Title

A temporal extracellular transcriptome atlas of human pre-implantation development.

Journal

Cell Genomics, 4(1)

Authors

Wu, Qiuyang

Zhou, Zixu

Yan, Zhangming

et al.

Publication Date

2024-01-10

DOI

10.1016/j.xgen.2023.100464

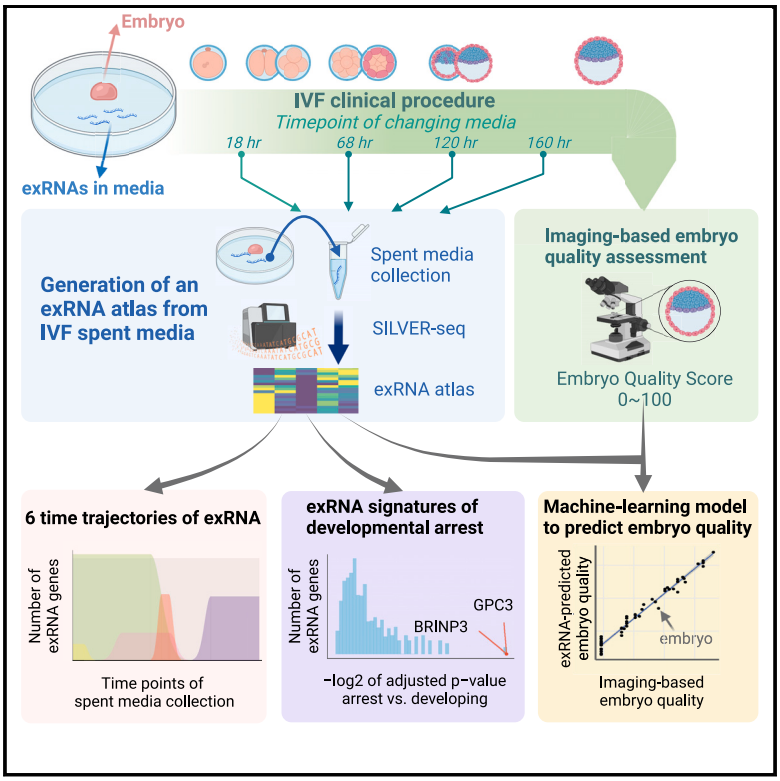
Peer reviewed

Cell Genomics

Short Article

A temporal extracellular transcriptome atlas of human pre-implantation development

Graphical abstract



Highlights

- A non-invasive method for sequencing exRNA from IVF embryo culture media
- A TETA of human pre-implantation development
- exRNA panels correlated with developmental arrest and embryo quality
- A machine-learning model to evaluate embryo quality based on exRNA levels

Authors

Qiuyang Wu, Zixu Zhou, Zhangming Yan, ..., Wei Zhang, H. Irene Su, Sheng Zhong

Correspondence

hisu@health.ucsd.edu (H.I.S.), szhong@ucsd.edu (S.Z.)

In brief

Non-invasive characterization of gene expression from IVF human embryos can provide a molecular-level understanding of the embryo. Wu et al. sequence the exRNAs in the embryo culture media and develop a machine-learning model to predict embryo quality based on exRNA profiles.

Wu et al., 2024, Cell Genomics *4*, 100464
January 10, 2024
https://doi.org/10.1016/j.xgen.2023.100464

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Short Article

A temporal extracellular transcriptome atlas of human pre-implantation development

Qiuyang Wu,[1,6] Zixu Zhou,[2,6] Zhangming Yan,[1] Megan Connel,[3] Gabriel Garzo,[3] Analisa Yeo,[3] Wei Zhang,[3] H. Irene Su,[3,4,5,*] and Sheng Zhong[1,2,7,*]

[1]Shu Chien-Gene Ley Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
[2]Genemo, Inc., San Diego, CA 92130, USA
[3]Reproductive Partners San Diego, La Jolla, CA 92037, USA
[4]Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Diego, La Jolla, CA 92093, USA
[5]Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA
[6]These authors contributed equally
[7]Lead contact
*Correspondence: hisu@health.ucsd.edu (H.I.S.), szhong@ucsd.edu (S.Z.)
https://doi.org/10.1016/j.xgen.2023.100464

SUMMARY

Non-invasively evaluating gene expression products in human pre-implantation embryos remains a significant challenge. Here, we develop a non-invasive method for comprehensive characterization of the extracellular RNAs (exRNAs) in a single droplet of spent media that was used to culture human *in vitro* fertilization embryos. We generate the temporal extracellular transcriptome atlas (TETA) of human pre-implantation development. TETA consists of 245 exRNA sequencing datasets for five developmental stages. These data reveal approximately 4,000 exRNAs at each stage. The exRNAs of the developmentally arrested embryos are enriched with the genes involved in negative regulation of the cell cycle, revealing an exRNA signature of developmental arrest. Furthermore, a machine-learning model can approximate the morphology-based rating of embryo quality based on the exRNA levels. These data reveal the widespread presence of coding gene-derived exRNAs at every stage of human pre-implantation development, and these exRNAs provide rich information on the physiology of the embryo.

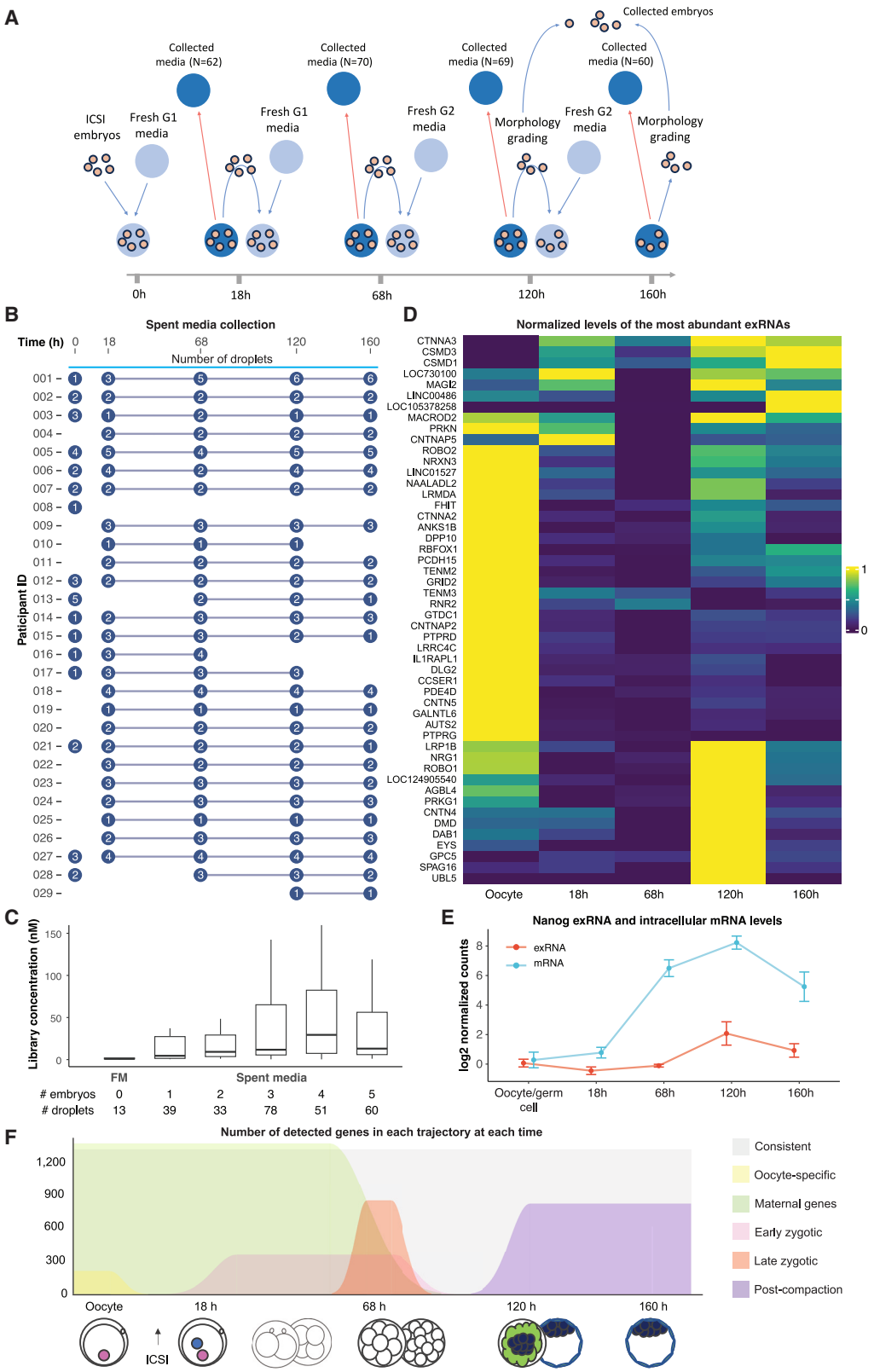
INTRODUCTION

The development of mammalian pre-implantation embryos is regulated by the coordinated expression of many genes at different developmental stages.[1,2] The maternally deposited RNA must be present and depleted at specific developmental stages, and the zygotic genes must also be activated and repressed in specific orders.[3–6] Human embryos express up to 12,000 genes at each major developmental stage and exhibit vast transcriptomic changes as the embryos develop,[7–10] suggesting human pre-implantation development also requires the coordinated expression of many genes. Gene expression data, however, are rarely used for evaluation of the proper development of human *in vitro* fertilized embryos, due to the invasive procedures required by transcriptomic characterization.

