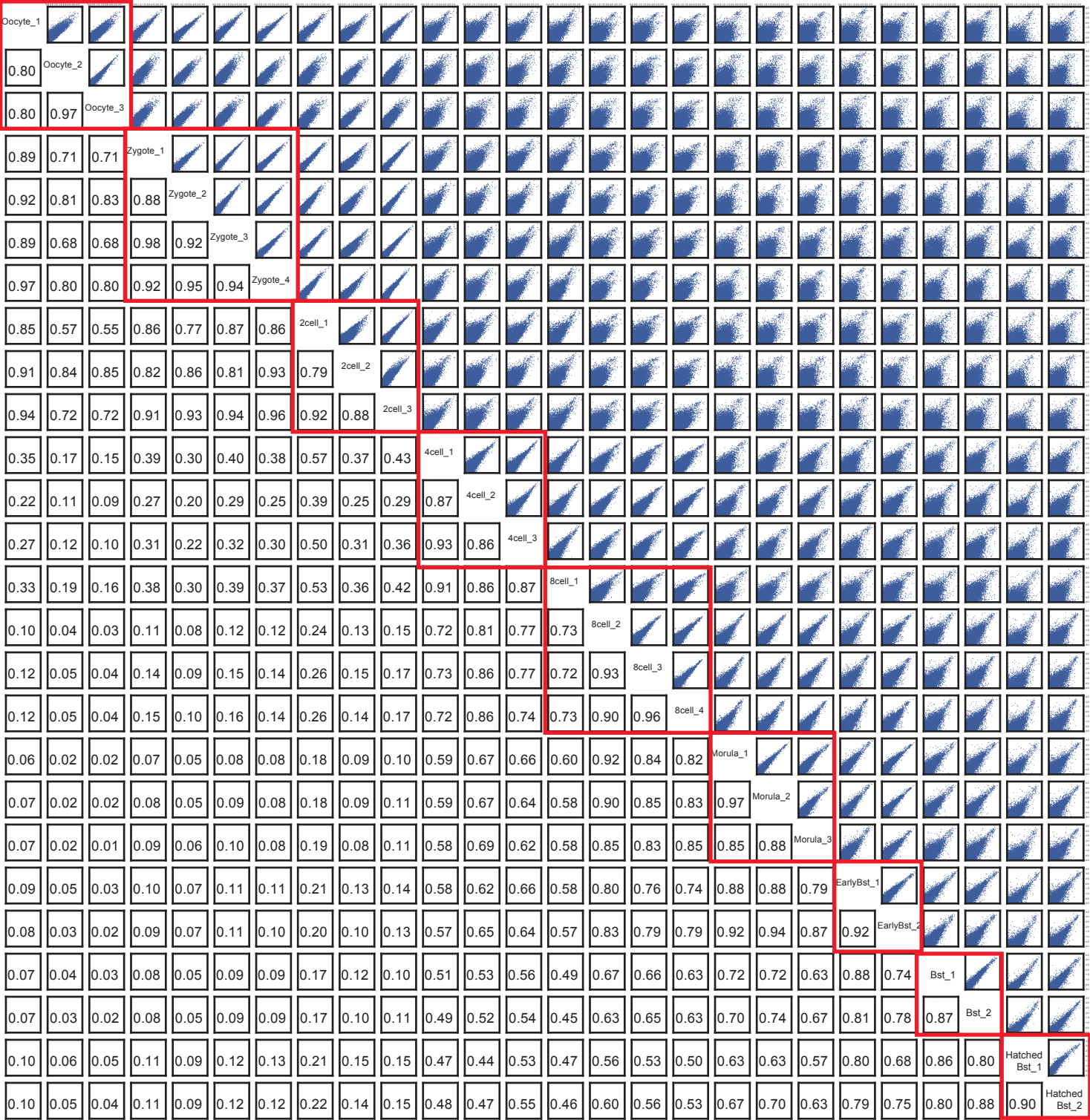
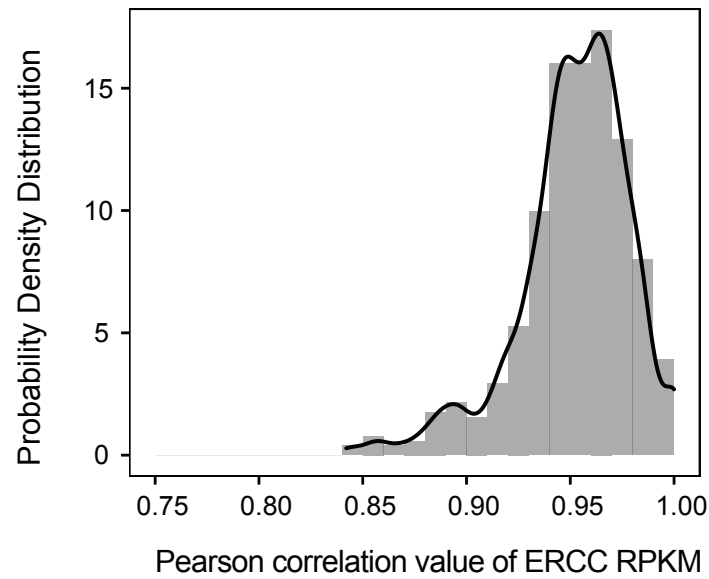


