

DNA methylome reveals cellular origin of cell-free DNA in spent medium of human preimplantation embryos

Yidong Chen,^{1,2,3} Yuan Gao,^{1,2,3} Jialin Jia,^{1,2,4,5} Liang Chang,^{1,2,4,5} Ping Liu,^{1,2,4,5} Jie Qiao,^{1,2,3,4,5} Fuchou Tang,^{1,2,3} Lu Wen,^{1,2} and Jin Huang^{1,2,4,5}

¹Beijing Advanced Innovation Center for Genomics, Department of Obstetrics and Gynecology, and ²Biomedical Pioneering Innovation Center and Center for Reproductive Medicine, Third Hospital, Peking University, Beijing, China. ³Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ⁴Key Laboratory of Assisted Reproduction and Key Laboratory of Cell Proliferation and Differentiation, Ministry of Education, Beijing, China. ⁵Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing, China.

The discovery of embryonic cell-free DNA (cfDNA) in spent embryo culture media (SECM) has brought hope for noninvasive preimplantation genetic testing. However, the cellular origins of SECM cfDNA are not sufficiently understood, and methods for determining maternal DNA contamination are limited. Here, we performed whole-genome DNA methylation sequencing for SECM cfDNA. Our results demonstrated that SECM cfDNA was derived from blastocysts, cumulus cells, and polar bodies. We identified the cumulus-specific differentially methylated regions (DMRs) and oocyte/polar body-specific DMRs, and established an algorithm for deducing the cumulus, polar body, and net maternal DNA contamination ratios in SECM. We showed that DNA methylation sequencing accurately detected chromosome aneuploidy in SECM and distinguished SECM samples with low and high false negative rates and gender discordance rates, after integrating the origin analysis. Our work provides insights into the characterization of embryonic DNA in SECM and provides a perspective for noninvasive preimplantation genetic testing in reproductive medicine.

Introduction

A key challenge in reproductive medicine is the choice of an embryo for a healthy live birth. Preimplantation genetic testing for aneuploidies (PGT-A) has been applied for analysis of chromosome aneuploidy, which occurs frequently in human embryos and cannot be accurately assessed by morphology alone (1). The current PGT-A approaches include embryo biopsies of polar bodies (2), blastomeres (3, 4), or trophectoderm (TE) (5, 6), among which TE biopsy has been shown to be superior to others and is most commonly used. However, TE biopsy does have limitations. It requires separation of 5–10 cells from the external TE layer of a blastocyst, which involves embryo handling and requires specialized equipment (7). Additionally, it can lead to misdiagnosis due to embryo mosaicism (8, 9). Furthermore, the TE cell removal process is invasive and it is still controversial whether it has adverse effects on implantation potential of the embryo (10, 11). Zhang et al. have recently shown that TE biopsy is associated with a 3-fold increase in the risk of preeclampsia (12). Thus, the long-term potential risk to offspring safety should be considered.

Recent studies have demonstrated that embryonic chromosomal aneuploidy can be detected using cell-free DNA (cfDNA) in spent embryo culture media (SECM) (13–18). Because noninvasive PGT-A (niPGT-A) does not affect the embryo itself, it is a promising approach for PGT (13). One issue that is not completely understood is

the origin and composition of SECM. SNP sequencing and sex chromosome analysis have revealed the existence of maternal DNA in SECM, which leads to gender discordance and false negative results (19, 20). However, the cellular origins of these maternal DNAs have not been elucidated by genomic approaches in these studies.

We and others have profiled the genome-wide dynamic DNA methylation reprogramming of human preimplantation embryos using single-cell DNA methylome sequencing (21–25). DNA methylomes of oocytes and sperm are quite different, and both of them are different from somatic cells (25). The polar bodies, which are byproducts of an oocyte meiotic division, have a DNA methylation pattern similar to that of metaphase II (MII) oocytes (21). After fertilization, genome-wide DNA demethylation occurs in female and male pronuclei and all through the cleavage stage to the blastocyst stage; the DNA methylation level is lowest at the blastocyst stage. Different DNA methylation signatures of embryos at distinct preimplantation stages, and germ and somatic cell types, should help tracing cellular origins of SECM, thus resembling plasma cfDNA tissue mapping (26, 27).

In this study, we performed post-bisulfite adaptor tagging-based single-cell whole-genome DNA methylation sequencing (scBS-seq) on 194 SECM samples as well as cumulus cell samples. We demonstrated that cfDNA in SECM was derived from blastocysts, cumulus cells, and polar bodies. We further examined chromosome aneuploidy using scBS-seq, aiming to increase the diagnostic accuracy of niPGT-A by integrating the cellular origin and chromosome aneuploidy information (Figure 1).

Results

Cumulus contamination in SECM. We performed scBS-seq on 194 SECM cfDNA samples using a protocol that does not require extracting cfDNA (see Methods). For each sample, we sequenced an average

Authorship note: YC and YG contributed equally to this work.

Conflict of interest: The authors have declared that no conflict of interest exists.

Copyright: © 2021, American Society for Clinical Investigation.

Submitted: November 13, 2020; **Accepted:** April 28, 2021; **Published:** June 15, 2021.

Reference information: *J Clin Invest.* 2021;131(12):e146051.

<https://doi.org/10.1172/JCI146051>.

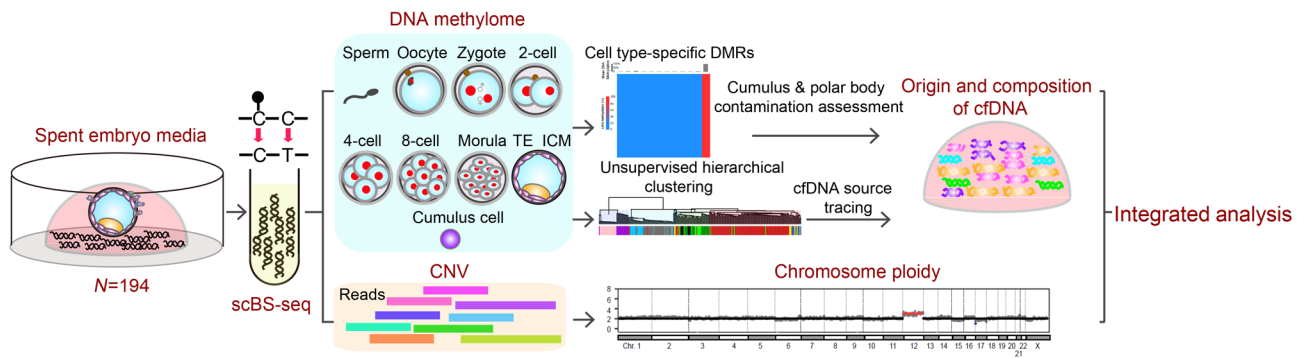


Figure 1. Study outline. We performed scBS-seq of SECM, which provided 2 layers of information: DNA methylation and chromosome aneuploidy. We deduced the origin and composition of SECM using the DNA methylome maps of human preimplantation embryos, germ cells, and cumulus cells. scBS-seq was also used to detect chromosome aneuploidy. By calculating the maternal DNA contamination ratio, we can identify the samples with low false negative and gender discordance rates.

of 5 Gb, generating 3.6 Gb of clean data, which covered an average of 5.3 million CpG sites ($\geq 1\times$) (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/JCI146051DS1>). Using the quality criterion of the number of unique mapping reads being greater than 1 million, 191 (98.5%) good-quality samples were obtained for subsequent analysis. We also performed scBS-seq on cumulus cells ($n = 12$) obtained from 4 individuals and sequenced an average of 8 Gb for each sample.

We retrieved the scBS-seq data of the preimplantation embryonic cells and the germ cells published in our previous study (23). The whole-genome DNA methylation levels of the SECM cfDNAs ranged from 13% to 74%, with a median value of 36%, and these levels were significantly higher than the reported levels in the inner cell mass (ICM) and TE (24% and 24% for ICM and TE, respectively; $P < 0.01$, two-tailed Mann-Whitney-Wilcoxon [MWW] test). Clustering analysis showed that a portion (50 of 191) of the SECM was clustered with cumulus cells (cluster III, Figure 2A). These samples displayed high DNA methylation levels (average 60%), which were close to those of the cumulus cells (average 71%).