Morphological assessment is non-invasive and currently the primary tool for evaluating embryo quality. Recent developments leveraged time-lapse multiplanar imaging to assess the kinetics of the embryo's morphology.[11] Morphological and morphokinetic features reflect certain critical aspects of the embryo.[12] However, the total amount of information is limited due to the limited number of known physiology-associated and prognostic morphological features.[13,14] Considering the significant association of gene expression and embryonic development, it is conceivable that, if any subset of the RNA transcripts of the genes with known importance to regulating development can be non-invasively quantified, such molecular measurements may provide additional information about the developmental status of the human embryo.[13]

A subset of the intracellular RNA is released to the extracellular space to form exRNA.[15–17] We reasoned that a comprehensive analysis of the extracellular transcriptome may offer a way to non-invasively evaluate the subset of the RNA released from the embryos. To date, comprehensive characterization of the extracellular transcriptome has been carried out from body fluids, which revealed the exRNAs from thousands of coding and non-coding genes from each type of body fluid.[18,19] In comparison, only a small number of noncoding RNAs in the *in vitro* fertilization (IVF) media have been studied.[20] Without a systematic characterization, it remains unknown what RNA species are present in the IVF culture media at what developmental stage. Importantly, it remains unknown whether the RNA of any coding gene is present in the IVF media.

The technical challenges in a comprehensive characterization of the extracellular transcriptome are 2-fold. First, the volume of each droplet of media used for embryo culture is small (20–40 μL).[21] Second, the number of embryos cultured in each media



(legend on next page)

droplet is also small (typically 1–6 embryos), limiting the number of exRNA-releasing cells. Most exRNA sequencing methods are tested with body fluids, and they require far more fluid (500–1,000 μL) as the input than the volume of an IVF media droplet. To address these challenges, we apply the recently developed small input liquid volume extracellular RNA sequencing (SILVER-seq), a technology designed for unbiased sequencing of the extracellular transcriptome from a tiny volume (5–10 μL) of liquid biopsy,[22,23] to spend IVF media.

With SILVER-seq, we characterize the extracellular transcriptomes at five time points, approximately corresponding to the oocyte, zygote, 8-cell, morula, and blastocyst stages of embryonic development. These data reveal a rich extracellular transcriptome at each developmental stage, composed of exRNAs transcribed from approximately 4,000 genes, in which more than one-half are coding genes. The temporal changes of the exRNAs during development can be summarized by a small number of temporal patterns, which we term ''time trajectories.'' A machine-learning model using the measured exRNA levels can predict embryo quality based on morphology assessment, suggesting the possibility of using the extracellular transcriptome to assist in evaluating embryo quality.

Ensuring ethical practices in human embryo-related analyses is of utmost importance. In our study, ethical considerations were integral to our research approach. We made a deliberate decision not to undertake any experiments that could potentially impact the clinical outcomes of human embryos. It is essential to emphasize that our study adhered to the clinical policy and standard of care followed in the majority of human IVF clinics, where embryos are co-cultured whenever feasible.

The practice of embryo co-culture is deemed the standard of care in clinical IVF settings due to its demonstrated benefits in enhancing embryo cleavage, morphology, and viability compared with individual embryo culture.[24–26] Throughout our study, we strictly adhered to this standard of care without altering any clinical policy or practice. Consequently, the majority of our sequencing libraries were constructed from the spent media of co-cultured embryos, reflecting the clinical reality of human IVF procedures. By aligning our study with the existing clinical policy and standard of care, we upheld ethical considerations and ensured that our research did not interfere with the established practices aimed at optimizing the outcomes of human embryos in clinical IVF settings.

RESULTS

Generation of an exRNA sequencing atlas from IVF spent media

We collected the spent media used in the IVF for a total of 29 participants (spent media donors). The spent media were collected at each time when the culture media were changed. We call each collected spent media a droplet. A total of 295 droplets were collected. Among them, 34 droplets were used for culturing MII oocytes, and the other 261 droplets were used for embryo culturing, which include 62, 70, 69, and 60 droplets collected at 18 h, 68 h, 120 h, and 160 h post insemination (Figures 1A and 1B; Tables S1 and S2).

We asked whether there was any detectable RNA in the spent media at each collection time. To this end, we used SILVER-seq to convert media-contained exRNA, if any, to a sequencing library. Briefly, SILVER-seq takes the media droplet as input, removes any DNA by DNase, lyses any extracellular vehicle and digests any proteins, synthesizes complementary DNA (cDNA) from RNA, and amplifies the cDNA into a sequencing library.[22]

As controls, we prepared 13 droplets of media that did not culture any cell (fresh media controls). These control droplets were planted and harvested in parallel with those droplets used for embryo culture. We subjected every control droplet to SILVER-seq to construct a sequencing library. None of the fresh media droplets yielded sufficient DNA content in the constructed library for DNA sequencing, based on the same number of PCR cycles as that used for library construction from the spent media droplets (fresh media [FM] column, Figure 1C; Table S3).

We subjected every spent media droplet to SILVER-seq to construct a sequencing library. A total of 245 droplets (83.05%) yielded high-concentration sequencing libraries (Tables 1 and S2), whereas the remaining 50 droplets (16.95%) yielded medium-concentration libraries (low-quality libraries) (Table S2). These data suggest the presence of exRNA in the culture media of oocytes and pre-implantation embryos at every analyzed developmental stage.

We asked whether the number of embryos co-cultured in the same droplet correlates with the chance for this droplet to yield a high-quality exRNA-sequencing library. The fresh media (number of embryos = 0) yielded no qualified library (FM column, Figure 1C). As the number of embryos cultured in the droplet increased, the DNA yield of the constructed library increased ($p = 0.022$, one-way ANOVA) (Figure 1C). Thus, the more

Figure 1. Sample overview and exRNA characteristics in IVF spent media

(A) An illustration of the clinical procedure for human IVF and spent media collection. N, number of samples collected. ICSI, intra-cytoplasmic sperm injection.

(B) An illustration of the clinical procedure for human IVF and spent media collection. Spent media were collected at 18 h, 68 h, 120 h, and 160 h when embryologists changed the culture media.

(C) The distribution of the concentration of the sequencing library constructed from each droplet (y axis) vs. the number of embryos cultured in each droplet (columns). The control droplets (FM) are denoted as # embryos = 0 (first column). Spent media droplets are grouped by the number of embryos cultured in each droplet (# embryos). The number of collected droplets (# droplets) of each group is provided in the last row (# droplets), which is the number of data points used for plotting the boxplot.

(D) A heatmap of exRNA levels of the top 50 most abundant exRNAs (rows). Color intensity shows the normalized levels across time points (columns) for each exRNA (row).

(E) The exRNA (red) and intracellular mRNA (blue) levels of the Nanog gene, as quantified by log2(normalized counts) (y axis) at each stage (x axis), with error bar denoted the standard error.

(F) The representative time trajectories. The six time trajectories are noted with different colors, with the number of detected genes of each trajectory at each time denoted on the y axis. See also Figures S1 and S2 and Tables S1, S2, S3, and S4.

Table 1. Overview of the exRNA sequencing libraries

Developmental Stage	Collection time post insemination	No. of droplets	No. of high-quality libraries	Sequencing library size
Oocyte	N/A	34	28	5,249,914
Zygote	18 h	62	57	5,904,028
2-cell to 16-cell	68 h	70	66	4,890,635
Morula to blastocyst	120 h	69	51	6,179,027
Blastocyst	160 h	60	43	5,534,728

Developmental stage: the approximate developmental stages of the cultured oocytes or embryos. Collection time post insemination: the amount of time passed since the oocytes were screened for insemination. The duration for which any media was used for culturing equals the denoted collection time subtracting the previous collection time. For example, the media collected at 68 h was used for 68 – 18 = 50 h of culturing. Number of droplets: the number of media droplets collected. Number of high-quality libraries: the number of droplets that resulted in high-quality sequencing libraries. Sequencing library size: the average number of sequenced read pairs per library. See also Table S2.

embryos cultured in a droplet, the higher the concentration of the constructed sequencing library.

We sequenced these libraries to yield 5–6 million sequencing read pairs per library. We provide the 245 high-quality libraries as a resource called a temporal extracellular transcriptome atlas of human pre-implantation development (TETA).