Supplementary Information

A



B



C

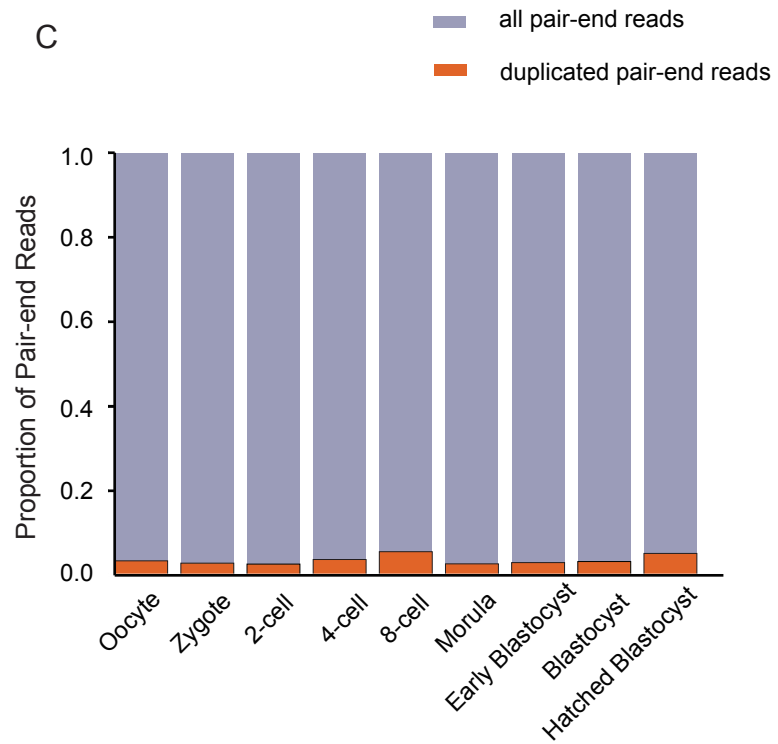
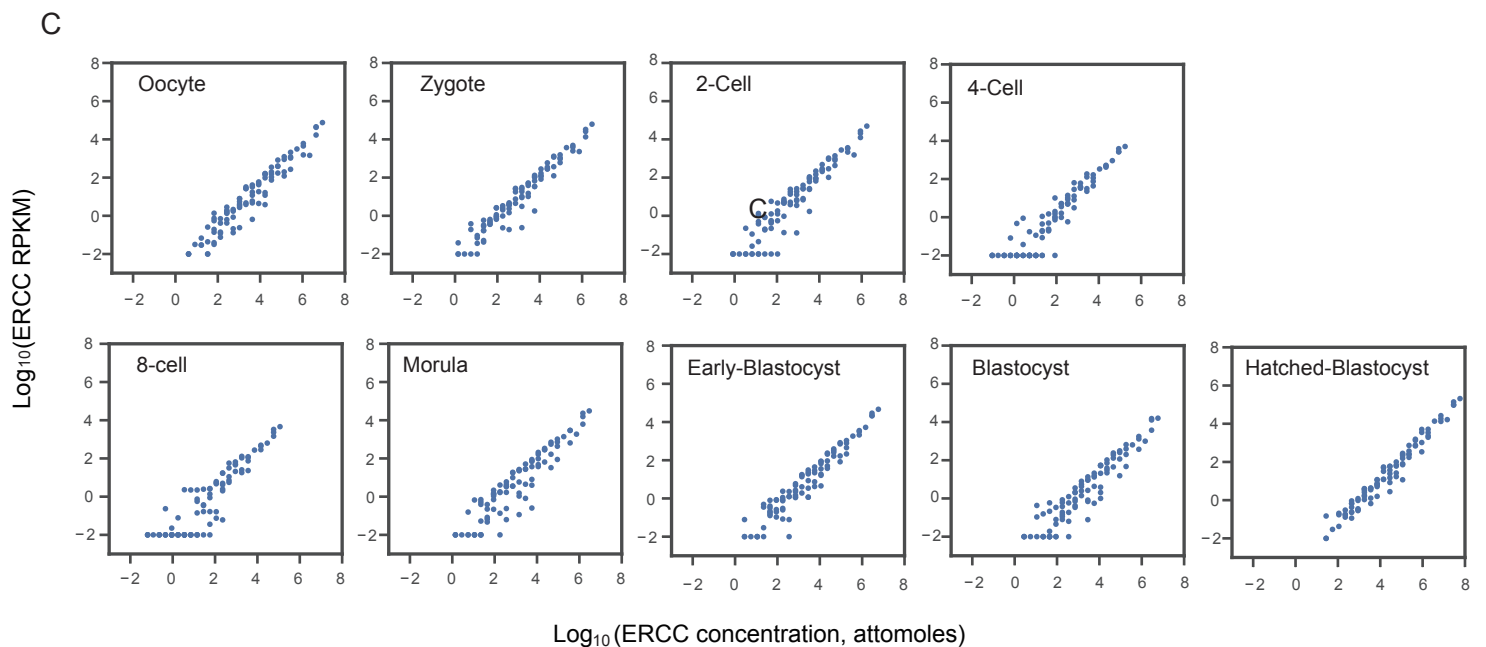
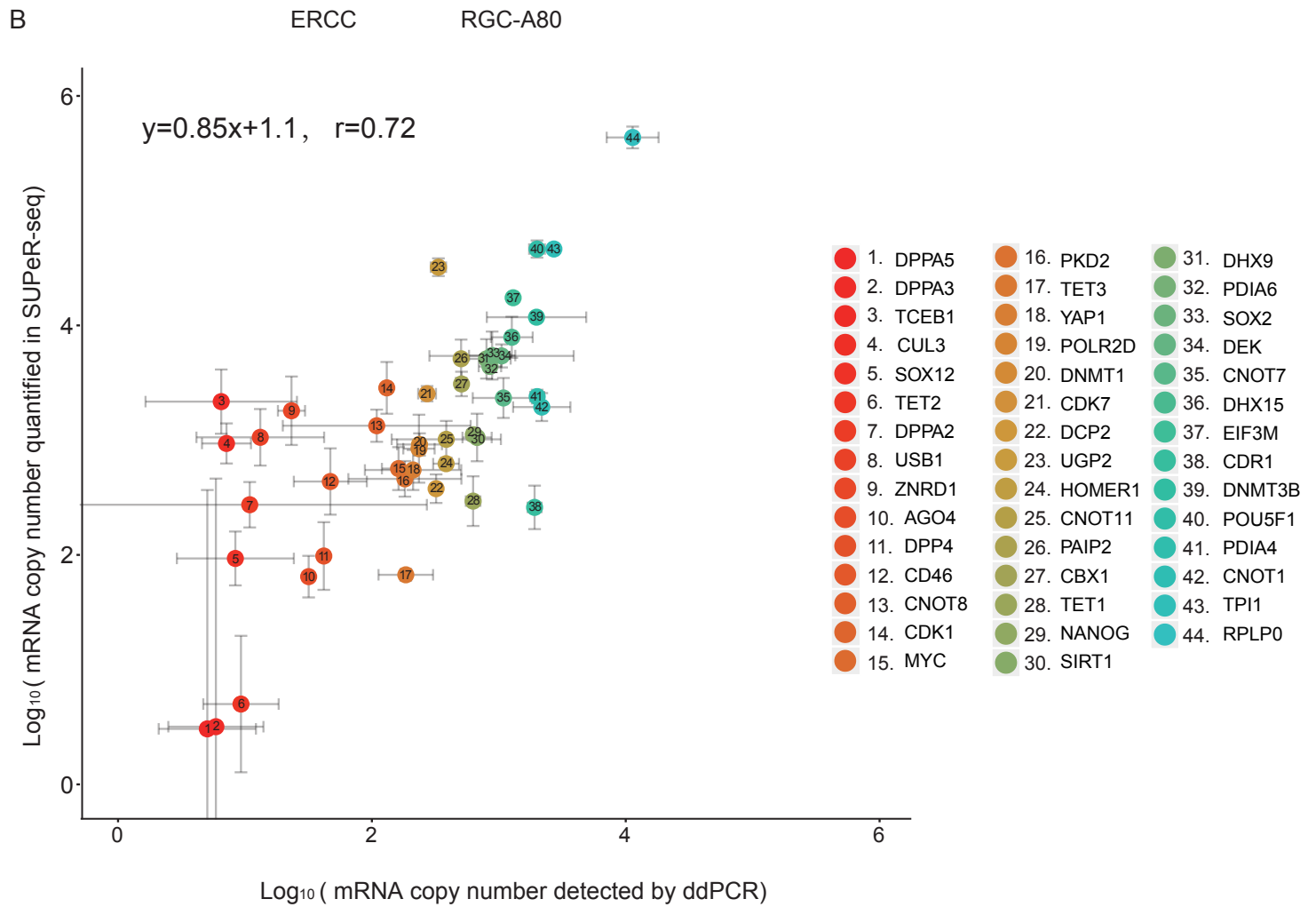
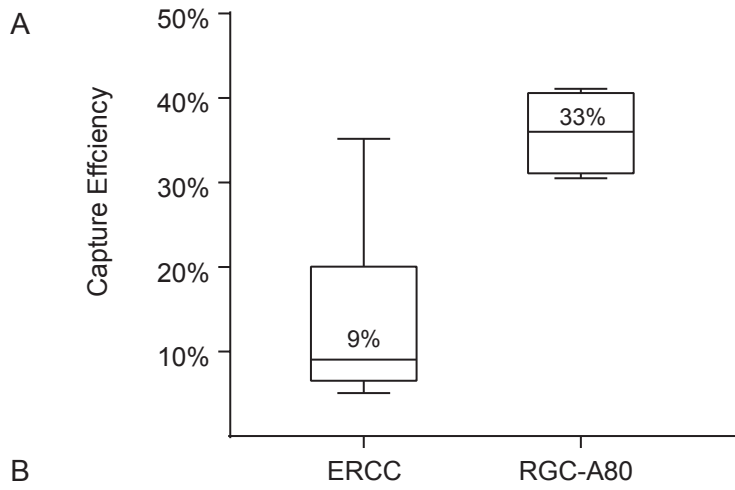


Figure S1. Reproducibility of the SUPeR-seq method.

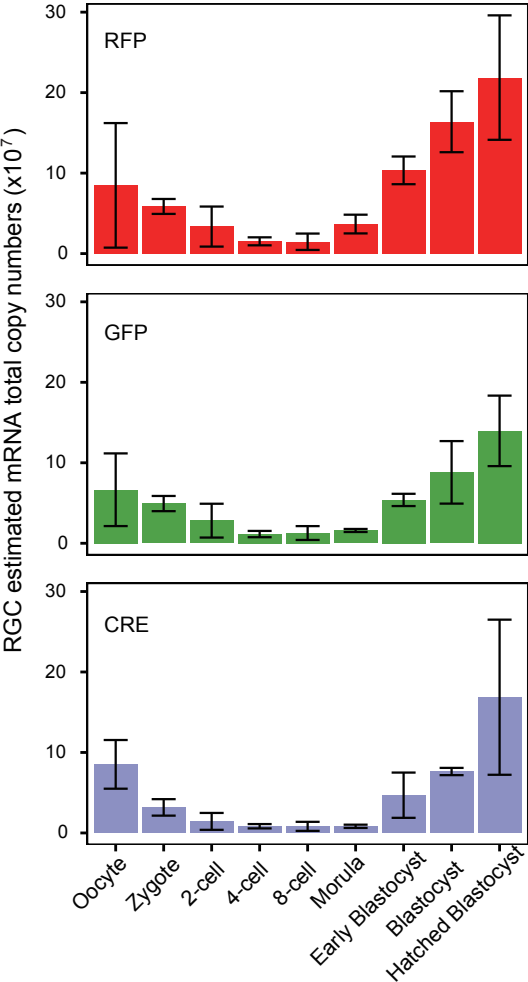
(A) Pearson correlation coefficients of oocytes and embryos during human pre-implantation development. R squared value of pair-wise correlation between each two samples are shown. Blastocyst is abbreviated to Bst.

