microscope (Nikon) every 24 h. At 72 h, the embryos were collected after imaging, and fixed for immunostaining.

The number of blastomeres within a single embryo in the exported bright field images was counted manually to determine the embryonic stage. At least 40 embryos per group were microinjected and assessed. Percentages of normal embryos of different stages at 3 time points were presented.

Immunostaining

The fixed embryos were washed twice with PBS containing 3% goat serum, blocked in blocking buffer (PBS containing 3% goat serum and 0.1% Triton X-100) at RT for 1 h, and then incubated with primary antibodies diluted in blocking buffer at 4°C overnight. Afterwards, embryos were washed three time with blocking buffer (5 min for each wash), and then incubated with Alexa fluorophore-conjugated secondary antibodies diluted in blocking buffer at RT for 1 h. After three washes as described above, embryos were counter-stained with DAPI at RT for 5 min, and mounted on confocal culture dishes.

HeLa cells were immunostained as previously described (Li et al., 2015).

Immunoprecipitation

24 h after plasmid transfection, cells were harvested with ice-cold scrapping buffer (1 mM Na_3VO_4, 50 mM NaF, 50 mM β-glycerophosphate, 1 mM PMSF, 10 μg/ml antipain, 10 μg/ml chymostatin, 100 μg/ml pepstatin A, 2 μg/ml leupeptin, 200 μg/ml AEBSF-HCl and 2 μg/ml aprotinin in PBS), and centrifuged at 4°C, 1160 g for 15 minutes. Pelleted cells were re-suspended in 100 μl of 2 x extraction buffer (20 mM Tris-HCl, 150 mM NaCl, 1% IGEPAL, 1 mM EGTA and protease and phosphatase inhibitors as above), and rotated at 4°C for 1 h. Cell extracts were cleared by centrifuging at 4°C, 21380 g for 15 min. Protein G Dynabeads (Thermo Fisher Scientific, cat. # 10003D) conjugated with an anti-GFP antibody (Roche, cat. # 11814460001) were washed with PBS-T and then with 1x extraction buffer, mixed with the cleared extracts, and incubated at 4°C for 4 h with constant rotation. The beads were washed 5 times with 1 x extraction buffer, and then the precipitated proteins were eluted with 2 x Laemmli buffer. The eluates were denatured at 100°C for 10 min and stored at −20°C.

Protein identification by mass spectrometry

The precipitated proteins were analyzed by mass spectrometry at BGI Genomics. Briefly, 100 μg of total proteins was digested with trypsin at 37°C for 4 h. The digested peptides were desalted and freeze-dried. The dried peptides were reconstituted and separated on a Thermo UltiMate 3000 UHPLC system (Thermo Fisher Scientific). The separated peptides were ionized by a nanoESI source and then passed to a tandem mass spectrometer Q-Exactive HF X (Thermo Fisher Scientific) for detection in the data‐dependent acquisition (DDA) mode. More details and the parameters were described by Ye and co-authors (Ye et al., 2021).

The raw mass data were converted into a peak list, and searched against the Uniprot database using Mascot 2.3.02 (Matrix Science, London, UK). The identified peptides were pre-processed and re-scored using Percolator (http://percolator.ms/) to improve the matching accuracy. The peptide spectrum matches (PSMs) were then filtered at a false discovery rate

(FDR) of 1% to obtain a list of significantly identified spectra and peptides, which were used for protein inference.

SDS-PAGE and immunoblotting

Immunoblotting was performed as described before (Li et al., 2015).

RNA-sequencing of 4-cell embryos

A protocol for scRNA-seq was employed. Zygotes were microinjected and cultured as described above. 48 h later, the 4-cell embryos were collected and transferred into RNA lysis buffer provided by Annoroad Gene Technology Co. Ltd. Seven embryos were pooled as one replicate for each group and stored at −80°C before use.

RNAs were reversely transcribed into first strand cDNAs by the Smart-seq2 method (Picelli et al., 2014). Briefly, the RT mix, supplemented with a reverse transcriptase, oligo-dT primers with a common sequence (5'-AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN-3'), TSO primers, an RNA inhibitor, betaine and dithiothreitol, was added to the lysed samples to obtain a final volume of 10 μl for each reaction, and then the samples were incubated at 42°C for 90 min on a thermal cycler. The second strand cDNAs were synthesized by PCR amplification using IS PCR primers. The cDNAs were purified with Beckman Ampure XP magnetic beads. The concentration and the quality of the cDNAs were evaluated with Agilent 2100 RNA Nano 6000 kit (Agilent).

RNA-seq was performed at Annoroad Gene Technology Co. Ltd. Briefly, a total amount of 20 ng of the amplified cDNAs was used for library construction. Full-length cDNAs were fragmented to the size of ~200 bp by a Bioruptor sonication system (Diagenode), followed by end-repair, adapter ligation and PCR amplification. The libraries were purified and assessed as described above. The indexed libraries were sequenced on a DNBSEQ T7 platform (MGI) to generate 350-bp paired-end reads.

Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was performed using the Single Cell Sequence Specific Amplification kit (Vazyme). Five to ten embryos of each developmental stage were collected and transferred into the reaction mix, containing the reaction buffer, RT/Taq enzymes and primer pool (0.1 μM). The samples were stored at -80°C before use. cDNAs were synthesized on a thermal cycler according to the instruction of the kit and diluted at 1:10 prior to qPCR. qPCR was performed with Taq Pro Universal SYBR qPCR Master Mix (Vazyme) on a Quantagene q225 real-time PCR system (Kubo Technology). The relative expression levels of genes were calculated using the $2^{-\Delta Ct}$ method, with *Tuba1a* as the internal control for normalization. The primers used in qPCR are listed in Table S1.

Embryo scoring

Embryos were categorized as cortical NUDT21 positive or negative (corNUDT21+ve, corNUDT21−ve), on the basis of the intensity of NUDT21 at the cortex that was plotted using the Profile function of ZEN.

Utilizing the Z-stack images, human embryos were scored according to the morphological criteria purposed by De Placido and colleagues (De Placido et al., 2002). In total, 40 D3 human

embryos were evaluated. Embryos with less than 8 blastomeres were defined as no division. Multinucleation was determined by DAPI and anti-Lamin A/C staining, whereas fragmentation and blastomere sizes were determined by the cytoplasmic signal of anti-NUDT21 staining.

Mouse embryos were scored similarly based on the morphological criteria. At least 30 embryos per groups collected at 72 h post microinjection were assessed.

Imaging and analysis

Whole embryos were imaged on a LSM780 or LSM900 confocal microscope coupled with a PMT detector (Zeiss), at an interval of 2.2 μm via the Z-stack function (20 slices per embryo). Maximum projection was applied to the stacks to obtain a composite image for each embryo. The representative Z-stack image of each embryo was exported using the software ZEN 3.1 (Zeiss), and resized to 300 ppi without re-sampling using Photoshop 2021 (Adobe).

The fluorescence intensities of NUDT21 and CPSF6 within the nucleus were quantified using CellProfiler (Jones et al., 2008). Briefly, all the Z-stack images were exported using ZEN 3.1; images of blastomeres were manually selected and loaded onto CellProfiler. The nuclear area was defined by DAPI intensity. The average intensity of NUDT21 and CPSF6 within the nuclear area of a single blastomere was measured. The mean intensity in blastomeres of each embryo was calculated.

Bioinformatics analysis

Raw sequencing data were either retrieved from the GEO repository or generated in this study. Reads were re-mapped to the reference genomes, hg38 for the human and mm10 for the mouse, respectively. Low-quality reads were filtered out; the remained were analyzed by APAtrap to quantify the changes of 3′UTR lengths. Log_2 value of percentage difference (PD) between two pre-implanted stages was calculated. To avoid the clipping of PD values with the axes, a value of 0.0009 (slightly smaller than the minimal PD of 0.001) was added to those PD of 0 (data points lined in parallel to the axes in Fig.1C, D).

DESeq2 was used to identify differentially expressed genes with a threshold of "log_2 fold change ≥ 1" and an adjusted p value < 0.05. ClusterProfiler and DAVID were employed for pathway enrichment analysis (Dennis et al., 2003; Yu et al., 2012). The STRING database was used for protein network analysis (Szklarczyk et al., 2019). Cytoscape was used for network visualization (Shannon et al., 2003).

Supplemental References

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4, P3.

De Placido, G., Wilding, M., Strina, I., Alviggi, E., Alviggi, C., Mollo, A., Varicchio, M.T., Tolino, A., Schiattarella, C., and Dale, B. (2002). High outcome predictability after IVF using a combined score for zygote and embryo morphology and growth rate. Human reproduction (Oxford, England) *17*, 2402-2409. 10.1093/humrep/17.9.2402.

Jones, T.R., Kang, I.H., Wheeler, D.B., Lindquist, R.A., Papallo, A., Sabatini, D.M., Golland, P., and Carpenter, A.E. (2008). CellProfiler Analyst: data exploration and analysis software for complex image-based screens. BMC Bioinformatics *9*, 482. 10.1186/1471-2105-9-482.

Li, H., Frappart, L., Moll, J., Winkler, A., Kroll, T., Hamann, J., Kufferath, I., Groth, M., Taudien, S., Schütte, M., et al. (2016). Impaired Planar Germ Cell Division in the Testis, Caused by Dissociation of RHAMM from the Spindle, Results in Hypofertility and Seminoma. Cancer research *76*, 6382-6395. 10.1158/0008-5472.Can-16-0179.

Li, H., Moll, J., Winkler, A., Frappart, L., Brunet, S., Hamann, J., Kroll, T., Verlhac, M.H., Heuer, H., Herrlich, P., and Ploubidou, A. (2015). RHAMM deficiency disrupts folliculogenesis resulting in female hypofertility. Biology open *4*, 562-571. 10.1242/bio.201410892.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature protocols *9*, 171-181. 10.1038/nprot.2014.006.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498-2504. 10.1101/gr.1239303.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47, D607-d613. 10.1093/nar/gky1131.

Ye, X., Zhao, S., Wu, M., Ruan, J., Tang, X., Wang, X., and Zhong, B. (2021). Role of sericin 1 in the immune system of silkworms revealed by transcriptomic and proteomic analyses after gene knockout. FEBS open bio *11*, 2304-2318. 10.1002/2211-5463.13239.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16, 284-287. 10.1089/omi.2011.0118.