Correspondence of intracellular and extracellular RNAs of human embryos

The majority of the sequencing reads were uniquely mapped to the human genome (hg38), with a median per library unique mapping rate of greater than 64% at every collection time, confirming that the sequenced RNAs originate from the human genome (Table S4). These unique mapping rates are in the same range as those of single-cell RNA sequencing (scRNA-seq) datasets of human pre-implantation embryos,[8,9] indicating the satisfactory quality of this exRNA-sequencing dataset (Table S4). We quantified the exRNA levels of each human gene (hg38) at each collection time based on the normalized counts of the mapped read pairs[27] (Figure 1D, red curves in Figures 1E and S1A–S1D). We detected approximately 4,000 exRNA genes at each collection time, totaling 6,730 genes (detected genes).

The collection time 18 h, 68 h, 120 h, and 160 h approximately corresponds with the developmental stages of zygote, 2- to 16-cell, morula to blastocyst, and blastocyst (Figure 1B; Table 1). We asked whether the detected exRNAs are a subset of the genes expressed at the corresponding developmental stages of the human pre-implantation embryos. To this end, we downloaded two scRNA-seq datasets of human pre-implantation embryos.[8,9] The two datasets together included 116 human oocytes and embryos, assayed at oocyte, pronucleus, zygote, 2-cell, 4-cell, 8-cell, morula, and late blastocyst developmental stages (Table 2).

To establish a reference, we asked how consistent the detected genes between these two intracellular RNA sequencing, i.e., the two scRNA-seq, datasets are. To this end, we detected the expressed genes from each dataset and compared the expressed genes at each developmental stage between the two scRNA-seq datasets. At each developmental stage, we detected approximately 9,000–15,000 expressed genes from each dataset. The percentage of overlap of the detected genes between the two scRNA-seq datasets ranged from 78% (the morula stage) to 98% (the zygote stage). These overlap percentages represent the variability between the two studies and provide a reference for our subsequent comparison of the extracellular (SILVER-seq) and intracellular RNA sequencing (scRNA-seq) datasets.

Next, we compared the detected exRNA genes (SILVER-seq) with the expressed genes detected by intracellular RNA sequencing (scRNA-seq) at each developmental stage. To this end, we merged the two scRNA-seq datasets and called the expressed genes from the merged dataset at each developmental stage from oocyte to morula (“No. of detected genes” column under “Intracellular RNA Sequencing,” Table 2a) (see below for comparison with blastocysts). Considering the differences in lengths and RNA integrity between exRNA and intracellular RNA, as well as the differences in the experimental steps between SILVER-seq and scRNA-seq, we expect the percent overlap of the genes detected from the two scRNA-seq datasets to be greater than the percent overlap of the genes detected from SILVER-seq (exRNA) and scRNA-seq (intracellular RNA). It turns out that 66%–83% of the detected exRNA genes in each stage were detected as expressed genes by scRNA-seq (Overlap percentage, Table 2), suggesting most of the detected exRNA genes are a subset of the detected intracellular RNA genes.

These two scRNA-seq datasets did not provide sufficient data at the blastocyst stage. One did not include blastocysts and the other assayed sufficient trophectoderm (TE) cells but only a small number of cells from the inner cell mass, including 7 primitive endoderm (PE) cells and 5 epiblast (EPI) cells. Thus, we downloaded another scRNA-seq dataset from Blakeley et al.[28] that contained 36 EPI, 26 PE, and 23 TE cells to compare with the detected exRNA genes. Of the exRNA genes detected in the 160-h droplets, 66.7% overlapped with the scRNA-seq detected genes at the blastocyst stage (Table 2b). This percentage is within the range of overlaps at the earlier developmental stages.

Taken together, the range of overlaps between SILVER-seq and scRNA-seq (66%–83%) is not completely separated from the range of the overlaps between two scRNA-seq studies (78%–98%). Thus, most SILVER-seq-detected exRNA genes overlap with scRNA-seq-detected genes at the corresponding stage, confirming that most exRNAs in the spent media correspond with the intracellular RNAs in human embryos.

We analyzed the sample-to-sample variability within and between time points using Uniform Manifold Approximation and Projection analysis on the scRNA-seq datasets[8,9] and our

Table 2. Comparison of the detected genes by extracellular and intracellular RNA-sequencing

exRNA-sequencing			Overlap		Intracellular RNA-sequencing		
Stage	No. of droplets	No. of detected genes	No. of genes	Percentage	Developmental stage	No. of single cells	No. of detected genes
(a) Comparison with pronucleus to morula stages scRNA-seq data by Xue et al. and Yan et al.							
Oocyte	28	4,877	3,656	74.96%	oocyte	6	11,755
0–18 h embryo	57	5,089	4,233	83.18%	pronucleus	3	14,193
					zygote	5	14,512
18–68 h embryo	66	6,801	5,581	82.06%	2-cell	9	14,589
					4-cell	16	11,740
					8-cell	31	11,408
68–120 h embryo	51	2,233	1,506	67.44%	morula	19	11,664
exRNA-sequencing			Overlap		Intracellular RNA-sequencing		
Stage	No. of droplets	No. of detected genes	No. of genes	Percentage	Cell type	No. of single cells	No. of detected genes
(b) Comparison with blastocyst stage scRNA-seq data by Blakeley et al.							
120–160 h embryo	43	3,616	2,414	66.76%	epiblast	36	6,455
					primitive endoderm	26	7,369
					trophectoderm	23	7,692

"exRNA-sequencing" columns include the developmental stages during which the spent media was used for culturing (stage), the number of media droplets collected (number of droplets), and the number of exRNA genes detected (number of detected genes). "Intracellular RNA-sequencing" columns include the developmental stage (developmental stage), the number of single cells (number of single cells) of the collected embryos, and the number of genes detected by single-cell RNA-sequencing (number of detected genes). Number of single cells and number of detected genes are based on the union of two scRNA-seq datasets from Xue et al.[8] and Yan et al.[9] in (a), and one scRNA-seq dataset from Blakeley et al.[28] in (b). The "Overlap" columns include the number of the exRNA-sequencing detected genes that are also detected by scRNA-sequencing (number of genes), and the percentage of these overlapping genes among the exRNA-sequencing detected genes (percentage).

exRNA sequencing (SILVER-seq) data (Figure S1E). The early-stage (18-h and 68-h) samples and the late-stage (120-h and 160-h) samples are separately clustered in both SILVER-seq and scRNA-seq datasets, indicating a larger sample-to-sample variation between an early-stage and a late-stage sample than two samples of the same stage. Thus, time is a variable in explaining the sample-to-sample to variability. Second, we recognize that the embryos do not always exhibit synchronized developmental progress. For example, the 18-h media contain both successfully and unsuccessfully fertilized oocytes. Additionally, both 120-h and 160-h media contain an array of embryos with different cell numbers and morphology. Thus, we anticipated to see a limited ability of time points to explain the sample-to-sample variability. Consistent with this expectation, the exRNA profiles did not exhibit clear separations between the time points within the early stage (18 h, 68 h) or within the late stage (120 h, 160 h).

Characteristic time trajectories of exRNA during human pre-implantation development

We defined the time trajectory of a gene as the detection status of this gene at each collection time and identified six trajectories in the TETA dataset. The trajectory with the largest number of genes is "present from oocyte to 68 h and absent from 120 h to 160 h." We called the genes in this trajectory the "maternal" genes (1,293 genes, green curve, Figure 1F). Based on the stages when the exRNA genes were detected, we called the next four trajectories as "oocyte-specific" (present at oocyte and absent from 18 h to 160 h), "early zygotic" (present from 18 h to 68 h), "late zygotic" (present at 68 h), and "post compaction" (present at 120 h or 168 h) (yellow, pink, red, purple curves, Figure 1F). In addition, a group of genes are "present at every stage," which we call the "consistent" trajectory (1,265 genes, gray curve, Figure 1F). For each time trajectory, we identified the subset of exRNA genes such that their corresponding intracellular RNAs exhibited similar temporal patterns as the exRNAs (Figures S2A and S2B). This subset of overlapping genes is larger than the expected size of overlap for each trajectory (largest p value = 5.04e−10) (Figure S2A), suggesting an enrichment of the overlapping genes. Gene Ontology analysis revealed that the overlapping genes in the oocyte-specific and the maternal genes trajectories are enriched with transcriptional coregulators and histone modifiers, whereas the overlapping genes in the late zygotic and post compaction trajectories are enriched in components of the membrane, transport vesicle, and glycoprotein and membrane lipid metabolic processes (Figure S2B).

exRNA signatures of the developmentally arrested embryos

We tested if the exRNA in the culture media correlates with whether the embryos are developmentally arrested. An embryologist marked every embryo at 120 h and 160 h as either developmentally arrested or not. A total of 117 media droplets were collected at 120 h and 160 h. Among them, 86 droplets each contained at least one developing embryo (developing media), and the other 31 droplets in which every embryo was arrested (arrest media) (Arrest column, Table S2).