To accurately assess the fraction of cumulus cell-derived DNA in SECM, we identified 769 CpG islands as cumulus differentially methylated regions (C-DMRs) that were highly methylated in cumulus cells and nearly unmethylated in preimplantation embryonic cells, including the ICM, TE, and oocytes (average methylation levels of 92%, 4%, and 3% for cumulus cells, ICM/TE, and MII oocytes, respectively; Supplemental Table 2 and Figure 2B). Notably, the average methylation levels of these C-DMRs were positively correlated with the whole-genome DNA methylation levels of SECM, indicating that the high whole-genome methylation levels of the SECM could largely be attributed to contamination of the cumulus cells ($R = 0.93$, $P < 2.2 \times 10^{-16}$, Pearson's correlation; Figure 2C).

We determined that approximately half of the SECM samples (95 of 191) were contaminated with cumulus cells (C-DMR methylation levels higher than 8% [mean 4% + 3 SD ($3 \times 1.3\%$) of the C-DMR methylation level in ICM/TE]). Among them, approximately half (50 of 95) displayed moderate contamination (C-DMR methylation levels 8% to 40%), and the other half (45 of 95) displayed severe contamination (C-DMR methylation levels $>40\%$). As expected, the whole-genome methylation levels increased from the no- to severe-contamination groups (Figure 2D).

Together, these results show that our DNA methylation analysis confirmed the assumption of cumulus contamination in SECM.

Detection of chromosome aneuploidy by scBS-seq. We have previously shown that scBS-seq is capable of assessing copy number (CN) variations (CNVs) (28, 29). We analyzed HCT116 cells, and the results showed that scBS-seq and multiple annealing and looping-based amplification cycles (MALBAC; ref. 30) gave the same expected CN profiles (Supplemental Figure 1A). To estimate the lower limit of the sequencing depth for accurate CNV calling, we downsampled the data, and the results showed that the coefficient of variation remained as low as 2 Mb (Supplemental Figure 1B). The majority of the SECM samples (182 of 191) gave informative CN profiles; the remaining 9 showed more than 6 aneuploidies and were defined as “aneuploid-chaotic” and were not used for further analysis. According to consistency between SECM and TE biopsies, the embryos were divided into 4 categories: (a) euploid in SECM versus euploid in TE biopsy (Euploid-Euploid), (b) euploid in SECM versus aneuploid in TE biopsy (Euploid-Aneuploid), (c) aneuploid in SECM versus euploid in TE biopsy (Aneuploid-Euploid), and (d) aneuploid in SECM versus aneuploid in TE biopsy (Aneuploid-Aneuploid). The Aneuploid-Aneuploid samples were further grouped into “Full ploidy concordance,” “Partial ploidy concordance (overlapping),” “Partial ploidy concordance (complementary),” and “Partial ploidy concordance (nonoverlapping)” (Supplemental Figure 2A). Figure 3 and Supplemental Figure 2B show the representative samples in each category.

Notably, SECM with no cumulus cell contamination showed the highest general concordance rate (GCR) (73.9%, 68 of 92) and the lowest false negative rate (FNR) (13.7%, 7 of 51), while SECM with severe cumulus cell contamination showed the lowest GCR (46.5%, 20 of 43) and the highest FNR (90.0%, 18 of 20) (Figure 2E). The false positive rates (FPRs) were 41.5%, 35.0%, and 21.7% in the no-, moderate-, and severe-contamination groups, respectively (Figure 2E). These “false positive” cases should mainly reflect CNV mosaicism of the embryo that was detected in the SECM cfDNA but not detected by the TE biopsy. Since the cumulus cell is mostly euploidy, an increase in the cumulus DNA fraction, i.e., contamination, will result in an increase in the euploid DNA fraction and thus a reduction in the “false positive” aneuploidy rate, which is not a technical artifact but indeed embryonic mosaicism.

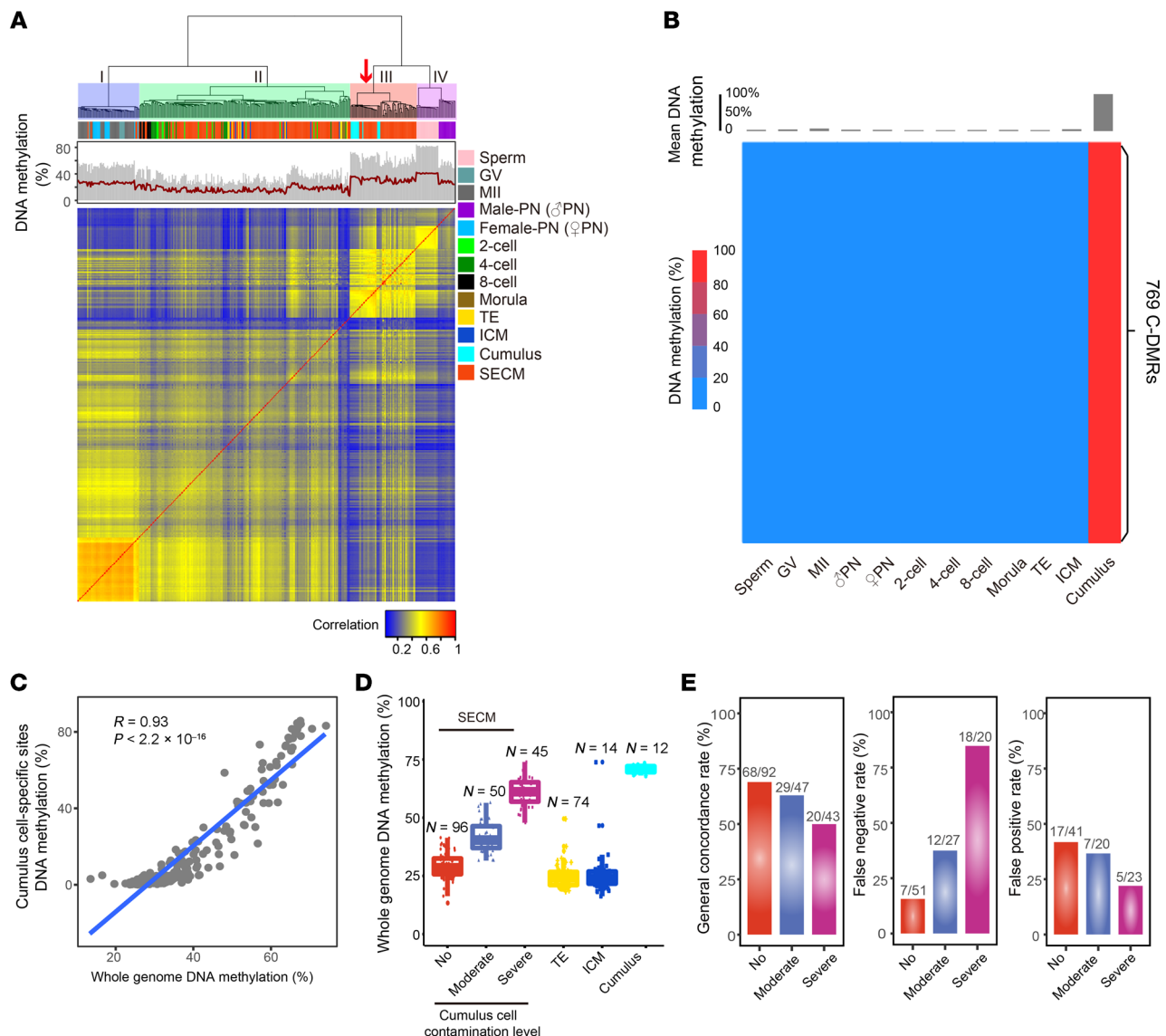


Figure 2. Assessment of cumulus contamination in SECM cfDNA. (A) Unsupervised hierarchical clustering analysis of DNA methylation levels in the SECM cfDNA samples, human preimplantation embryos, germ cells, and cumulus cells. GV, germinal vesicle oocytes; MII, metaphase II oocytes; PN, pronuclei. (B) Heatmap of 769 CpG islands (C-DMRs) that are specifically hypermethylated in cumulus cells. (C) Scatter plot showing a positive correlation between the whole-genome DNA methylation levels and the C-DMR methylation levels in the SECM cfDNAs. The 2-tailed Mann-Whitney-Wilcoxon test was used to assess significance. (D) Box-and-whisker plot showing the whole-genome DNA methylation levels of the ICM, TE, cumulus cells, and 3 SECM cfDNA groups with no, moderate, and severe cumulus contamination degrees as estimated by the C-DMR methylation levels. (E) Bar plots showing the general concordance rate, false negative rate, and false positive rate of the 3 groups of SECM compared with TE biopsy.