(B) Distribution of Pearson correlation coefficients of ERCC spike-in RNAs during human pre-implantation development.

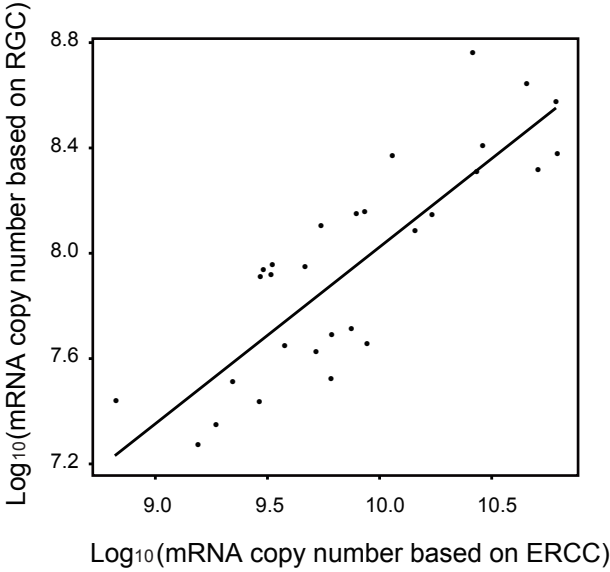
(C) Proportion of duplicated pair-end reads in all pair-end reads by SUPeR-seq of human early embryos.



D



E



F

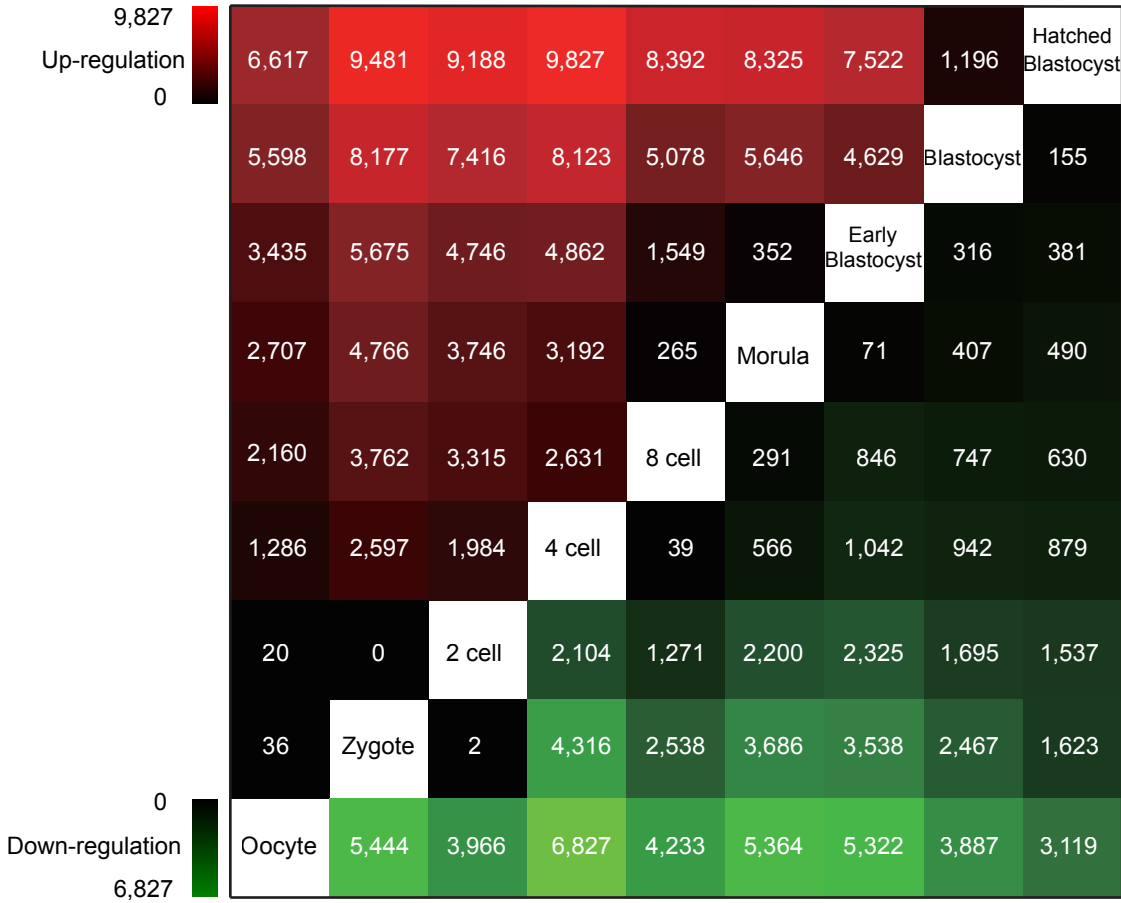
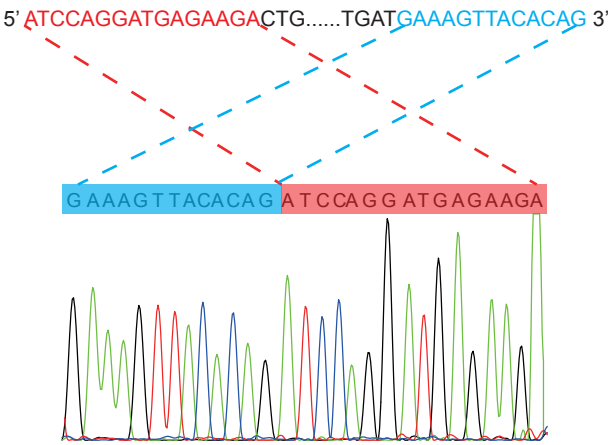


Figure S2. mRNA copy numbers quantified by ERCC/RGC-A80 normalization algorithm.

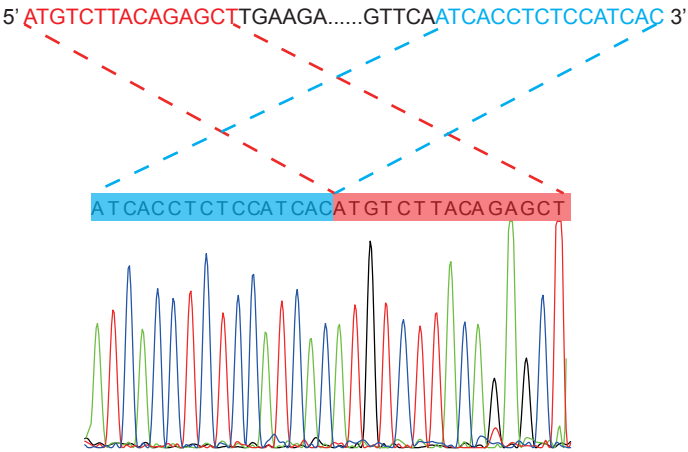
- (A) Capture efficiency of ERCC and RGC-A80 during reverse transcription in SUPeR-seq tested by ddPCR. These two spike-ins were reverse transcribed at same masses following the protocol of SUPeR-seq, and the unamplified cDNA were used as templates in ddPCR reaction for ten ERCC genes, and RFP and GFP twice independently.
- (B) Validation of the quantification in SUPeR-seq through detection the copy number of 44 genes in 1ng hESCs total RNAs by ddPCR. The copy numbers of these 44 genes estimated in SUPeR-seq were distributed from the total mRNA copy number by the ratio of each gene's RPKM to the total RPKM of all RefSeq genes. The correlation between the data from ddPCR and SUPeR-seq is 0.72. The linear regression formula based on these genes is $y=0.85x+1.1$. Errors bars show the 95%CI for individual ddPCR measurements (horizontal bars, $n=2$) or across SUPeR-seq (vertical bars, $n=3$).
- (C) Linear regression of 92 ERCCs between the copy number and RPKM at each stage.
- (D) mRNA copy number of RefSeq genes per embryo estimated based on RGC-A80 spike-ins.
- (E) Linear regression of the mRNA copy numbers per embryo based on two sets of spike-in molecules, $r=0.71$. The linear regression equation is $y=0.67x+1.3$, and x represents the log10-transformed ERCC estimated mRNA copy number, and y represents the log10-transformed RGC-A80 estimated mRNA copy number. The final mRNA copy number per embryo was the value for y that was calculated using the linear regression equation for each value of x .
- (F) Number of known RefSeq genes showing up-regulation and down-regulation during human pre-implantation development after size-factoring the sequencing depth and the mRNA quantity.