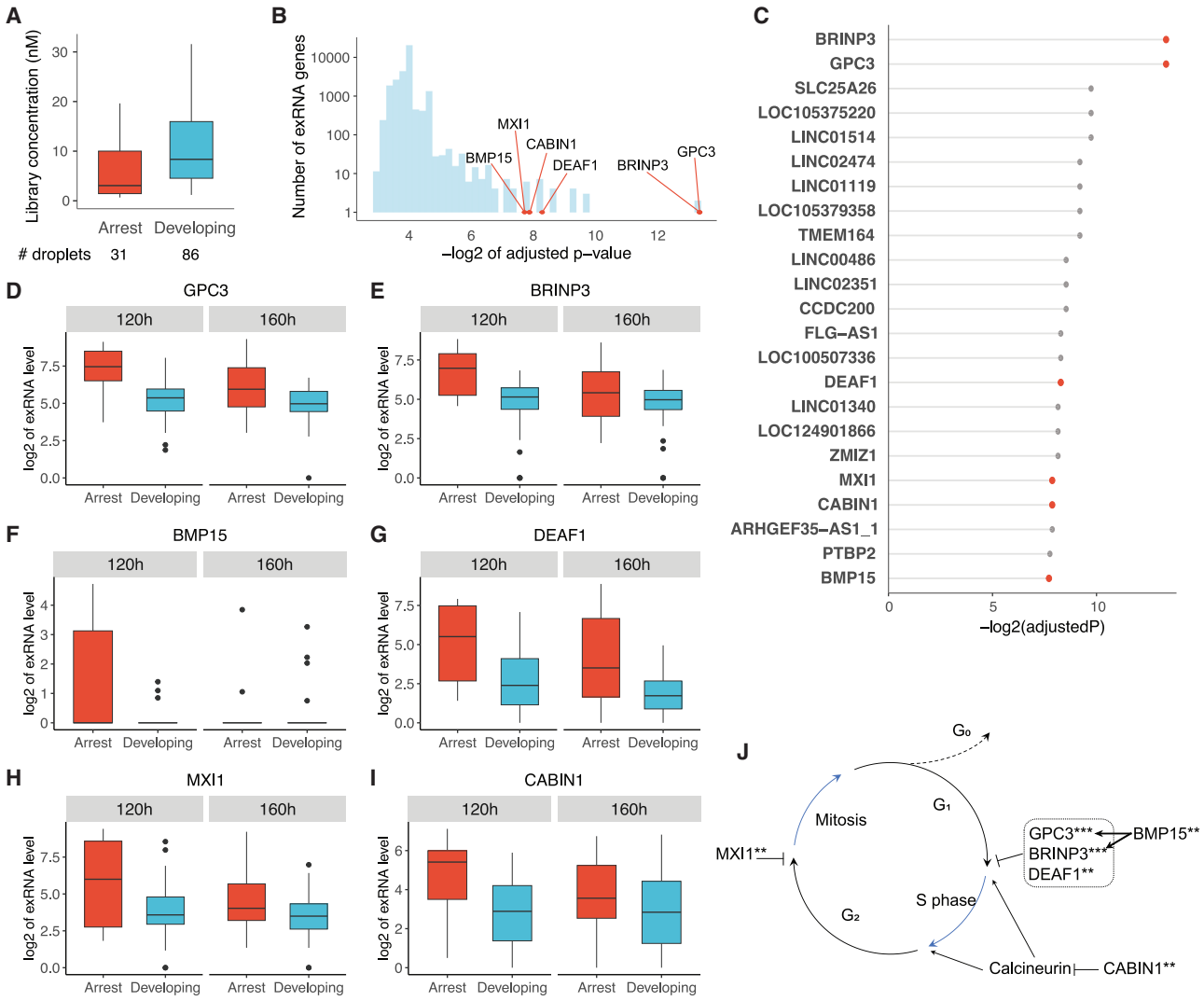


Figure 2. exRNA features of embryo arrest

(A) The distribution of the concentration of the sequencing library constructed from each droplet (y axis) of the arrest media (arrest column) and developing media (developing column). The number of droplets (# droplets) of each group is provided in the last row (# droplets), which is the number of data points used for plotting the boxplot. Outliers are not shown in these boxplots.

(B) The distribution of the adjusted p values (the multiple testing adjusted p value is estimated by Storey-Tibshirani procedure, which controls the positive false discovery rate) of all detected exRNA genes. A large –log2 adjusted p value (x axis) indicates a strong difference in the exRNA levels between the arrest and developing media.

(C) The genes (rows) with the most significant difference between the arrest and developing media (adjusted $p < 0.005$). The x axis, –log2 adjusted p value.

(D and E) The exRNA levels of GPC3 and BRINP3 in arrest media and developing media, presented by boxplots.

(F) The exRNA levels of BMP15, presented by boxplot.

(G–I) The exRNA levels of DEAF1, MXL1, and CANBIN1, presented by boxplots.

(J) The roles of ACE genes in inhibiting cell cycle progression. ***Adjusted p value < 0.0001, **adjusted p value < 0.005.

We asked whether the library concentrations are the same between the arrest and developing media. The average library concentration of the droplets of the arrest media was smaller than that of the developing media ($p = 0.0069$, two-way ANOVA controlling for embryo number) (Figure 2A). These data suggest the total amount of exRNA negatively correlates with developmental arrest.

Next, we asked whether the levels of any specific exRNAs are associated with developmental arrest. Our null hypothesis is that there is no association between the exRNA level of a gene and the arrest media. A total of 23 exRNA genes were associated with arrest media (Multiple hypothesis-testing adjusted $p < 0.005$, logistic regression controlling for embryo number) (Figure 2B). All these identified exRNAs exhibited higher levels in the arrest media than in the developing media (Figure 2C). Hereafter we call these genes the arrest-correlated exRNA (ACE) genes.

We examined the known functions of every identified ACE gene. Ten of the 23 ACE genes are coding genes, and the other 13 are noncoding genes. The top two most significant ACE genes are Glypican 3 (GPC3) and BMP/Retinoic Acid Inducible Neural Specific 3 (BRINP3) (adjusted $p < 0.0001$, logistic regression) (Figures 2B, 2D, and 2E). Both GPC3 and BRINP3 negatively regulate cell-cycle progression at the G1/S phase transition, and both respond to BMP signaling.[29,30] Consistent with the top 2 ACE genes both responding to BMP, BMP15, an important member of the BMP family, is also one of the 10 coding ACE genes (adjusted $p < 0.005$) (Figure 2F). The other seven coding ACE genes are SLC25A26, TMEM164, DEAF1, ZMIZ1, MXI1, CABIN1, and PTBP2 (adjusted $p < 0.005$), in which DEAF1,[31] MXI1,[32] and CABIN1[33] are also negative regulators of cell cycle (Figures 2G–2I). Thus, the ACE genes exhibit two functional features, namely (1) being part of or responding to BMP signaling and (2) being negative regulators of the cell cycle (Figure 2J). Taken together, developmentally arrested embryos exhibit fewer total amounts of exRNA, but higher levels of exRNA of specific genes.

A predictive model of embryo quality based on exRNA profiles

We asked if any exRNA in the culture media correlates with embryo quality. An embryologist rated every embryo at 120 h and 160 h by morphology. Each blastocyst stage embryo received a rating of good, fair, or poor. Each morula stage embryo received a label of complete or incomplete, corresponding with complete or incomplete compaction. We provided a numeric value to each morphological rating (Table S5) and used the mean of these values to represent the average quality of the co-cultured embryos in a droplet. Hereafter, we call this mean value as the average quality score of this droplet. When a single embryo is cultured in a droplet, the quality score of this individually cultured embryo is, by definition, the average quality score of this droplet.

We set off to identify any exRNA that is predictive of the average quality score. First, we independently tested every gene for its correlation with the average quality score, which identified 0 genes (adjusted p cutoff = 0.01). It suggests that, when the genes are used independently, no gene by itself is strongly predictive of the average embryo quality in a droplet.

Next, we asked if any set of genes is collectively predictive of the average quality score. To this end, we trained a random forest (RF) regression model using the exRNA levels of all the detected genes as inputs and the average quality score as the outcome. The predicted average quality scores by RF regression strongly correlated with the observed average quality score of every droplet (Pearson correlation $r = 0.989$, $p < 2.2e{-}16$) (Figure 3A). This result suggests that average embryo quality is associated with the collective behavior of a subset of genes, namely those genes assigned with large weights in RF regression.