Together, our results demonstrated that scBS-seq is sensitive for detecting chromosome aneuploidy in SECM. The cumulus contamination led to an increased FNR, a decreased FPR, and a decreased GCR.

Polar body contamination in SECM. To further explore the cellular origins of SECM, we performed clustering analysis for the samples with no cumulus contamination ($n = 96$), as well as for the preimplantation embryonic cells and germ cells. The results showed that most SECM samples (92 of 96) were clustered with the ICM and TE, while 1 (S167) and 2 (S176 and S193) samples were notably clustered with the MII oocytes and female pronuclei, respectively (Figure 4A). Since the genomic

DNA of oocytes and pronuclei should not be released, these SECM most likely contained components of polar bodies, which are produced by the oocyte during meiosis.

To further assess polar body contamination, we identified 548 oocyte/polar body-specific DMRs (O-DMRs) with high methylation in MII oocytes but low methylation in preimplantation embryonic cells and cumulus cells (average methylation levels of 19%, 22%, and 82% for cumulus cells, ICM/TE, and oocytes, respectively; Supplemental Table 2 and Figure 4B), assuming that polar bodies have similar DNA methylation profiles to those of oocytes. The 3 SECM samples indeed displayed significantly higher methylation levels for the O-DMRs than the other SECM samples

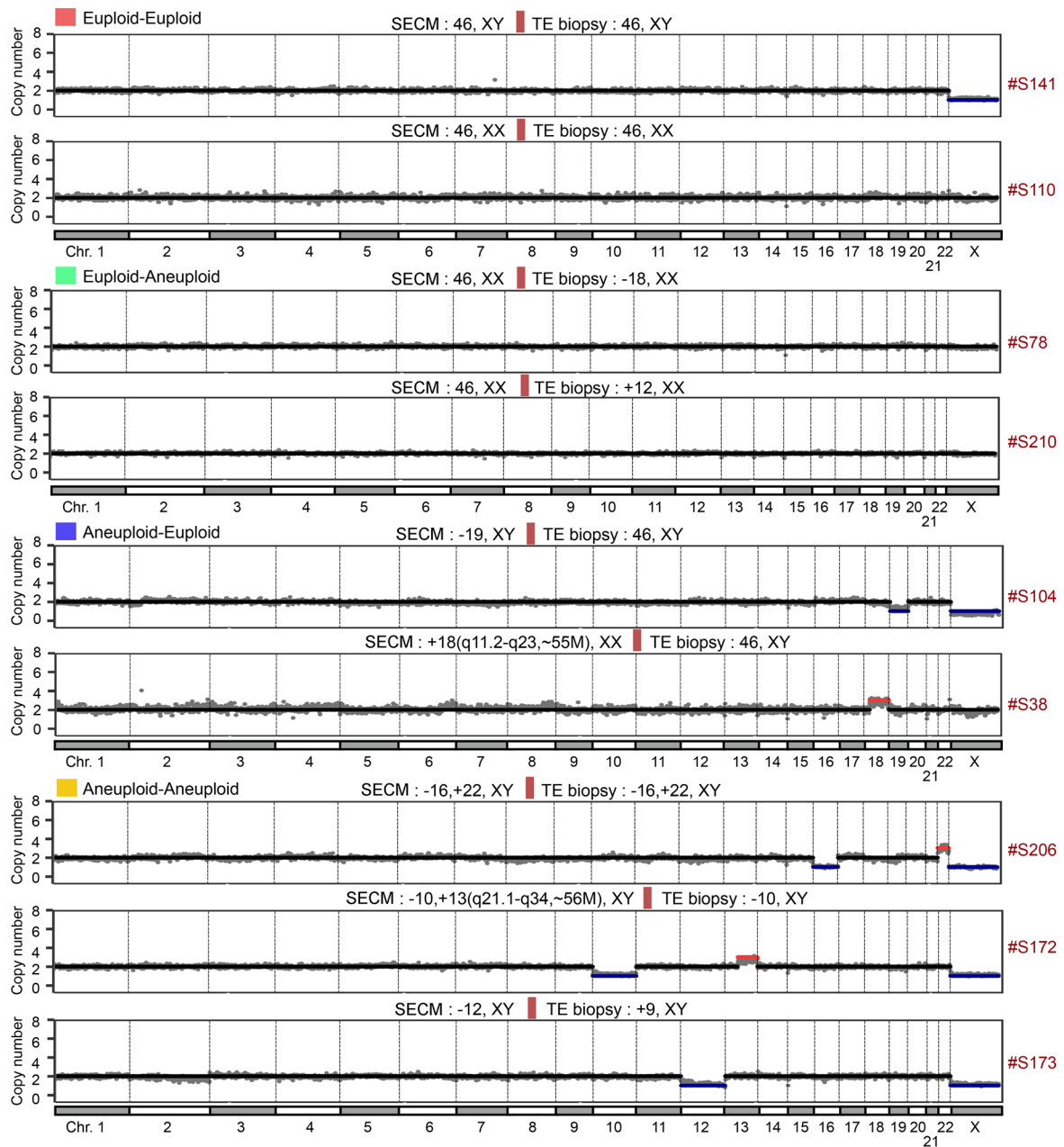


Figure 3. scBS-seq detects chromosome aneuploidy in SECM. Representative CN profiles of SECM in different categories. The results of SECM versus TE biopsy are presented.

(median methylation levels 100%, 56%, and 79% for S167, S176, and S193, respectively, versus a median of 14% for the other SECM samples, $P < 0.01$; Supplemental Figure 3A).

Remarkably, the chromosome CN profiles showed that all 3 SEM samples were false negative or gender discordant; the TE biopsy results were “46, XY” for S176 and S193 and “-21, XX” for S167, but all 3 SECM samples were “46, XX” (Figure 4C and Supplemental Figure 3B). They were clearly not contaminated by cumulus cells, as shown by the C-DMR methylation levels.

We determined that approximately one-third (27%, 53 of 191) of the SECM samples were contaminated with polar bodies (O-DMR methylation levels higher than 31% [mean 22% + 3 SD (3 × 3%) of the O-DMR methylation level in ICM/TE]). We also exam-

ined the non-CpG methylation level, which is higher in oocytes than in embryonic cells of other preimplantation stages (31). The results showed that the methylation levels in both the CHG and CHH (non-CpG) contexts were positively correlated with the O-DMR methylation levels (CHG: $R = 0.52$, $P = 4.6 \times 10^{-8}$; CHH: $R = 0.55$, $P = 6.8 \times 10^{-9}$; Pearson’s correlation, 2-tailed MWW test; Figure 4D).

We also explored whether the SECM cfDNA was derived from the ICM or TE. Our recent study profiled the DNA methylation patterns of epiblast (EPI) and TE samples using single-cell triple-omics sequencing (32). Principal component analysis (PCA) showed that the EPI and TE can be roughly separated based on DNA methylation profiles. We focused on the day 6 SECM samples with no cumulus cell or polar body contamination ($n = 61$).