A

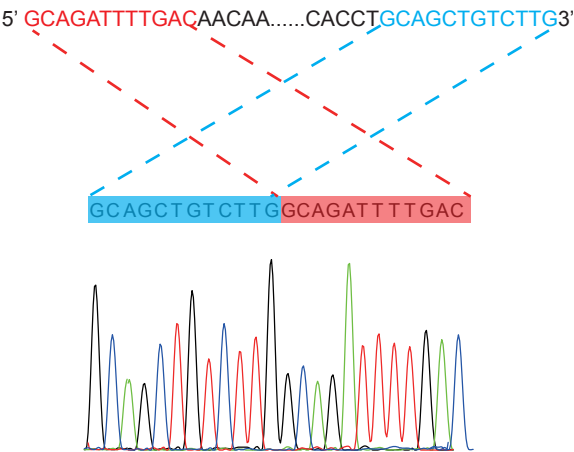
chr18:9182379-9221997 ANKRD12



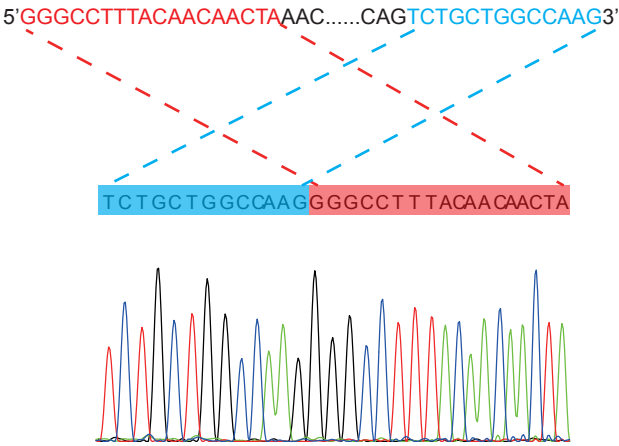
chr21:40619626-40630598 BRWD1



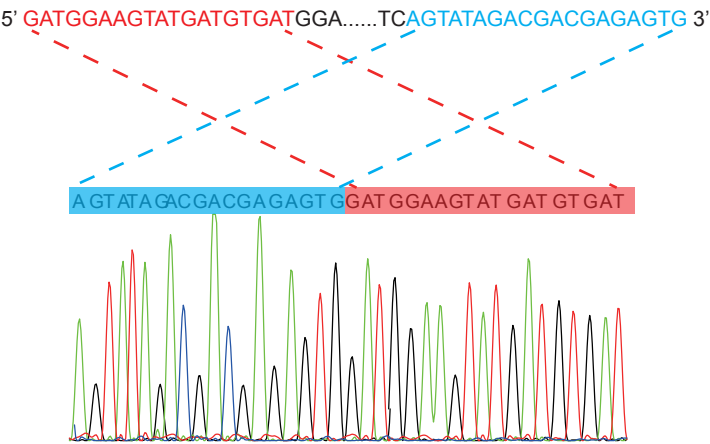
chr10:70404454-70406762 TET1

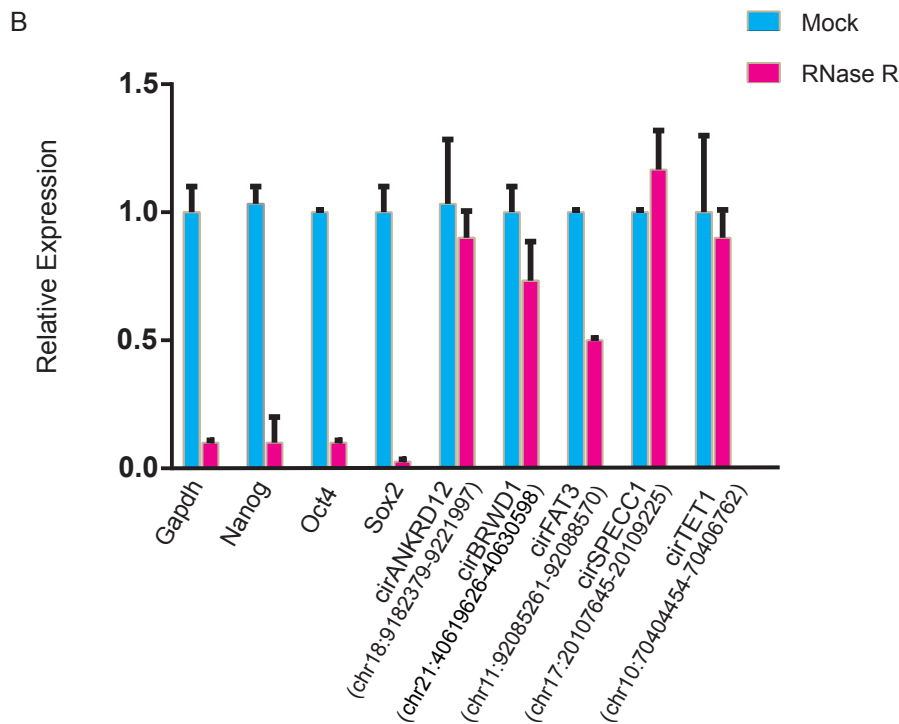


chr17:20107645-20109225 SPECC1

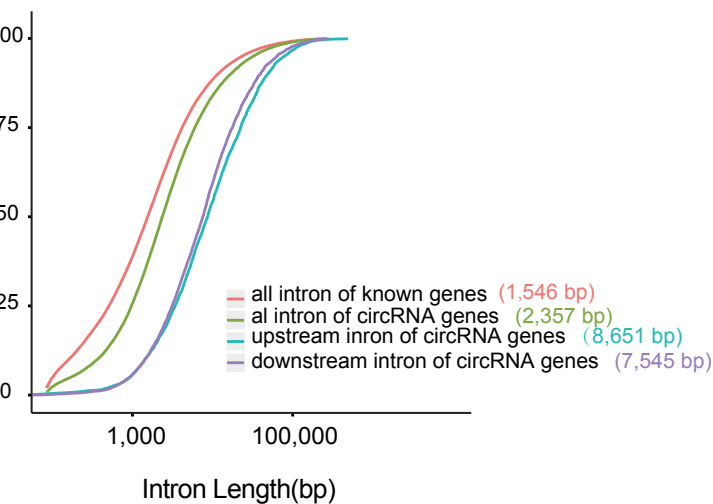


chr11:92085261-92088570 FAT3