Hereafter, we call the top-ranked genes with the largest weights given by the RF regression as the embryo quality predictive (EQUIP) genes. We asked if the EQUIP genes share any functional features. The top 100 EQUIP genes are enriched with two Gene Ontology (GO) terms (adjusted p value cutoff = 0.1), which are Regulation of GTPase Activity (adjusted $p = 0.083$) and Spindle (adjusted $p = 0.098$) (Figure 3B). Both enriched GO terms are related to the regulation of asymmetric cell division. For example, an EQUIP gene associated with both GO terms is Microtubule-Associated Protein RP/EB Family Member 2 (MAPRE2), which is required for asymmetric cell division due to its GTPase activity and role in modulating spindle symmetry.[34] In summary, the exRNA profiles of a group of genes in a droplet correlate with the average embryo quality in this droplet.

To provide a simple and intuitive quantification of the extent to which exRNAs can predict embryo quality, we simplified the task of predicting embryo quality to a classification problem. To this end, we dichotomized the average quality score into high ($\geq$25) and low (<25) to train a RF classification model and evaluated the model's ability to classify these binary scores. To avoid overfitting, we used the average area under the curve (AUC) of cross-validations to indicate the model performance. This average AUC offers a simple quantity that reflects how predictive the exRNAs can be to the average quality of the embryos co-cultured in a droplet.

We analyzed the impact of the number of input genes on the performance of the RF classification model. The performance of an effective classification model is expected to first increase and then decrease as the number of input genes increases. Such behavior shows the greater signal (benefit) than noise (harm) that each additional input gene initially brings to the model until the optimal number of input genes and the greater harm than benefit thereafter. We plotted the average AUC of 10-fold cross-validations as a function of the number of input genes (Figure 3C). The average AUC increased from 0.71 to 0.88 as the input genes increased from 5 to 42 and decreased as the input genes further increased (Figure 3C). With 42 input genes, the classification model achieved an average AUC of 0.89 in cross-validations (Figure 3D). These data indicate that the model is effective in classifying the average quality of co-cultured embryos.

We conducted an evaluation to assess the predictive capability of our RF classification model in determining the quality of individually cultured embryos. According to our dichotomy, embryos were classified as high quality if their quality score exceeded 25 and low quality otherwise (Table S3). Our dataset contained 39 droplets used for culturing single embryos, in which 26 were collected at 120 h or 160 h. These 26 individually cultured embryos have been rated by an embryologist for morphology. Among these 26 individually cultured embryos, 5 were rated as high quality and the rest 21 were rated as low quality.

To evaluate the performance of our RF classification model, we employed leave-one-out cross-validation (LOOCV) on each individually cultured embryo. In each round of LOOCV, one single-embryo culture droplet was reserved for testing purposes, while the other single-embryo culture droplets, as well as all the co-culture droplets, were utilized for training. The LOOCV analysis yielded an impressive AUC of 0.94 (Figure 3E), indicating that our model, based on exRNA profiles, effectively predicts the quality of individually cultured embryos. To account for the uneven distribution of high- and low-quality single embryos, we calculated the harmonic mean of precision and recall, resulting in an average F1 score of 0.92 for the

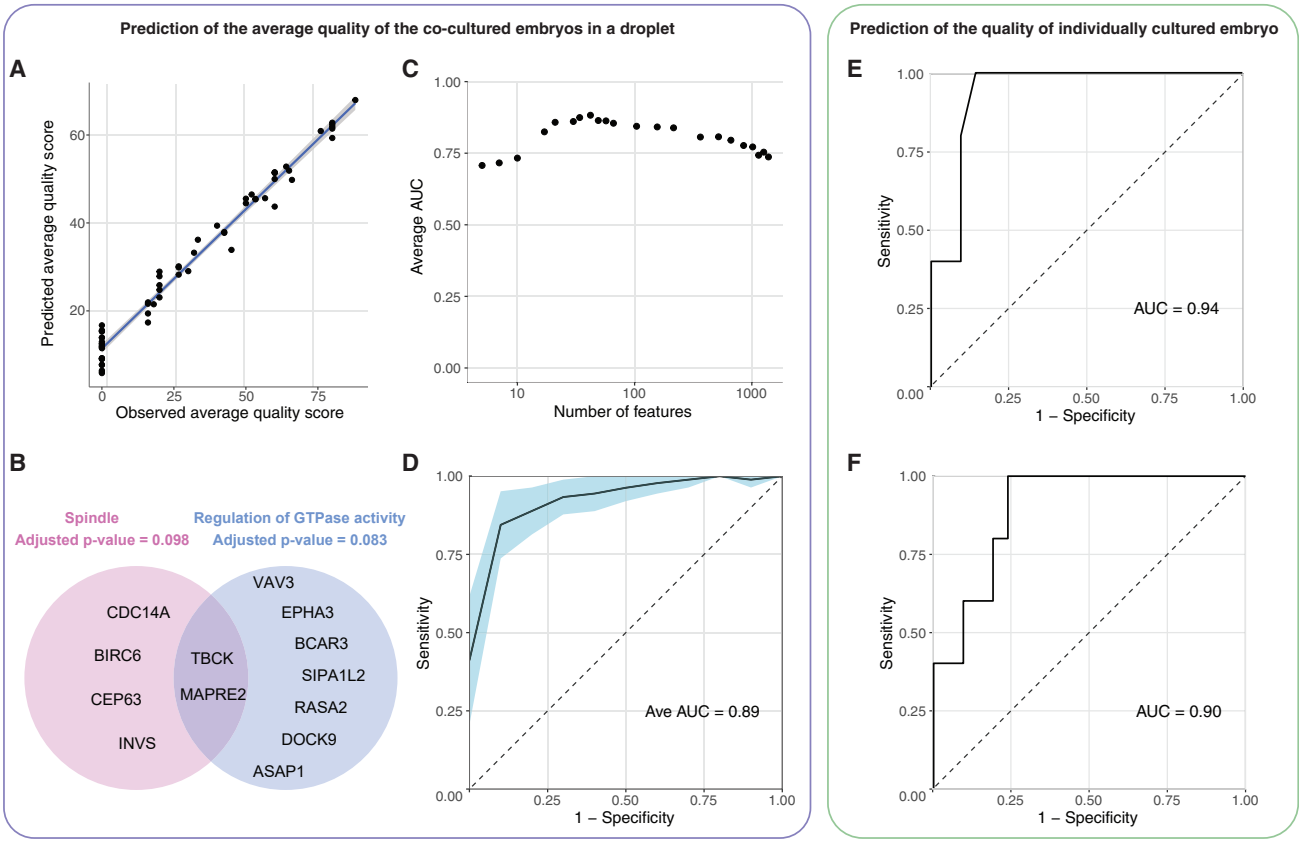


Figure 3. A machine-learning model of embryo quality prediction

(A) The observed quality score (x axis) of every droplet (dot) vs. the predicted quality score by the RF model (y axis).

(B) A Venn diagram of the enriched GO terms (called with Benjamini-Hochberg p value < 0.1) and the EQUIP associated with each GO term.

(C) The average AUC of 10-fold cross-validations (y axis) with respect to the number of input genes used by the model (x axis).

(D) The average of 10 cross-validation receiver operating characteristic (ROC) curves with 42 input genes. The blue shade denoted the 95% confidence interval.

(E) The ROC curve based on LOOCV for the individually cultured embryos.

(F) The ROC curve based on training with co-cultured embryos and testing on individually cultured embryos. See also Table S5.

LOOCV. This suggests that the classification model's strong performance is not due to imbalanced numbers of high- and low-quality embryos.

We hypothesized that the effectiveness of our classification model is partly attributed to the model's ability to extract informative features from the co-culture droplets, which are indicative of the quality of individually cultured embryos. To test this hypothesis, we split the data into training sets consisting of co-culture droplets and testing sets consisting of individual-culture droplets. In contrast with LOOCV, this analysis excluded any single-embryo culture data from the training set entirely. The results yielded an AUC of 0.90 (average F1 score = 0.88) (Figure 3F), indicating that the RF classification model trained on co-cultured embryos successfully predicts the qualities of individually cultured embryos. These performance scores are slightly lower than those obtained from LOOCV, which is expected since LOOCV incorporates both co-culture droplets and all-except-one individual-culture droplets. In summary, these data demonstrate that a machine-learning model utilizing exRNA profiles can accurately predict morphology-based embryo quality.

DISCUSSION

As demonstrated by RNA-seq and scRNA-seq analysis of the blastocyst TE, the mRNA expression reflects the embryo's physiology and competence.[35,37] It remains largely unknown what coding genes' transcripts are present in the extracellular space of cultured human embryos. This study characterizes several thousand exRNAs. More than one-half of the detected exRNAs were derived from coding genes, revealing a widespread presence of many coding genes' RNA transcripts in IVF culture media. This dataset includes the extracellular transcriptomes of human oocytes and zygotes, which have not been reported to date. This dataset should provide a rich resource for future studies of human pre-implantation embryonic development.