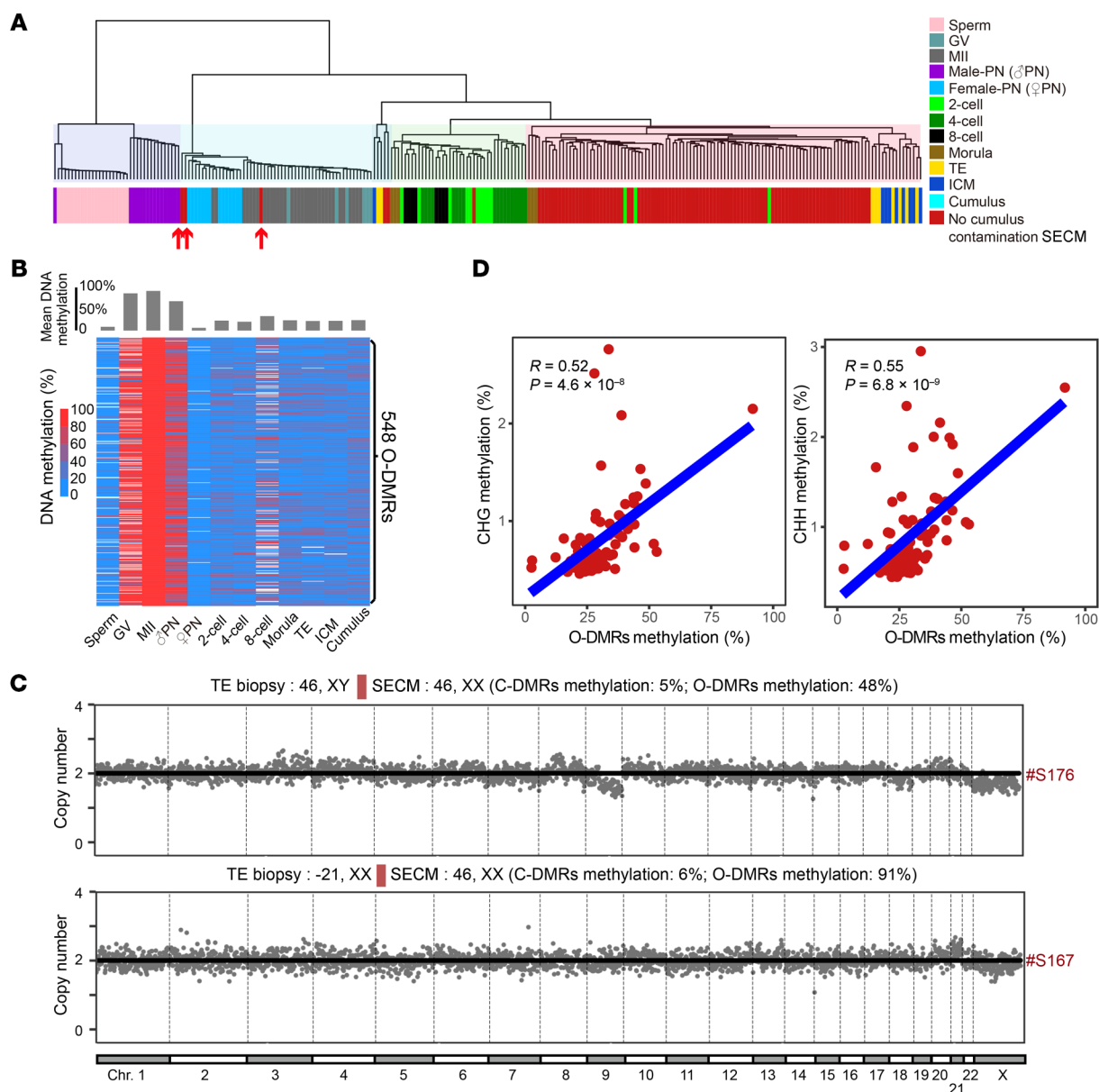


Figure 4. Polar body contamination in SECM. (A) Unsupervised hierarchical clustering of whole-genome DNA methylation for the SECM samples with no cumulus cell contamination, human preimplantation embryos of different stages, germ cells, and cumulus cells. GV, germinal vesicle oocytes; MII, meta-phase II oocytes; PN, pronuclei. (B) A total of 548 regions (O-DMRs) were specifically hypermethylated in the MII oocytes. (C) Chromosome CN profiles of 2 SECM samples clustered with the female pronuclei (upper, S176) or the MII oocytes (lower, S176). The chromosome aneuploidy results of TE biopsy and SECM are indicated, along with the methylation levels of the C-DMRs and the O-DMRs. (D) Correlations between non-CpG (left, CHG; right, CHH) DNA methylation levels and the O-DMR DNA methylation levels. CHG/CHH are short for methylation levels on non-CpG islands; H represents A (adenine) or T (thymine). The 2-tailed Mann-Whitney-Wilcoxon test was used to assess significance.

The results showed that approximately one-third (18 of 61) of SECM samples were positioned with TE and that approximately two-thirds (43 of 61) were positioned with EPI ($P < 0.01$, χ^2 test) (Supplemental Figure 3C). The promoter methylation levels of EPI differentially expressed genes (DEGs) divided by those of TE DEGs can distinguish between EPI and TE. The distribution suggested that SECM can be derived from both the TE and ICM ($P < 0.01$, two-tailed MWW test; Supplemental Figure 3D).

Together, the DNA methylation clustering, DMR, non-CpG, and chromosome CN analyses demonstrated the presence of polar body contamination in SECM.

Deducing the maternal DNA contamination ratio and integrated chromosome aneuploidy analysis. We next sought to deduce the maternal DNA contamination ratio. The methylation levels of the C-DMRs and O-DMRs were used to set up an algorithm for deducing the cumulus and polar body DNA fractions in SECM, respectively, and then, 2 fractions were added to obtain the net maternal DNA contamination ratio (see Methods). To test the accuracy of the approach, we performed simulation analysis by generating a series of synthetic data sets with different cumulus and polar body percentages mixed with the ICM/TE (Figure 5A). As shown in Figure 5B, the estimated