C



D

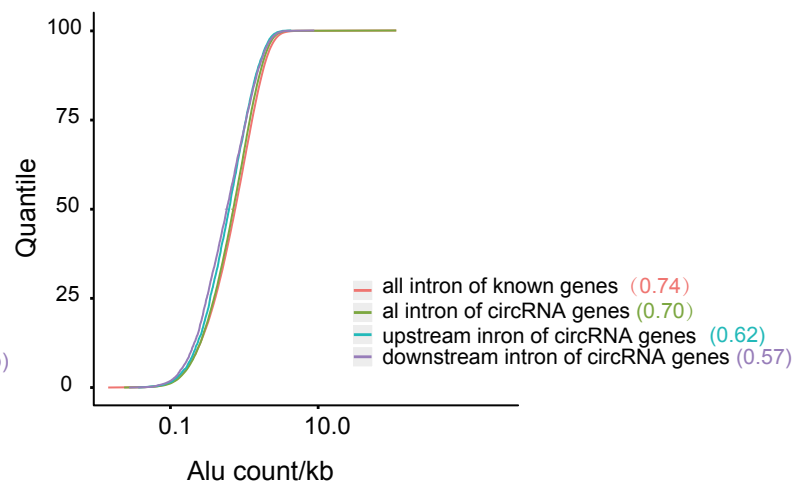


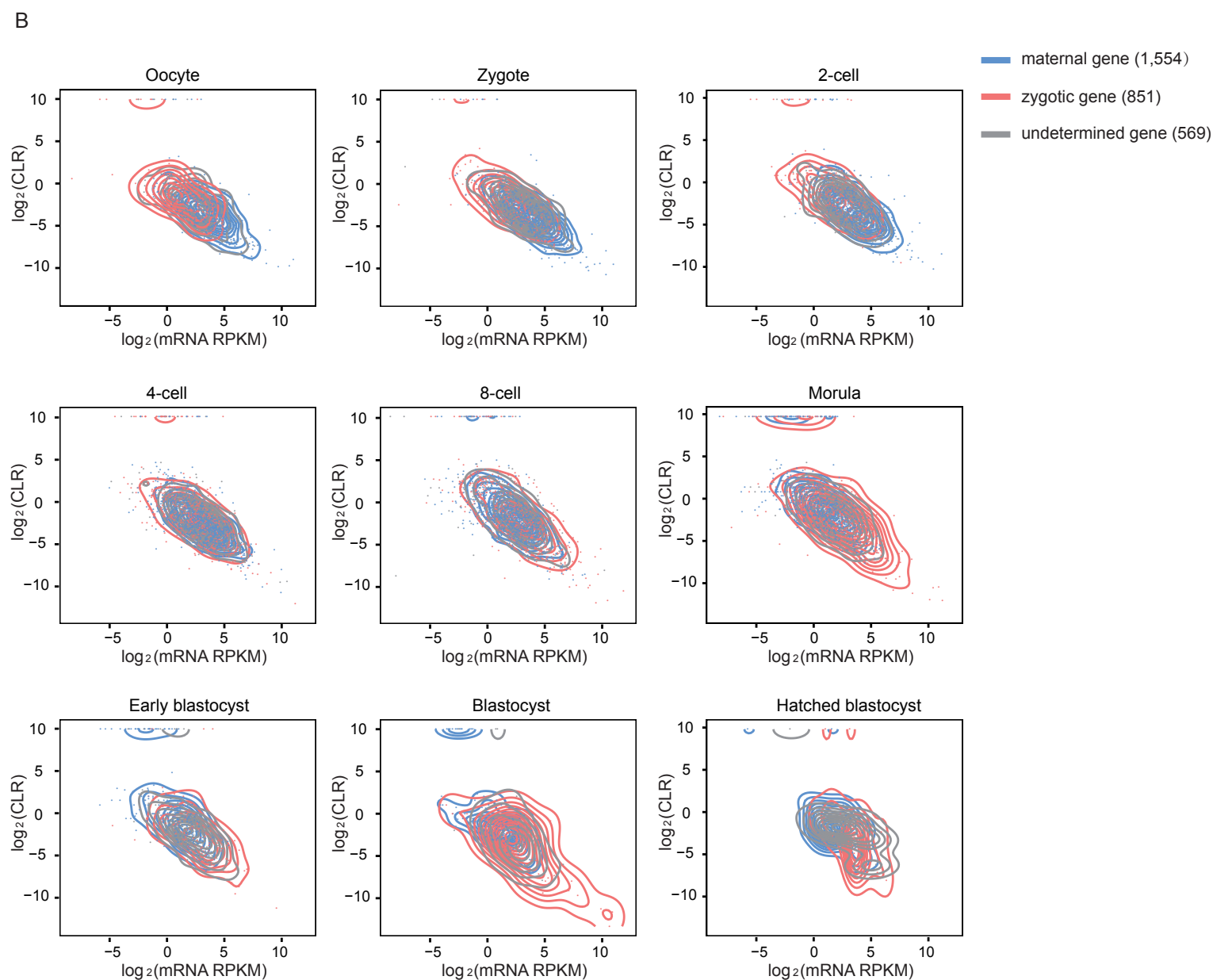
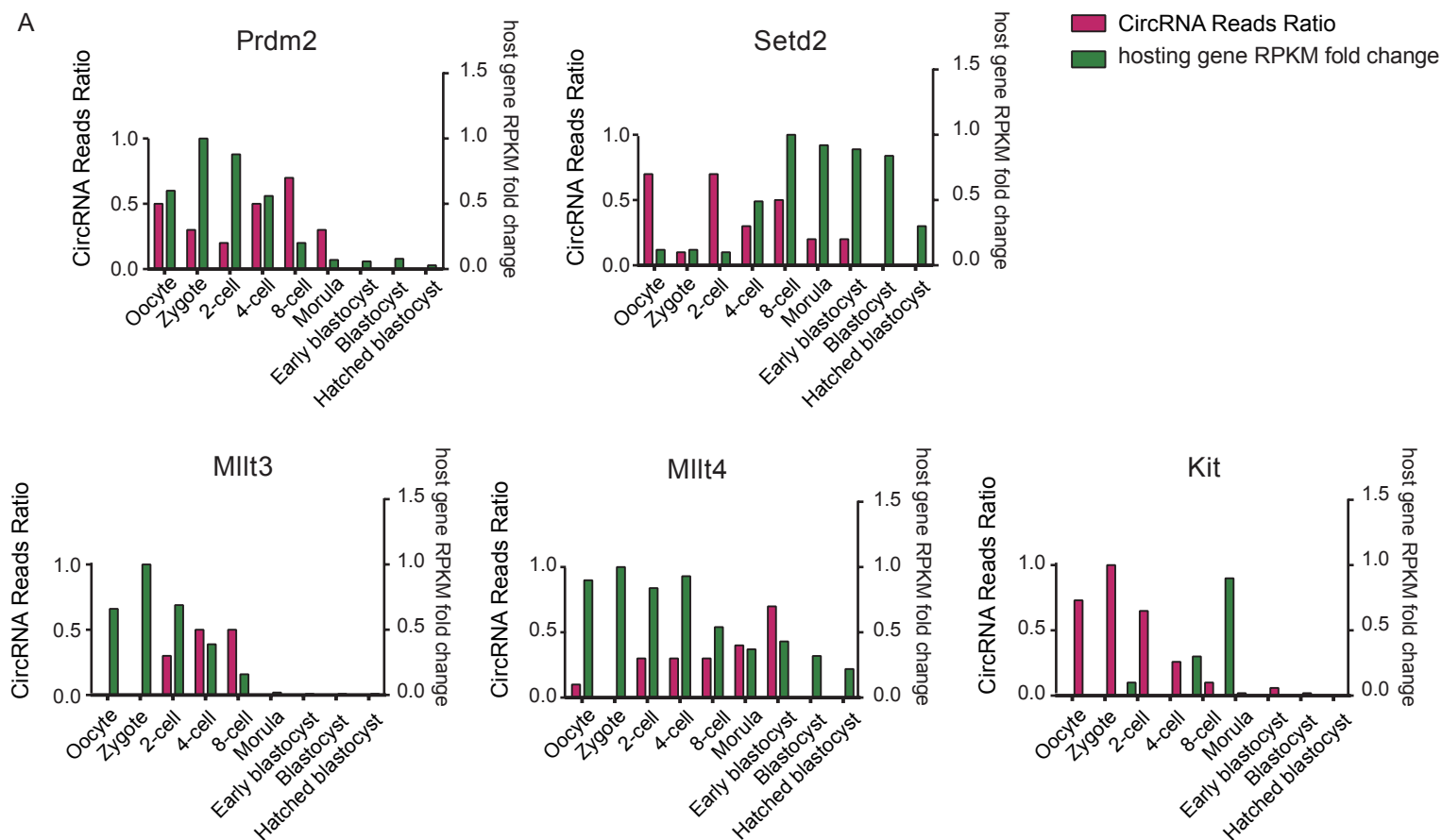
Figure S3. Feature analysis of the flanking introns of circRNAs.

(A) Validation of circRNA candidates in hESCs by Sanger seq. The back-spliced sites of these circRNAs are detected as a reversed order in the genome.