The time of human zygotic genome activation remains debatable.[4,7] More generally, many molecular events marking the human maternal-zygotic transition and their precise timing remain to be characterized. This work reveals a transition from oocyte to zygote in their extracellular exRNAs (Figure 1F). Approximately 300 oocyte-specific exRNAs become undetectable in zygotes (yellow curve), whereas approximately 600 early zygotic

exRNAs become detectable from the 1-cell (18-h) stage (pink curve, Figure 1F). These data are more compatible with the notion that zygotic genome activation is initiated at the 1-cell stage in humans.[4,7]

Relatively little is known about the biogenesis processes of embryo-derived exRNAs. Human embryos at the blastocyst stage present a sufficient amount of extracellular DNA for whole-genome sequencing.[38] Such extracellular DNA is likely deposited by the apoptotic blastomeres during the late stages of pre-implantation development.[39] Apoptotic cells can also release RNA,[40] making apoptosis a conceivable contributing factor to the exRNAs at the late stages of pre-implantation development. Although the early-stage embryos do not contain apoptotic cells, at least one of the polar bodies is often eliminated through an apoptosis-like process.[41] Human polar bodies contain mRNAs,[42] making the disintegrating polar body(ies) a plausible source of the exRNA at the early stages of pre-implantation development. RNA-containing extracellular vesicles (EVs) are detected from day 3 and day 5 post-fertilization from human IVF embryos.[43] Thus, EV-containing RNA may account for a fraction of the extracellular transcriptome, at least in the late stages of pre-implantation development. Taken together, apoptosis, polar body disintegration and EVs are among our perceived plausible means for human embryos to release exRNA. Further studies are warranted to test these hypothetical pathways.

This study identifies a correlation between the exRNAs of ACE genes and embryonic arrest. The most significant ACE genes, GPC3 and BRINP3, both respond to BMP signaling and suppress cell-cycle progression, highlighting BMP as a plausible signal for inducing embryonic arrest. In line with this idea, BMP15 is also a highly ranked ACE gene. Consistent with the prior knowledge that BMP15 is maternally expressed and promotes oocyte developmental competence,[44] BMP15's exRNA is detected from the oocyte and the zygote stages but becomes undetectable in most spent media at the late stages (120 h–160 h) (Figure S1F). The small number of BMP15 exRNA-presenting droplets at the late stages are mostly arrest media at 120 h (Figure 2F), explaining the correlation between BMP15 exRNA and arrest media. These data support a hypothetical model where the maternal BMP15 transcripts are depleted as the embryos properly develop, whereas those embryos that did not progress to compaction kept a high level of BMP15.

DEAF1 Transcription Factor (DEAF1) exRNA is also correlated with embryonic arrest at the morula stage. DEAF1 interacts with GSK3A and GSK3B.[45,46] GSK3 inhibitors are used in the 2i and 5i cocktails to push human pluripotent stem cells to the naive state.[47–49] GSK3 inhibitors increase DEAF1's transcriptional activity on target promoters.[45] These data suggest the possibility that the upregulation of DEAF1 mimics some of the effects of 2i or 5i cocktails on stalling the developmental process prior to the first cellular differentiation that specifies the inner cell mass and the TE.

The aforementioned examples highlight the rich information in the coding genes' exRNAs. They help to connect the noninvasive measurements with the vast prior knowledge about the coding genes' roles in orchestrating embryonic development. Further supporting this connection, a simple machine-learning model based on exRNA levels achieved decent accuracy in predicting embryo quality. Additionally, this machine-learning model prioritized the genes involved in asymmetric cell division as the most predictive features of embryo quality, coinciding with the prior knowledge that the specification of the first two cell lineages is an essential task of the embryos at the morula-to-blastocyst transition. Taken together, thousands of exRNAs originate from the coding, and the noncoding genome of the human embryo are quantifiable, which offers a unique opportunity for noninvasively studying human pre-implantation development.

Limitations of this study

During the conceptualization stage of this study, all authors unanimously agreed upon a principle that prioritized the well-being of the participants and aimed to avoid any negative impact on the standard of care. As a result, we did not request patients to donate embryos specifically for research purposes. Additionally, in accordance with clinical policy, embryos were co-cultured whenever feasible. Consequently, most sequencing libraries in the TETA resource are based on the pooled exRNA from embryos co-cultured within the same droplet. This nature of our data presents limitations that preclude us from addressing certain intriguing questions. For example, we cannot explore whether our technology can predict the competence of individual embryos,[35] as most transferred embryos are chosen from co-culture droplets. Similarly, our technology cannot be employed to determine the genotype and ploidy of each embryo.[36] Our limited number of individually cultured embryos is insufficient to analyze an association between any genetic variation and exRNA levels. These limitations stem from our commitment to maintaining the integrity of the clinical practices without altering the established clinic policy.

Because our study did not mandate the donation of embryos for research purposes and it did not alter any procedures associated with the standard of care, enough patients willingly consented to participate in the study. This sufficient number of participants was essential for generating this TETA resource and demonstrating the predictive capability of exRNA profiles in determining morphology-based embryo quality in a clinical setting.

Including co-culture droplets helped to derive a reference set of exRNAs at each developmental stage. The co-culture droplets had higher concentrations in the constructed sequencing libraries than the individual culture droplets. Thus, including co-culture droplets allowed for greater sensitivity in identifying exRNAs at each developmental stage. The identified exRNAs provide a reference to future studies using either co-culture or individual culture. As demonstrated in the tests with individual culture droplets, co-culture droplets are not required in future studies for evaluating embryo quality.

The correspondence between the media change time and the embryo's developmental stage is not perfect. Previous scRNA-seq studies were able to analyze the embryos at the exact developmental stages, for example, at the 8-cell stage. However, in this study, an embryo can develop faster or slower than the average speed, and a media droplet contains up to five embryos that do not necessarily synchronize in their developmental progress. Thus, some detected exRNAs at a time point may come from an embryo at earlier or later developmental stages than expected. More generally, any difference in the exRNA (this study) and intracellular RNA (previous data) profiles may arise from the

difference between the actual and the expected developmental progress of the embryos.

This work provided only a small number of speculations on how the ACE genes may work together or with other genes to cause developmental arrest. While a topic of interest, it is outside the scope of the current work and represents a separate research topic that requires additional investigation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human subjects
- METHOD DETAILS
  - Human In vitro fertilization (IVF) and collection of spent media
  - Constructing SILVER-seq libraries
  - Processing SILVER-seq libraries
  - Detection of expressed genes from scRNA-seq data
  - Identification of exRNA signatures of the developmental arrest
  - The random forest regression model
  - The random forest classification model
  - Singularly cultured embryo quality evaluation
  - Cluster analysis
- QUANTIFICATION AND STATISTICAL METHODS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100464.

ACKNOWLEDGMENTS

This work is supported by NIH grants R01HD107206, R01GM138852, and DP1DK126138 and a Kruger research grant. The graphical abstract is created with BioRender.com.

AUTHOR CONTRIBUTIONS

S.Z. and H.I.S. conceived the project; M.C., G.G., A.Y., and W.Z. collected the IVF samples; Z.Z. performed SILVER-seq experiments; Q.W., S.Z., H.I.S., Z.Z., and Z.Y. analyzed and interpreted the data; S.Z., Q.W., and Z.Y. wrote the paper with input from A.Y., W.Z., and H.I.S.

DECLARATION OF INTERESTS

S.Z. is a founder and board member of Genemo, Inc.

Received: August 9, 2023
Revised: October 9, 2023
Accepted: November 19, 2023
Published: January 10, 2024