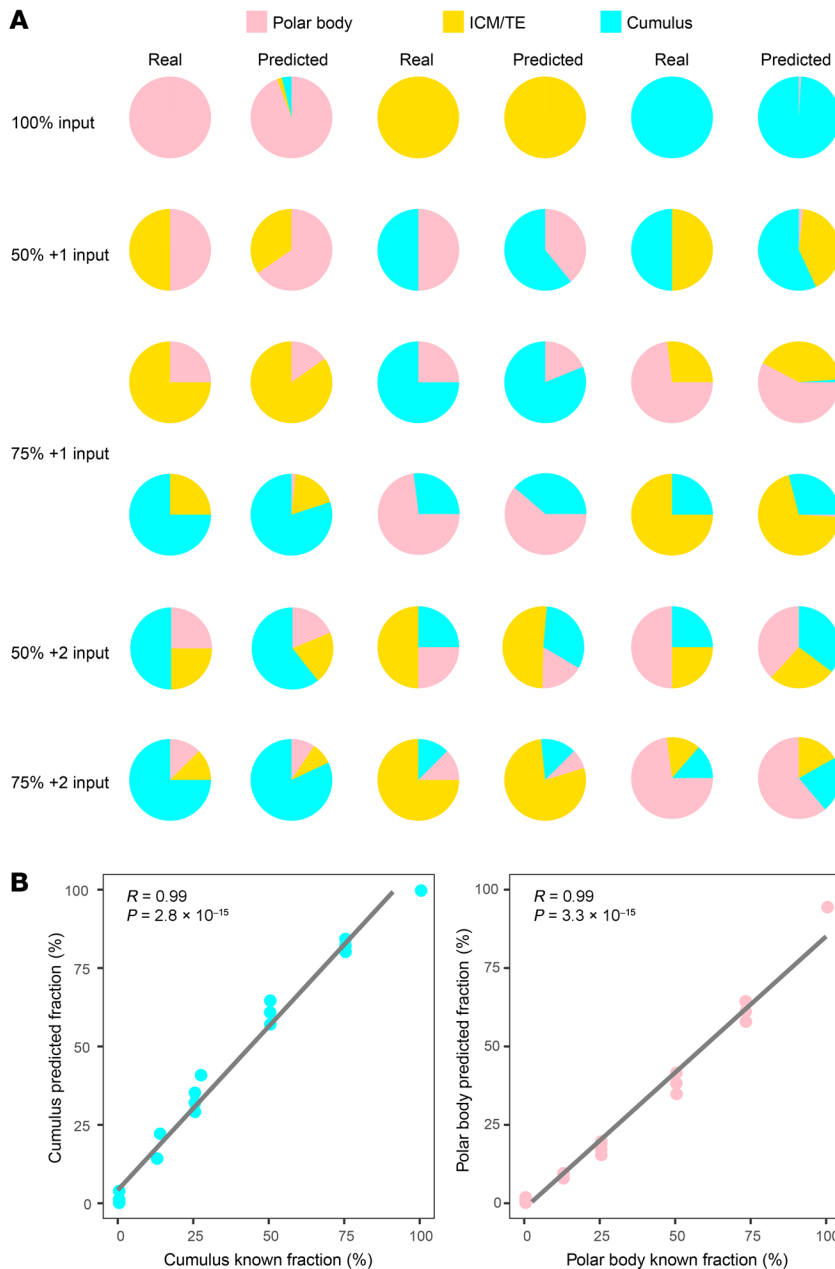


Figure 5. DNA mixing analysis. (A) Pie charts depicting the results of the simulated DNA mixing experiment. Different percentages of DNA methylation data of the polar body (the MII oocyte), ICM/TE, and cumulus cells were mixed, including 100% input from 1 of the 3 components (100% input), 50% input from each of 2 components (50% + 1 input), 75% input of 1 component plus 25% input of 1 other component (75% + 1 input), 50% input of 1 component plus 25% each of the other 2 components (50% + 2 input), and 75% input of 1 component plus 12.5% each of the other 2 components (75% + 2 input). The input percentages and the predicted percentages are shown for comparison. (B) Correlations between the predicted and input component fractions of the simulated DNA mixing experiment. The 2-tailed Mann-Whitney-Wilcoxon test was used to assess significance.

To investigate the effect of maternal contamination, we calculated sensitivity, specificity, positive and negative predictive value, as well as the gender discordance rate (GDR) and FNR, using the TE biopsy as the reference (Supplemental Figure 4B). Notably, the GDR reached zero (0%, 0 of 24) when the net maternal ratio was less than 20%, indicating that this SECM group indeed had minimal maternal contamination (Figure 6C). In contrast, the GDR remained at 18% (9 of 49) when only the cumulus cell ratio was less than 20% and remained at 42% (24 of 57) when only the polar body ratio was less than 20%. Examination of the chromosome CN profiles confirmed that these samples were affected by contamination from the corresponding maternal components, as shown in Figure 4C. This further confirmed that both the cumulus cells and polar body contributed to maternal contamination.

Interestingly, the FNR was still high (16%, 6 of 37) when the net maternal ratio was less than 20%. Close examination of the chromosome CN profiles suggested that these SECM FNRs were mosaic aneuploidy with signs

percentages correlated well with the input percentages of the DNA mixtures, which gave linear regression lines ($R = 0.99$, Pearson's correlation, 2-tailed MWW test).

We then assessed SECM. The cumulus cells contributed to more severe contamination (cumulus cell ratio > 60%: 39 of 182, 22%) than the polar body (polar body ratio > 60%: 7 of 182, 4%; Supplemental Figure 4A). Two fractions were slightly correlated ($R = -0.19$, Pearson's correlation, 2-tailed MWW test), possibly reflecting that situations such as lower embryonic fractions lead to higher maternal fractions from both origins (Figure 6A). It was clear that high polar body ratios occurred in SECM samples with no or mild cumulus cell proportions and vice versa (Figure 6A). For net maternal DNA contamination, approximately one-third (31.3%, 57 of 182) of the samples showed a ratio greater than 60%, and one-third (34.1%, 62 of 182) showed a ratio less than 20% (Figure 6B).

of CN gain or loss matching or complementing the TE biopsy results in most (5 of 6) cases (Figure 6, C and D, and Supplemental Figure 4C). This suggested that these embryos contained both aneuploid and euploid cells, with the euploid cells not sampled by TE biopsy.

Both the GDR and FNR increased with increasing cumulus cell, polar body, and net maternal contamination ratios. Remarkably, when the net maternal ratio was higher than 60%, the GDR and FNR increased to 100% (31 of 31) and 75% (6 of 8), respectively (Figure 6C).

We also examined the sampling time and found that the cumulus ratios, GDR, and FNR were significantly lower in the day 6 samples than in the day 5 samples (Supplemental Figure 4, D and G). The amplified DNA amounts were significantly higher in the day 6 samples than in the day 5 samples, indicat-