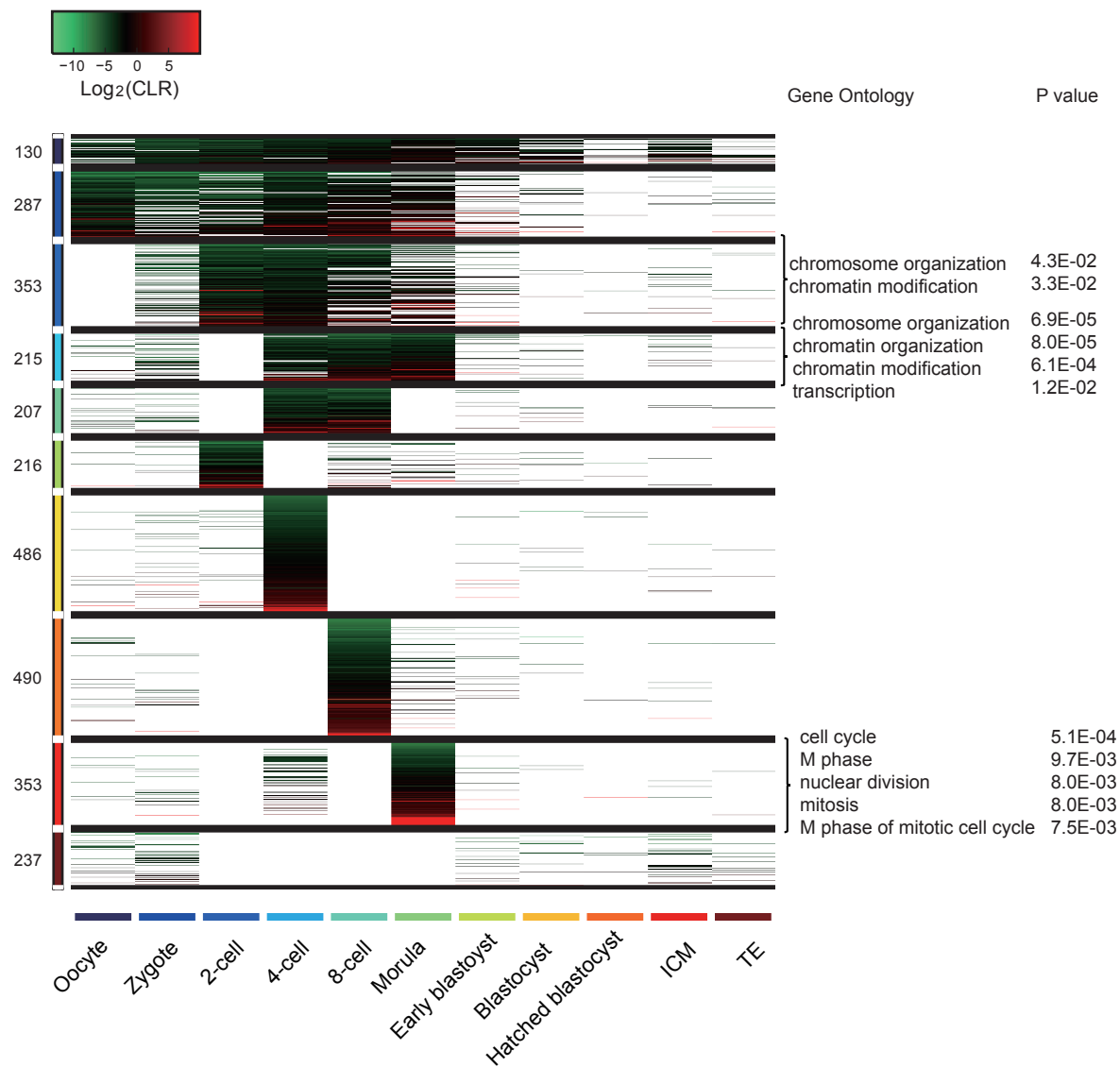
(B) Verification of circRNA candidates by RNase R digestion. The linear transcripts from four genes are degraded after RNase R treatment. Differently, these circRNA transcripts are resistant to the RNase treatment, proving their circularized characteristics.

(C) Length distribution of the flanking introns of circRNAs and all other introns. The upstream (blue line) and downstream (purple line) flanking introns are much longer than the control introns (green and red line). The median number is given in the bracket.

(D) Density of Alu elements in the flanking introns of circRNAs and all other introns. The Alu element count per kilobase in the flanking introns (upstream in blue and downstream in purple) is comparable with the randomly selected control introns regardless of their locations in circRNA host gene (in green) or not (in red). The median number is given in the bracket.



C



D

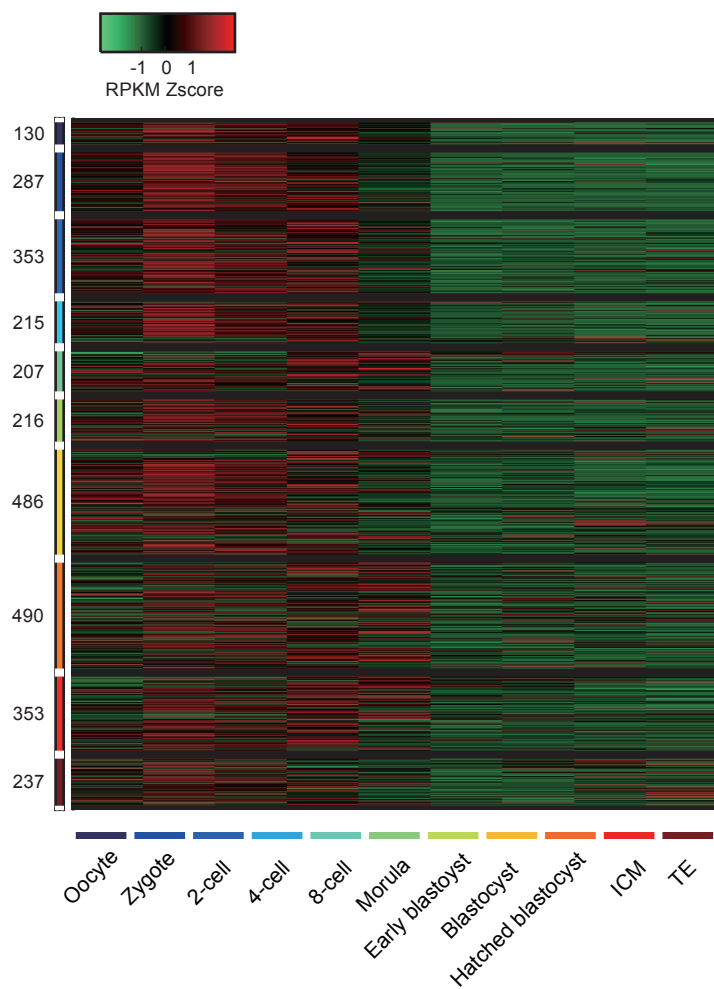


Figure S4. CircRNA and hosting gene expression patterns during human pre-implantation development.

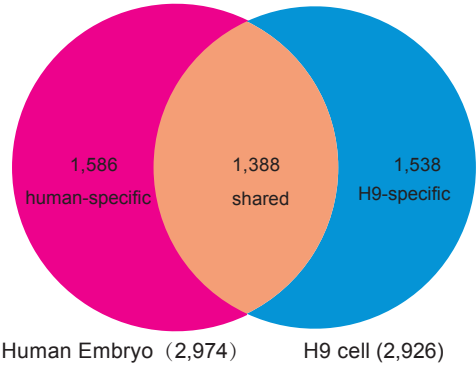
(A) Dynamics of the circRNA expression level and corresponding hosting gene RPKM fold changes during pre-implantation. Five genes, as PRDM2, SETD2, MLLT3, MLLT4 and KIT, were selected because their circular reads ratio exceeded 0.5 as their circular transcripts' levels were higher than the level of the linear counterparts. Histograms show the circular reads ratio (in red) and the hosting gene RPKM fold-change normalized by the highest value (in green).

(B) Plot of CLR (circular to linear ratio) against the RPKM of the hosting gene showing the significant negative relationship between the abundance of circRNA and the hosting gene. The plotted hosting genes are clustered as zygotic genes (pink dots), maternal genes (blue dots) and undetermined genes (grey dots) according to the differential expressed gene list (Figure 2B).

(C) The clusters of circRNA hosting genes based on CLR value during human pre-implantation development. The enriched GO terms and corresponding P values are shown at right.

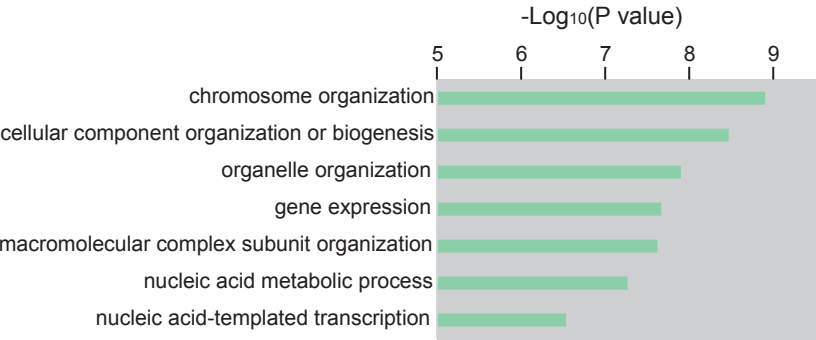
(D) The expression level of circRNA hosting genes. The genes are clustered according to Figure S4C.