REFERENCES

1. Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic Genome Activation in Vertebrates. Dev. Cell *42*, 316–332.
2. Canovas, S., Ivanova, E., Romar, R., García-Martínez, S., Soriano-Úbeda, C., García-Vázquez, F.A., Saadeh, H., Andrews, S., Kelsey, G., and Coy, P. (2017). DNA methylation and gene expression changes derived from assisted reproductive technologies can be decreased by reproductive fluids. Elife *6*, e23670.
3. Sha, Q.Q., Zheng, W., Wu, Y.W., Li, S., Guo, L., Zhang, S., Lin, G., Ou, X.H., and Fan, H.Y. (2020). Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans. Nat. Commun. *11*, 4917.
4. Perry, A.C.F., Asami, M., Lam, B.Y.H., and Yeo, G.S.H. (2023). The initiation of mammalian embryonic transcription: to begin at the beginning. Trends Cell Biol. *33*, 365–373.
5. Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome Res. *20*, 804–815.
6. Zhang, C., Wang, M., Li, Y., and Zhang, Y. (2022). Profiling and functional characterization of maternal mRNA translation during mouse maternal-to-zygotic transition. Sci. Adv. *8*, eabj3967.
7. Asami, M., Lam, B.Y.H., Ma, M.K., Rainbow, K., Braun, S., VerMilyea, M.D., Yeo, G.S.H., and Perry, A.C.F. (2022). Human embryonic genome activation initiates at the one-cell stage. Cell Stem Cell *29*, 209–216.e4.
8. Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature *500*, 593–597.
9. Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.
10. Leng, L., Sun, J., Huang, J., Gong, F., Yang, L., Zhang, S., Yuan, X., Fang, F., Xu, X., Luo, Y., et al. (2019). Single-Cell Transcriptome Analysis of Uniparental Embryos Reveals Parent-of-Origin Effects on Human Preimplantation Development. Cell Stem Cell *25*, 697–712.e6.
11. Bamford, T., Barrie, A., Montgomery, S., Dhillon-Smith, R., Campbell, A., Easter, C., and Coomarasamy, A. (2022). Morphological and morphokinetic associations with aneuploidy: a systematic review and meta-analysis. Hum. Reprod. Update *28*, 656–686.
12. Rocha, J.C., Passalia, F., Matos, F.D., Maserati, M.P., Jr., Alves, M.F., Almeida, T.G.d., Cardoso, B.L., Basso, A.C., and Nogueira, M.F.G. (2016). Methods for assessing the quality of mammalian embryos: How far we are from the gold standard? JBRA Assist. Reprod. *20*, 150–158.
13. Gardner, D.K., and Balaban, B. (2016). Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still important? Mol. Hum. Reprod. *22*, 704–718.
14. Armstrong, S., Bhide, P., Jordan, V., Pacey, A., Farquhar, C., and Farquhar, C. (2018). Time-lapse systems for embryo incubation and assessment in assisted reproduction. Cochrane Database Syst. Rev. *5*, CD011320.
15. Patton, J.G., Franklin, J.L., Weaver, A.M., Vickers, K., Zhang, B., Coffey, R.J., Ansel, K.M., Blelloch, R., Goga, A., Huang, B., et al. (2015). Biogenesis, delivery, and function of extracellular RNA. J. Extracell. Vesicles *4*, 27494.
16. Gruner, H.N., and McManus, M.T. (2021). Examining the evidence for extracellular RNA function in mammals. Nat. Rev. Genet. *22*, 448–458.
17. Hawke, D.C., Watson, A.J., and Betts, D.H. (2021). Extracellular vesicles, microRNA and the preimplantation embryo: non-invasive clues of embryo well-being. Reprod. Biomed. Online *42*, 39–54.
18. Murillo, O.D., Thistlethwaite, W., Rozowsky, J., Subramanian, S.L., Lucero, R., Shah, N., Jackson, A.R., Srinivasan, S., Chung, A., Laurent, C.D., et al. (2019). exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. Cell *177*, 463–477.e15.

19. Srinivasan, S., Yeri, A., Cheah, P.S., Chung, A., Danielson, K., De Hoff, P., Filant, J., Laurent, C.D., Laurent, L.D., Magee, R., et al. (2019). Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. Cell *177*, 446–462.e16.

20. Kamijo, S., Hamatani, T., Sasaki, H., Suzuki, H., Abe, A., Inoue, O., Iwai, M., Ogawa, S., Odawara, K., Tanaka, K., et al. (2022). MicroRNAs secreted by human preimplantation embryos and IVF outcome. Reprod. Biol. Endocrinol. *20*, 130.

21. Sigalos, G.A., Michalopoulos, Y., Kastoras, A.G., Triantafyllidou, O., and Vlahos, N.F. (2018). Low versus high volume of culture medium during embryo transfer: a randomized clinical trial. J. Assist. Reprod. Genet. *35*, 693–699.

22. Zhou, Z., Wu, Q., Yan, Z., Zheng, H., Chen, C.J., Liu, Y., Qi, Z., Calandrelli, R., Chen, Z., Chien, S., et al. (2019). Extracellular RNA in a single droplet of human serum reflects physiologic and disease states. Proc. Natl. Acad. Sci. USA *116*, 19200–19208.

23. Yan, Z., Zhou, Z., Wu, Q., Chen, Z.B., Koo, E.H., and Zhong, S. (2020). Presymptomatic Increase of an Extracellular RNA in Blood Plasma Associates with the Development of Alzheimer's Disease. Curr. Biol. *30*, 1771–1782.e3.

24. Moessner, J., and Dodson, W.C. (1995). The quality of human embryo growth is improved when embryos are cultured in groups rather than separately. Fertil. Steril. *64*, 1034–1035.

25. Thibodeaux, J.K., and Godke, R.A. (1995). Potential use of embryo coculture with human in vitro fertilization procedures. J. Assist. Reprod. Genet. *12*, 665–677.

26. Gardner, D.K., and Lane, M. (1997). Culture and selection of viable blastocysts: a feasible proposition for human IVF? Hum. Reprod. Update *3*, 367–382.

27. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

28. Blakeley, P., Fogarty, N.M.E., Del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson, P., and Niakan, K.K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development *142*, 3613.

29. Valsechi, M.C., Oliveira, A.B.B., Conceição, A.L.G., Stuqui, B., Candido, N.M., Provazzi, P.J.S., de Araújo, L.F., Silva, W.A., Calmon, M.D.F., and Rahal, P. (2014). GPC3 reduces cell proliferation in renal carcinoma cell lines. BMC Cancer *14*, 631.

30. Kawano, H., Nakatani, T., Mori, T., Ueno, S., Fukaya, M., Abe, A., Kobayashi, M., Toda, F., Watanabe, M., and Matsuoka, I. (2004). Identification and characterization of novel developmentally regulated neural-specific proteins, BRINP family. Brain Res. Mol. Brain Res. *125*, 60–75.

31. LeBoeuf, R.D., Ban, E.M., Green, M.M., Stone, A.S., Propst, S.M., Blalock, J.E., and Tauber, J.D. (1998). Molecular cloning, sequence analysis, expression, and tissue distribution of suppressin, a novel suppressor of cell cycle entry. J. Biol. Chem. *273*, 361–368.

32. Wechsler, D.S., Shelly, C.A., Petroff, C.A., and Dang, C.V. (1997). MXI1, a putative tumor suppressor gene, suppresses growth of human glioblastoma cells. Cancer Res. *57*, 4905–4912.

33. Masaki, T., and Shimada, M. (2022). Decoding the Phosphatase Code: Regulation of Cell Proliferation by Calcineurin. Int. J. Mol. Sci. *23*, 1122.

34. Thues, C., Valadas, J.S., Deaulmerie, L., Geens, A., Chouhan, A.K., Duran-Romaña, R., Schymkowitz, J., Rousseau, F., Bartusel, M., Rehimi, R., et al. (2021). MAPRE2 mutations result in altered human cranial neural crest migration, underlying craniofacial malformations in CSC-KT syndrome. Sci. Rep. *11*, 4976.

35. Groff, A.F., Resetkova, N., DiDomenico, F., Sakkas, D., Penzias, A., Rinn, J.L., and Eggan, K. (2019). RNA-seq as a tool for evaluating human embryo competence. Genome Res. *29*, 1705–1718.

36. Knouse, K.A., Wu, J., Whittaker, C.A., and Amon, A. (2014). Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. Proc. Natl. Acad. Sci. USA *111*, 13409–13414.

37. Liu, D., Chen, Y., Ren, Y., Yuan, P., Wang, N., Liu, Q., Yang, C., Yan, Z., Yang, M., Wang, J., et al. (2022). Primary specification of blastocyst trophectoderm by scRNA-seq: New insights into embryo implantation. Sci. Adv. *8*, eabj3725.

38. Huang, L., Bogale, B., Tang, Y., Lu, S., Xie, X.S., and Racowsky, C. (2019). Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy. Proc. Natl. Acad. Sci. USA *116*, 14105–14112.

39. Brison, D.R. (2000). Apoptosis in mammalian preimplantation embryos: regulation by survival factors. Hum. Fertil. *3*, 36–47.

40. Hardy, M.P., Audemard, É., Migneault, F., Feghaly, A., Brochu, S., Gendron, P., Boilard, É., Major, F., Dieudé, M., Hébert, M.J., and Perreault, C. (2019). Apoptotic endothelial cells release small extracellular vesicles loaded with immunostimulatory viral-like RNAs. Sci. Rep. *9*, 7203.

41. Fabian, D., Čikoš, Š., Rehák, P., and Koppel, J. (2014). Do embryonic polar bodies commit suicide? Zygote *22*, 10–17.

42. Klatsky, P., Wessel, G., Robins, J., Wittmyer, J., and Carson, S. (2010). Analysis of mRNA in human polar bodies. Fertil. Steril. *94*, S30.

43. Giacomini, E., Vago, R., Sanchez, A.M., Podini, P., Zarovni, N., Murdica, V., Rizzo, R., Bortolotti, D., Candiani, M., and Viganò, P. (2017). Secretome of in vitro cultured human embryos contains extracellular vesicles that are uptaken by the maternal side. Sci. Rep. *7*, 5210.

44. Persani, L., Rossetti, R., Di Pasquale, E., Cacciatore, C., and Fabre, S. (2014). The fundamental role of bone morphogenetic protein 15 in ovarian function and its involvement in female fertility disorders. Hum. Reprod. Update *20*, 869–883.