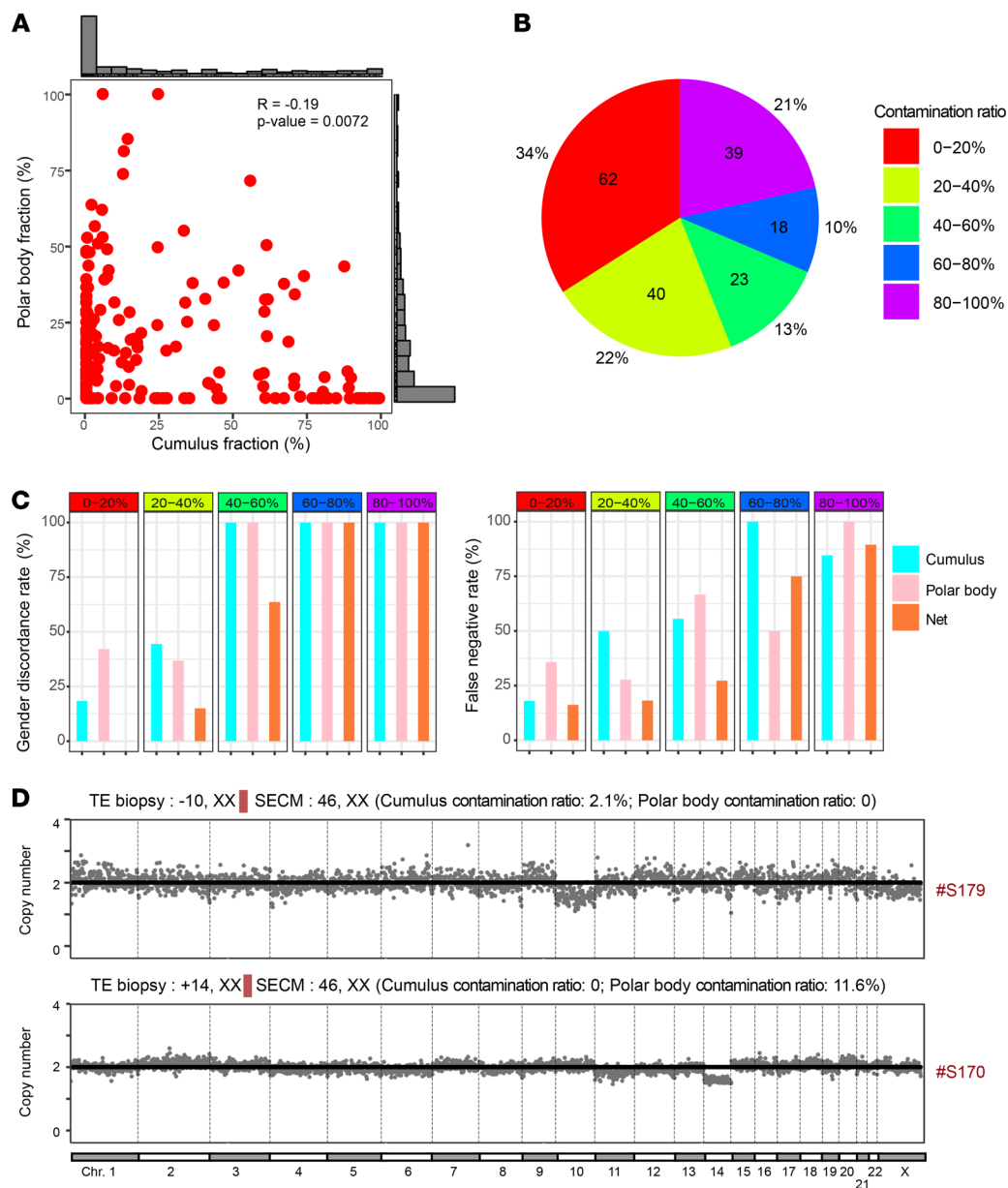


Figure 6. Maternal DNA ratio in SECM and integrated chromosome aneuploidy analysis. (A) Scatter plot showing the correlation between the cumulus cell and polar body contamination fractions in SECM. The percentage distribution of each fraction is shown. The 2-tailed Mann-Whitney-Wilcoxon test was used to assess significance. **(B)** Pie chart showing the numbers and percentages of the SECM samples with different net maternal DNA contamination ratios. **(C)** Histograms showing GDRs (left) and FNRs (right) for different ratios of cumulus cell, polar body, and net maternal contamination. **(D)** Representative CN profiles for false negative SECM with nearly no maternal DNA contamination.

ing that the day 6 samples had more embryonic DNA (Supplemental Figure 4F). Interestingly, the polar body ratios were not different between these 2 groups, suggesting that the polar body DNA continued to be released between day 5 and day 6 (Supplemental Figure 4E).

Next, we wanted to determine the impact of maternal contamination and chromosome CN on DNA concentration in the culture medium. Our results showed that the amplified DNA amount decreased with increasing maternal contamination ratios, suggesting that the main variable was the amount of embryonic DNA (2-tailed MWW test; Supplemental Figure 5A). The amplified DNA amount was not different between embryos with and without CNVs (2-tailed MWW test; Supplemental Figure 5B).

In summary, we established an algorithm for deducing the maternal contamination ratio using scBS-seq, which allowed recognition of the SECM samples with a low GDR and FNR in the chromosome aneuploidy analysis.

Discussion

In this study, we conducted single-cell whole-genome DNA methylation analysis for a large cohort of SECM cfDNA samples. First, our results traced the cellular origins of the SECM cfDNA to blastocysts, cumulus cells, and polar bodies. An unexpected finding is the polar body origin. Polar bodies are small cells that are released by an oocyte during meiosis I (first polar body) or during meiosis II immediately after fertilization (second polar body). The first polar body usually undergoes degeneration before fertilization and is present in only a quarter of embryos at the zygotic stage in mice, while the second polar body is present in all zygotes and undergoes degeneration during preimplantation development, being only occasionally found in blastocysts in mice (33). Therefore, we propose that the polar body contamination in SECM mainly comes from the second polar body. Polar body contamination complicates the use of SECM for niPGT-A, as polar bodies are less likely to be removed than cumulus cells. Our results showed that

polar body contamination occurred in approximately one-third of the SECM samples in our cohort, but fortunately, the percentage of SECM samples with severe polar body contamination (>60%) was as low as 4%. The chromosome aneuploidy analysis clearly showed that polar body contamination increased the GDR and FNR and thus should be considered in SECM applications.

Second, we deduced maternal DNA contamination via cumulus cell- and oocyte-specific DMRs. Compared with the SNP sequencing approach, our approach does not require analysis of TE cells (to obtain information about the embryonic haplotype) and thus is applicable for a potential niPGT-A; it also does not require analysis of the follicular fluid DNA (to obtain information about the maternal haplotype) or amplification of target SNP loci and is thus more convenient (15). Although the scBS-seq data are generally sparse, the selected cell type-specific DMRs spanned several hundred megabases, allowing for accurate estimation of the methylation level. The lower limit of DNA input for estimating cumulus contamination was approximately 1 cell at a correlation coefficient of 0.97 between the predicated and known cumulus cell fractions. The correlation coefficient for polar body contamination was lower, reaching 0.91 for 2 cells (Supplemental Figure 5C).

Third, we demonstrated that the DNA methylation approach allowed simultaneous assessment of chromosome ploidy and maternal contamination. The maternal DNA contamination ratio provides important information for interpreting chromosome ploidy profiles. The FNR and GDR markedly increased to 75% and 100%, respectively, when the net maternal DNA ratio was higher than 60%. This indicated that the chromosome ploidy results of these samples, which accounted for approximately one-third (60 of 182) of the samples in our cohort, were highly unreliable. If the cumulus cells could be more thoroughly removed (cumulus ratio < 20%), we would expect that the results of only 10% of the samples would be unreliable due to polar body contamination. Under such a situation, the FNR and GDR were expected to be 10% to 20%. Interestingly, this seems to be the case for a recent SECM study showing an FNR of 5% and a GDR of 15% (16). This study showed an increase in ongoing implantation rates in the euploid TE/euploid SECM group compared with the euploid TE/aneuploid SECM group, showing promise for the use of SECM for PGT-A. The results may be further improved by recognizing the FNRs caused by polar body contamination.

Our approach is not without hurdles and challenges. First, maternal contamination is still the major technological challenge for the spent medium-based PGT-A. Although we are now able to infer the fraction of maternal contamination, we are not able to decrease it. Operationally removing cumulus cell contamination is complex, but we could collect day 6 or later SECM to decrease maternal contamination and increase the proportion of embryonic DNA. Second, single-cell DNA methylome technology still needs to be improved, including read mapping, library complexity, and coverage uniformity. Third, the mitotic error, which raises issues of mosaicism and self-correction, should limit prediction accuracy; this limitation is faced by both the TE biopsy- and SECM-based methods. Further, even though studies have reported high concordance between the genomes of SECM cfDNA and embryos, whether the SECM cfDNA is a better representation of the embryo than the TE biopsy is still an open question (17, 34).

Meiotic aneuploidy, which should be accurately predicted, occurs in older women who have fewer eggs and the harm of false positive diagnoses is especially great. Recent studies have tried to establish a prioritization system to tackle these limits (35). In such a system, subchromosomal abnormalities and abnormalities with high levels of mosaicism, which are more likely to be self-corrected, would be prioritized for transfer before other abnormalities, whereas whole-chromosome abnormalities with no mosaicism should not be transferred since they are highly likely to be meiotic aneuploidy, which can never be self-corrected. DNA methylation information should help improve the accuracy of prioritization.

In summary, we provide insights into the cellular origins of SECM cfDNA and have developed an approach for integrated analyses of both the maternal contamination ratio and chromosome aneuploidy. This DNA methylation-based approach has an advantage over traditional genomic approaches because it additionally provides maternal DNA information. With further investigations to improve its accuracy and resolution, we hope to achieve a DNA methylation-based niPGT-A.

Methods

Human SECM sample collection. A total of 194 PGT-A blastocysts and their corresponding culture media were included in this study. In all of these PGT-A cycles, fertilization was performed by intracytoplasmic sperm injection (ICSI) on the day of oocyte retrieval. ICSI involves injecting a single sperm into an egg to fertilize it using a microoperating system. On day 3, the embryos were moved to the blastocyst culture medium. On day 4, each compacted embryo or morula was carefully denuded of surrounding cumulus cells again, thoroughly washed, and then cultured individually in a new dish (15 μ L of each culture drop). The culture media were collected in polymerase chain reaction (PCR) tubes when the embryos reached a fully expanded blastocyst stage, generally between day 5 and day 6; three samples were collected on day 7. The samples were stored at -20°C . The TE of the corresponding blastocyst was biopsied, and each biopsy specimen was vitrified individually. The biopsied cells were analyzed with the SNP array for PGT analysis, as we previously described (36).