A



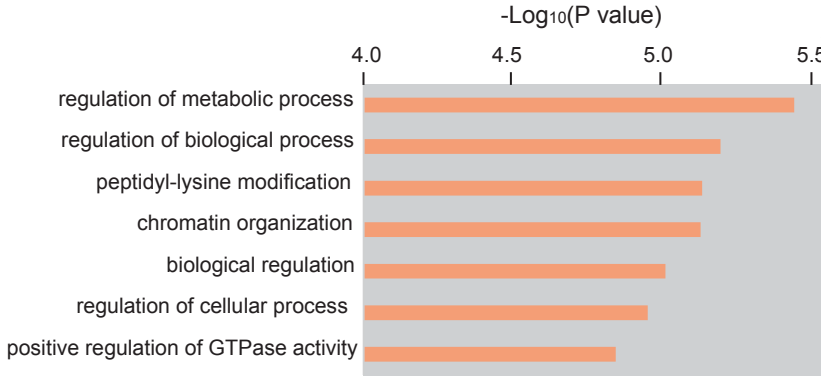
B

shared circRNA hosting genes (compared with mouse embryos, reads $\geq$ 2)

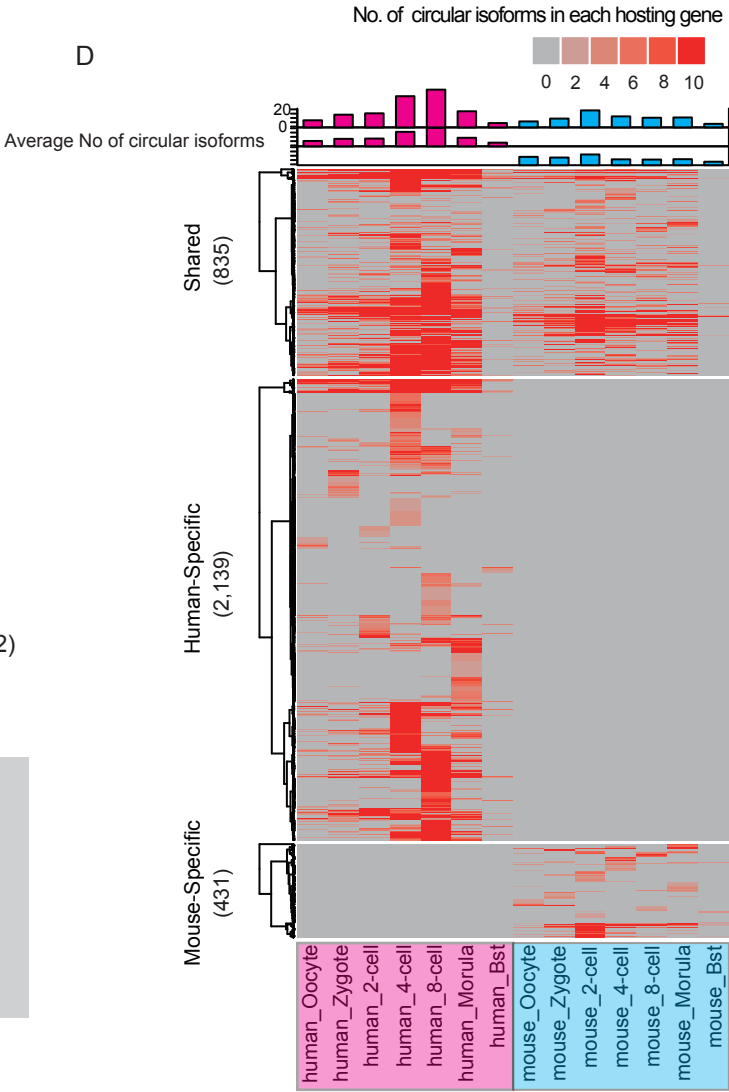


C

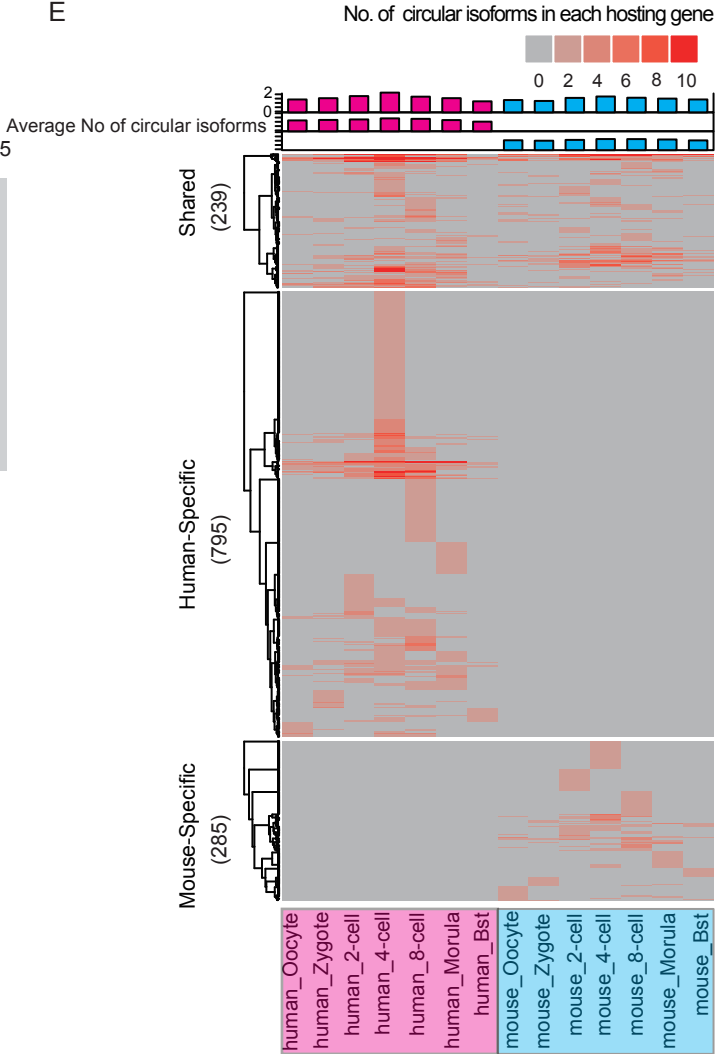
shared circRNA hosting genes (compared with H9 cells, reads $\geq$ 2)