45. Pilot-Storck, F., Chopin, E., Rual, J.F., Baudot, A., Dobrokhotov, P., Robinson-Rechavi, M., Brun, C., Cusick, M.E., Hill, D.E., Schaeffer, L., et al. (2010). Interactome mapping of the phosphatidylinositol 3-kinase-mammalian target of rapamycin pathway identifies deformed epidermal autoregulatory factor-1 as a new glycogen synthase kinase-3 interactor. Mol. Cell. Proteomics *9*, 1578–1593.

46. Freitas, M.J., Silva, J.V., Brothag, C., Regadas-Correia, B., Fardilha, M., and Vijayaraghavan, S. (2019). Isoform-specific GSK3A activity is negatively correlated with human sperm motility. Mol. Hum. Reprod. *25*, 171–183.

47. Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

48. Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. Proc. Natl. Acad. Sci. USA *111*, 4484–4489.

49. Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. Cell Stem Cell *15*, 524–526.

50. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

51. Jeong, H.H., Yalamanchili, H.K., Guo, C., Shulman, J.M., and Liu, Z. (2018). An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. Pac. Symp. Biocomput. *23*, 168–179.

52. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

53. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

54. Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R., and Krawetz, S.A. (2013). Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. *41*, 4104–4117.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human plasma	UCSD Shiley-Marcos Alzheimer's Disease Research Center	N/A
Deposited data		
Processed SILVER-seq data	This paper	GEO: GSE227442
Human reference genome GRCh38/hg38	NCBI RefSeq assembly	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40/
RefSeq gene annotation (GTF file)	NCBI RefSeq annotation	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40/
Single-cell RNA-seq data	Xue et al.[8]	https://www.nature.com/articles/nature12364
Single-cell RNA-seq data	Yan et al.[9]	https://www.nature.com/articles/nsmb.2660
Single-cell RNA-seq data	Blakeley et al.[28]	https://journals.biologists.com/dev/article/142/18/3151/46887/Defining-the-three-cell-lineages-of-the-human
Software and algorithms		
SILVER-seq data process	This paper	https://doi.org/10.5281/zenodo.10129614
Trimmomatic (version 0.36)	Bolger et al.[50]	http://www.usadellab.org/cms/?page=trimmomatic
STAR (version 2.5.4b)	Jeong et al.[51]	https://github.com/alexdobin/STAR
featureCounts (version 1.6.1)	Liao et al.[52]	https://subread.sourceforge.net/featureCounts.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Sheng Zhong (szhong@ucsd.edu).

Material availability

This study did not generate new unique reagents.

Data and code availability

- The raw sequencing data is withheld from public repositories to safeguard participant privacy and confidentiality. The processed SILVER-seq data have been deposited at Gene Expression Omnibus (GEO). Accession numbers are listed in the key resources table.
- Code of analysis is available at Zenodo: https://doi.org/10.5281/zenodo.10129614.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human subjects

The study was approved by the Institutional Review Board at the University of California, San Diego. Female patients ages 18–45 undergoing *in vitro* fertilization at one academic fertility center were approached and provided written consent to donate spent IVF media.

METHOD DETAILS

Human In vitro fertilization (IVF) and collection of spent media

IVF was performed by trained embryologists at Reproductive Partners Fertility Center (RPFC) according to the standard IVF procedure at RPFC. Briefly, the time of egg insemination is designated as the 0th hour. Up to 5 inseminated oocytes were cultured in one drop of G-1 PLUS media (G1). At the time of fertilization check (18th hour), culture media is removed from each drop, the pronuclei are checked, and the normally fertilized embryos are transferred to fresh G1 media, with up to 5 embryos in each drop of G1 media. The spent G1 media (18h spent media) are collected. At the 68th hour, the embryos are transferred to new G-2 PLUS media (G2), and the spent G1 media are collected (68h spent media). At the 120th hour, the embryos are evaluated, and the spent media are collected

(120h spent media). Depending on the evaluation, the embryos are either frozen for transfer to the uterus or transferred to fresh G2 media for additional culturing. Those embryos subjected to additional culturing are collected at the 160th hour (160h spent media). Additionally, a set of unfertilized oocytes was cultured in parallel with the inseminated embryos, following the same scheme of media change and spent media collection as the inseminated embryos. All media are pre-equilibrated for 18 to 20 h in the benchtop incubator (37C, 7.5% CO2, 5% O2, and bal N2) prior to use.

Constructing SILVER-seq libraries

The spent media were carefully aspirated out from the original culture dishes and centrifuged at 500 ×g for 3 min under 4°C. The supernatant was aspirated and centrifuged again. A 7 μL aliquot of the supernatant of the 2nd round of centrifugation was used as the input to SILVER-seq. SILVER-seq sequencing libraries were constructed as previously described,[22] with 21 PCR cycles before the rRNA depletion step. The constructed libraries were sequenced by an Illumina NovaSeq 6000 sequencer with 2 × 50 bp read length.

Processing SILVER-seq libraries

Adapters and low-quality bases were trimmed by Trimmomatic (version 0.36).[50] The trimmed sequences were aligned to the human reference genome (GRCh38/hg38) by STAR (version 2.5.4b)[51] and de-duplicated by UMI. Read counts per gene were calculated by featureCounts (version 1.6.1)[52] with the RefSeq gene annotation GTF file (GRCh38/p14).

Detection of expressed genes from scRNA-seq data

We downloaded human embryo scRNA-seq datasets from Xue et al.[8] and Yan et al.[9] and a scRNA-seq dataset of human sperm from Sendler et al.[54] The downloaded scRNA-seq datasets were first processed using STAR (mapping to GRCh38/p14). Read counts per gene were calculated by featureCounts (version 1.6.1) with the RefSeq gene annotation GTF file (GRCh38/p14).

Identification of exRNA signatures of the developmental arrest

We used regression to identify exRNA signatures of the developmentally arrested embryos in 120h and 160h stages as below,

$$Y_{ij} = \beta_o + \beta_{1i} A_j + R_{ij},$$

where i is the index of exRNAs, and j is the index of the stage ($j \in \{120h, 160h\}$). The response variable Y is the normalized counts of exRNA i at stage j. A a binary indicator of arrest vs. developing embryo. β_0 is the intercept term. R_{ij} is the residual term. The multiple hypothesis testing adjusted p value (adjusted p value) is derived for every exRNA.

The random forest regression model

We used the R package {randomForest} to implement the random forest model. All 63 120h samples were used for training the regression model. exRNAs that expressed (normalized counts >0) in over 70% samples were input as features.

The random forest classification model

We used the R package {randomForest} to implement the random forest model. We dichotomized the average quality score into high (score $\geq$25) and low (score <25) for classification purpose. The features of the model were exRNAs that expressed (normalized counts >0) in over 70% samples. In every cross-validation, the 63 samples were split into 44 training samples, including 22 arrest and 22 developing samples, as well as 19 testing samples, including 10 arrest and 9 developing samples. The average AUC was calculated based on 10 cross-validations. In evaluating the impact of the number of input genes on RF classification model performance, different cutoffs of Pearson correlation coefficients (between exRNA level and average quality score) were used to select the features. The cutoffs corresponding to the points in Figure 3C are 0.005, 0.008, 0.010, 0.014, 0.018, 0.022, 0.026, 0.030, 0.034, 0.038, 0.040, 0.060, 0.080, 0.100, 0.150, 0.200, 0.250, 0.300, 0.350, 0.400, 0.450, 0.500.

Singularly cultured embryo quality evaluation

We used the R package {randomForest} to implement the random forest model. We dichotomized the average quality score into high (score $\geq$25) and low (score <25) for each individually cultured embryo. We employed leave-one-out cross-validation (LOOCV) on each individually cultured embryo. In each round of LOOCV, one single-embryo culture droplet was reserved for testing, while the other single-embryo culture droplets, as well as all the co-culture droplets (62 samples in total), were utilized for training.

Cluster analysis

We performed a cluster analysis using uniform manifold approximation and projection (UMAP) method (R package {umap}) with stage genes called by one-way ANOVA (p value <0.01, 4334 genes in total).

QUANTIFICATION AND STATISTICAL METHODS

exRNA read counts were converted to normalized counts (NC) by DESeq2[53] in R (R package DESeq2). An exRNA gene is called as detected at a stage if this gene is detected (NC > 5) in at least 50% of the droplets collected at that stage.

Single-cell read counts were converted to normalized counts (NC) by DESeq2 in R (R package DESeq2). Genes detected at a developmental stage are called if a gene is detected (NC > 5) in at least 50% of the scRNA-seq libraries of that developmental stage.

Benjamini-Hochberg procedure was used to call the significant enriched GO terms with cutoff adjusted p value <0.1.

All other statistical analysis was performed with R and the details can be found in the figure legends, results, and corresponding methods details, including the statistical tests used, exact value of n, what n represents, and significance cutoffs.