Whole-genome DNA methylation sequencing of SECM. The method for detecting DNA methylation of SECM is based on the single-cell whole-genome methylation sequencing method (37). Briefly, SECM was replenished to a volume of 20 μ L with nuclease-free water, lysed with a corresponding volume of lysis buffer (20 mM Tris-EDTA, 20 mM KCl, 0.3% Triton X-100, and 1 mg/mL proteinase K) at 50°C for 1.5 hours, and then treated with bisulfite using an EZ-96 DNA Methylation-Direct MagPrep kit (Zymo Research). After purification, the first strand of DNA was synthesized using random primer P5-N9 (5'-CTACACGAC-GCTCTCCGATCTNNNNNNNNN-3') with Klenow polymerase. This step was performed 4 times. The second strand of DNA was synthesized using P7-N9 primers (5'-AGACGTGTGCTCTCCGATCTNNNNNNNNN-3'). The index primer and Illumina universal PCR primer were used for PCR amplification to obtain the library for sequencing.

DNA methylation data processing. First, we removed the sequencing adapters, amplification primers, and low-quality bases in the raw bisulfite sequencing read ends using the Perl script, which was previously published (38). Then, we discarded R2 reads that had more than 3 unmethylated CHs as well as the corresponding R1 reads. The clean reads were mapped to the human reference genome (hg19) using BS-Seeker2 v2.1.1

(<https://github.com/BSSeeker/BSSeeker2>) with the end-to-end alignment mode. The unaligned reads were rematched to the hg19 genome with local alignment mode, and the low confidence alignments within microhomology regions were removed. Next, PCR duplicates were removed using Picard tools v1.119 (<https://broadinstitute.github.io/picard/>). The DNA methylation level was calculated as the ratio of the number of reads with methylated C to that of total reads (methylated and unmethylated); only CpG sites covered by more than 3 reads were used for calculation. Samples with unique mapping reads greater than 1 million were considered for further analysis.

Determining cumulus cell or polar body origin. We used unsupervised hierarchical clustering to group the samples with similar methylation levels, so as to infer the source of SECM cfDNA. If SECM and cumulus cells were grouped together, it indicated that there was cumulus cell contamination in SECM. The same was true for polar body contamination. For unsupervised clustering, the whole genome was divided into 1-kb tiles, and the average DNA methylation level of each tile was calculated. Then, the correlation coefficient of the methylation levels between samples was calculated using the “cor” function with the parameter “method=‘spearman’, use=‘pairwise.complete.obs’” in R (<https://cran.r-project.org/web/packages/pheatmap/>).

Cumulus cell and polar body DMRs. The CpG island (CGI) ($n = 27,435$) data were downloaded from the University of California, Santa Cruz database (genome.ucsc.edu). CGIs on sex chromosomes were excluded to minimize the gender effect. The cumulus-specific CGIs were selected using the following criteria: (a) methylation level in cumulus cells greater than 80%; (b) methylation levels in other cell types (sperm, germinal vesicle, and MII oocytes, female and male pronuclei, 2-cell, 4-cell, and 8-cell, morula, ICM, and TE) less than 20%. The oocyte/polar body-specific CGIs were selected from 20,984 MII oocyte-specific hypermethylated regions (23) using the following criteria: (a) being detected in 3 or more MII oocytes with an average methylation level higher than 80%, (b) being detected in 3 or more ICM samples with a methylation level less than 20%, and (c) cumulus cells having a methylation level of less than 50%.

Inferring CNV. We evaluated CNV using the software Ginkgo (39) with a few modifications. First, we divided the whole genome into 2,705 length-variable bins with a median length of 1 Mb, with highly variable bins being excluded. The BED files, which were transformed from the aligned BAM files using bedtools v2.22.1 (<https://bedtools.readthedocs.io/>), were used as the input files. Genomic GC content bias was corrected by Lowess normalization. The BED file synthesized from randomly extracted normal diploid reads was applied as the reference. The CNV line plots were drawn by ggplot2 (<https://cran.r-project.org/web/packages/ggplot2/>), and the CNV circle plots were drawn by the circlize R package.

$$\begin{aligned} \text{GCR} &= (\text{SECM}_{\text{euploid}} \text{TE}_{\text{euploid}} + \text{SECM}_{\text{aneuploid}} \text{TE}_{\text{aneuploid}}) / \text{all SECM}; \\ \text{FNR} &= \text{SECM}_{\text{euploid}} \text{TE}_{\text{aneuploid}} / (\text{SECM}_{\text{euploid}} \text{TE}_{\text{aneuploid}} + \text{SECM}_{\text{aneuploid}} \text{TE}_{\text{aneuploid}}); \\ \text{FPR} &= \text{SECM}_{\text{aneuploid}} \text{TE}_{\text{euploid}} / (\text{SECM}_{\text{aneuploid}} \text{TE}_{\text{euploid}} + \text{SECM}_{\text{euploid}} \text{TE}_{\text{euploid}}); \\ \text{GDR} &= \text{SECM}_{\text{female}} \text{TE}_{\text{male}} / (\text{SECM}_{\text{female}} \text{TE}_{\text{male}} + \text{SECM}_{\text{male}} \text{TE}_{\text{male}}). \end{aligned}$$

PCA to distinguish the source of DNA. The top 300 highly DEGs of 3 lineages (EPI, PE, and TE) identified by RNA data in our previous study (32) were downloaded. We calculated the methylation levels of the promoters (3000 bp upstream and downstream of the transcription start site) of these genes in our SECMs, which had no cumulus cells or polar bodies, that were collected on day 6 and in our previously published EPI/TE DNA methylome data, also collected on day 6

(32). Then, we divided the methylation levels on the promoters by the methylation level of the whole genome to correct for the bias caused by sequencing coverage. The normalized promoter methylation levels of the top 300 DEGs were determined by PCA with the FactoMineR package (<https://cran.r-project.org/web/packages/FactoMineR/>).

Deducing maternal DNA fractions in SECM and simulation analysis. The mathematical relationship between the methylation levels of SECM and the corresponding methylation levels in each component of DMR_i can be expressed by the formula $MM_i = \sum MC_{ik} \times P_k \times a_{ik}$, in which MM_i represents the methylation levels of DMR_i in SECM cfDNA, MC_{ik} represents the methylation levels of DMR_i in component k , and P_k represents the proportional contribution of component k to SECM cfDNA. There are 2 types of DMRs, namely, C-DMRs and O-DMRs, and 3 components in SECM, namely, blastocysts, cumulus cells, and polar bodies, with the net proportions of the 3 components being 100%. The DNA methylation levels of the DMRs in the components are known as (a) C-DMRs, averaging 92% in cumulus cells, 4% in blastocysts, and 3% in oocytes/polar bodies; and (b) O-DMRs, averaging 19% in cumulus cells, 22% in blastocysts, and 82% in oocytes/polar bodies. The correction factor a_{ik} represents the PCR amplification efficiency of DMR_i in component k , as PCR amplification of the bisulphite-converted DNA is often biased toward the unmethylated allele (40). Our data showed that a_{ik} was approximately 0.6 for C-DMRs in cumulus cells and 0.6 for O-DMRs in polar bodies, after setting other values as 1 (Supplemental Table 2).

For the simulation analysis, we synthesized an “average” MII oocyte, blastocyte, and cumulus cell by sampling high-quality data of the MII oocytes ($n = 33$), ICM/TE (ICM, $n = 9$; TE, $n = 9$) and cumulus cells ($n = 12$). For example, we randomly sampled 3,030,303 unique mapping reads from each of 33 MII oocytes to synthesize the “average” MII oocyte with 1 million reads. Then, we randomly chose a certain proportion of reads from this MII oocyte, such as 50% (50,000,000 reads), and mixed it with a certain proportion of reads from other cell types, such as 50% (50,000,000 reads) from the blastocyte, to generate a mixed cell, with approximately 1 million uniquely mapped reads.