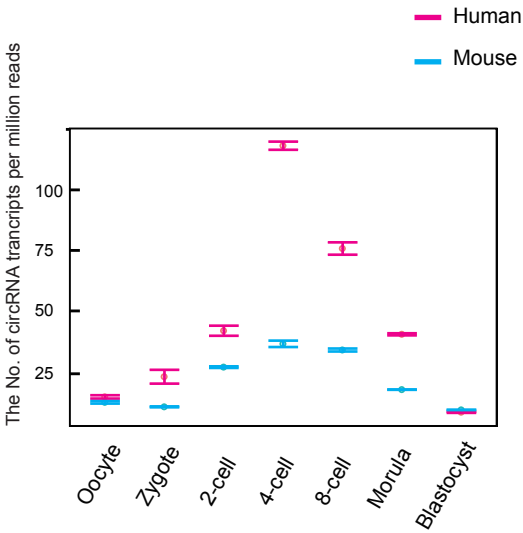
D



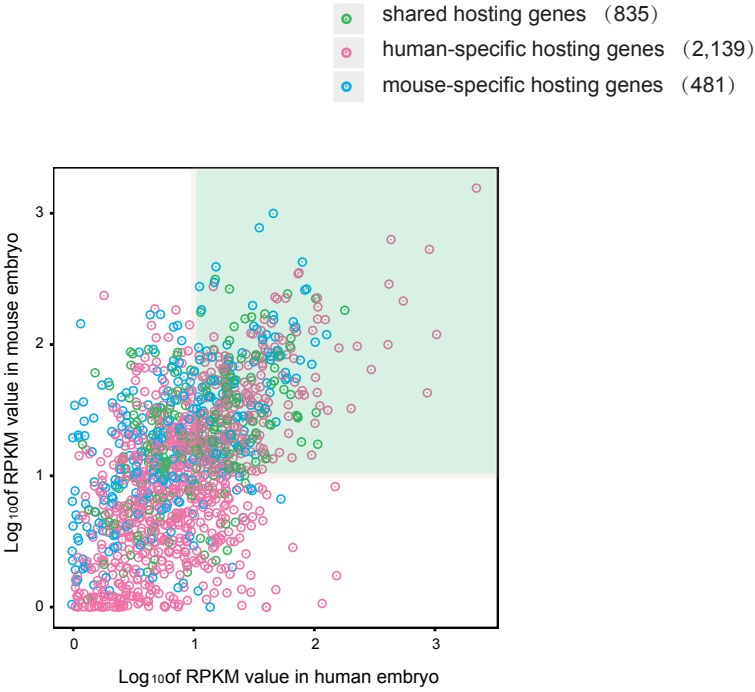
E



F



G



H

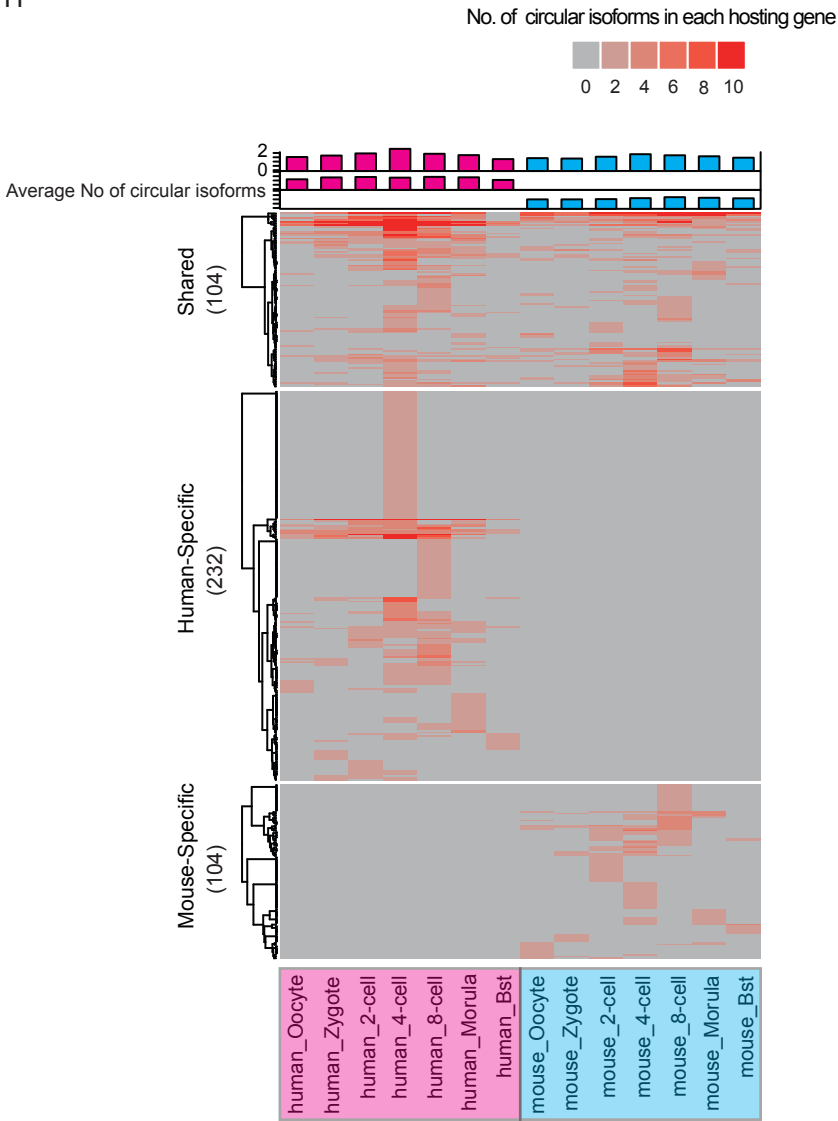


Figure S5. CircRNA hosting gene comparison.

- (A)** The Venn diagram shows that the majority of genes that express circRNAs in H9 cells also produce circular transcripts in human embryos.
- (B)** The top enriched gene ontology (GO) terms of the shared circRNA hosting genes between human embryos and mouse embryos are shown.
- (C)** The top enriched gene ontology (GO) terms of the shared circRNA hosting genes between human embryos and H9 cells are shown.
- (D)** Heat map of circularized level of all circRNA hosting genes during human and mouse pre-implantation development. The shared, human-specific and mouse-specific hosting genes are corresponding to the three parts of hosting genes from the Venn diagram in Figure 5A.
- (E)** Heat map of circularized level of circRNA hosting genes during human and mouse pre-implantation development after sub-sample the sequencing data.
- (F)** The box plots show the numbers of circular transcripts in human embryos (in red) and mouse embryos (in blue), during pre-implantation development, and the number was normalized to the total mapped reads (in millions) after subsampling the sequencing data.
- (G)** The scatter plot shows the distribution of the three kinds of circRNA hosting genes according to their expression level in human and mouse embryos. The values of $\log_{10}$ of (max RPKM during pre-implantation) are shown. And the highly expressed hosting genes are outlined as the RPKM >10, simultaneously in human and mouse embryos.
- (H)** Heat map of circularized level of circRNA hosting genes which are highly expressed during human and mouse pre-implantation development after subsampling the sequencing data.

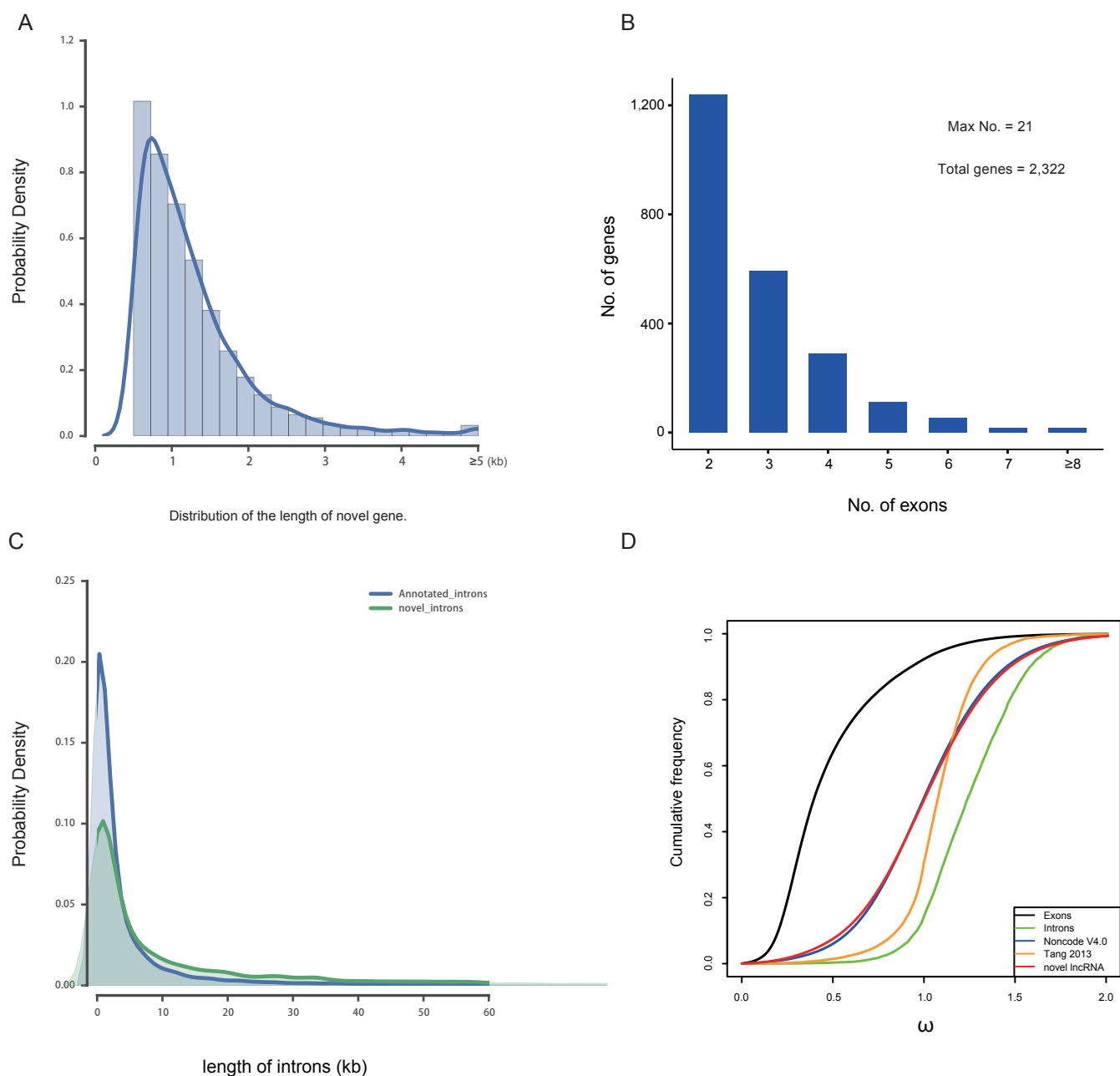


Figure S6. De novo assembled transcripts in the human pre-implantation embryos.

- (A)** Length distribution of the novel transcripts; candidates with lengths less than 500 bp are excluded.
- (B)** Exon number distribution of 2,322 novel genes; candidates with lengths less than 500 bp are excluded.
- (C)** The intron length distribution of the novel transcripts. The introns of novel genes are slightly longer than control introns of annotated genes.
- (D)** Conservation level (ω metric) of the novel transcripts compared with exons, introns, Noncode V4.0 database and novel lncRNAs identified previously, indicating that novel transcripts are potential lncRNAs.