Statistics. A 2-tailed MWW test was used in all figures that require significance testing, except Supplemental Figure 3C, for which we used the χ^2 test. Statistically significant comparisons are shown, with significance defined as P less than 0.05.

Study approval. This study was approved by the Reproductive Medicine Ethics Committee of Peking University Third Hospital (Research License 2019-393-02).

Data availability. All sequencing data of this manuscript have been deposited in the National Genomics Data Center of the China National Center for Bioinformation (<https://bigd.big.ac.cn/gsa/>), with accession number HRA000332.

Author contributions

JH, LW, FT, and JQ conceived the project. YC and YG conducted all the studies. YC performed the bioinformatics analysis. YG performed the experiments with the help of JJ, LC, and PL. YC, LW, YG, and JH wrote the manuscript, with contributions from all of the authors.

Acknowledgments

We thank the donors who participated in the studies. We sincerely appreciate the support from grants from the National Key R&D Program of China (no. 2018YFC1003100 to LW and JH) and the

National Natural Science Foundation of China (no. 82071721). We are also thankful for the support from the Beijing Advanced Innovation Center for Genomics at Peking University and from the Computing Platform of the Center for Life Science for data analysis.

Address correspondence to: Jin Huang, Center for Reproductive Medicine, Room 216, No. 49 Huayuan North Road, Beijing 100191, China. Phone: 010.82265033; Email: 13501313874@163.com.

Or to: Lu Wen, Biomedical Pioneering Innovation Center, Room 308, No. 5 Yiheyuan Road, Beijing 100871, China. Phone: 010.62755246; Email: wenlu@pku.edu.cn. Or to: Fuchou Tang, Biomedical Pioneering Innovation Center, Room 309, No. 5 Yiheyuan Road, Beijing 100871, China. Phone: 010.62744062; Email: tangfuchou@pku.edu.cn. Or to: Jie Qiao, Peking University Third Hospital, director's office, No. 49 Huayuan North Road, Beijing 100191, China. Phone: 010.82266886; Email: jie.qiao@263.net.

- Capalbo A, et al. Correlation between standard blastocyst morphology, euploidy and implantation: an observational study in two centers involving 956 screened blastocysts. *Hum Reprod.* 2014;29(6):1173–1181.
- Montag M, et al. Polar body biopsy. *Fertil Steril.* 2013;100(3):603–607.
- Ray PF, et al. Preimplantation genetic diagnosis of beta-thalassaemia major. *Lancet.* 1996;347(9016):1696.
- Handyside AH, et al. Biopsy of human preimplantation embryos and sexing by DNA amplification. *Lancet.* 1989;1(8634):347–349.
- Kokkali G, et al. Birth of a healthy infant following trophectoderm biopsy from blastocysts for PGD of beta-thalassaemia major. *Hum Reprod.* 2005;20(7):1855–1859.
- Schoolcraft WB, et al. Clinical application of comprehensive chromosomal screening at the blastocyst stage. *Fertil Steril.* 2010;94(5):1700–1706.
- Capalbo A, et al. Consistent and reproducible outcomes of blastocyst biopsy and aneuploidy screening across different biopsy practitioners: a multicentre study involving 2586 embryo biopsies. *Hum Reprod.* 2016;31(1):199–208.
- Vera-Rodriguez M, Rubio C. Assessing the true incidence of mosaicism in preimplantation embryos. *Fertil Steril.* 2017;107(5):1107–1112.
- McCoy RC. Mosaicism in preimplantation human embryos: when chromosomal abnormalities are the norm. *Trends Genet.* 2017;33(7):448–463.
- Scott RT Jr, et al. Cleavage-stage biopsy significantly impairs human embryonic implantation potential while blastocyst biopsy does not: a randomized and paired clinical trial. *Fertil Steril.* 2013;100(3):624–630.
- Guzman L, et al. The number of biopsied trophectoderm cells may affect pregnancy outcomes. *J Assist Reprod Genet.* 2019;36(1):145–151.
- Zhang WY, et al. Maternal and neonatal outcomes associated with trophectoderm biopsy. *Fertil Steril.* 2019;112(2):283–290.
- Xu J, et al. Noninvasive chromosome screening of human embryos by genome sequencing of embryo culture medium for in vitro fertilization. *Proc Natl Acad Sci U S A.* 2016;113(42):11907–11912.
- Hammond ER, et al. Nuclear and mitochondrial DNA in blastocoele fluid and embryo culture medium: evidence and potential clinical use. *Hum Reprod.* 2016;31(8):1653–1661.
- Vera-Rodriguez M, et al. Origin and composition of cell-free DNA in spent medium from human embryo culture during preimplantation development. *Hum Reprod.* 2018;33(4):745–756.
- Rubio C, et al. Embryonic cell-free DNA versus trophectoderm biopsy for aneuploidy testing: concordance rate and clinical implications. *Fertil Steril.* 2019;112(3):510–519.
- Huang L, et al. Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy. *Proc Natl Acad Sci U S A.* 2019;116(28):14105–14112.
- Ho JR, et al. Pushing the limits of detection: investigation of cell-free DNA for aneuploidy screening in embryos. *Fertil Steril.* 2018;110(3):467–475.
- Hammond ER, et al. Characterizing nuclear and mitochondrial DNA in spent embryo culture media: genetic contamination identified. *Fertil Steril.* 2017;107(1):220–228.
- Fang R, et al. Chromosome screening using culture medium of embryos fertilised in vitro: a pilot clinical study. *J Transl Med.* 2019;17(1):73.
- Guo HS, et al. The DNA methylation landscape of human early embryos. *Nature.* 2014;511(7511):606–610.
- Smith ZD, et al. DNA methylation dynamics of the human preimplantation embryo. *Nature.* 2014;511(7511):611–615.
- Zhu P, et al. Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet.* 2018;50(1):12–19.
- Wen L, Tang F. Human germline cell development: from the perspective of single-cell sequencing. *Mol Cell.* 2019;76(2):320–328.
- Smallwood SA, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet.* 2011;43(8):811–814.
- Sun K, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A.* 2015;112(40):E5503–E5512.
- Liu X, et al. Comprehensive DNA methylation analysis of tissue of origin of plasma cell-free DNA by methylated CpG tandem amplification and sequencing (MCTA-Seq). *Clin Epigenetics.* 2019;11(1):93.
- Hou Y, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 2016;26(3):304–319.
- Bian SH, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science.* 2018;362(6418):1060–1063.
- Zong CH, et al. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science.* 2012;338(6114):1622–1626.
- Li CR, et al. DNA methylation reprogramming of functional elements during mammalian embryonic development. *Cell Discov.* 2018;4:41.
- Zhou F, et al. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature.* 2019;572(7771):660–664.
- Fabian D, et al. Do embryonic polar bodies commit suicide? *Zygote.* 2014;22(1):10–17.
- Jiao J, et al. Minimally invasive preimplantation genetic testing using blastocyst culture medium. *Hum Reprod.* 2019;34(7):1369–1379.
- Rubio C, et al. Multicenter prospective study of concordance between embryonic cell-free DNA and trophectoderm biopsies from 1301 human blastocysts. *Am J Obstet Gynecol.* 2020;223(5):751.e1–751.e13.
- Huang J, et al. Validation of a next-generation sequencing-based protocol for 24-chromosome aneuploidy screening of blastocysts. *Fertil Steril.* 2016;105(6):1532–1536.
- Smallwood SA, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods.* 2014;11(8):817–820.
- Wu P, et al. Using local alignment to enhance single-cell bisulfite sequencing data efficiency. *Bioinformatics.* 2019;35(18):3273–3278.
- Garvin T, et al. Interactive analysis and assessment of single-cell copy-number variations. *Nat Methods.* 2015;12(11):1058–1060.
- Warnecke PM, et al. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.* 1997;25(21):4422–4426.