Embryo Genome Profiling by Single-Cell Sequencing for Preimplantation Genetic Diagnosis in a β-Thalassemia Family

Yanwen Xu,† Shengpei Chen,† Xuyang Yin,† Xiaoting Shen,† Xiaoyu Pan,† Fang Chen,† Hui Jiang,† Yu Liang, Wei Wang, Xun Xu, Jian Wang, Xiqing Zhang,* Canquan Zhou,* and Jun Wang*†

BACKGROUND: The embryonic genome, including genotypes and haplotypes, contains all the information for preimplantation genetic diagnosis, representing great potential for mendelian disorder carriers to conceive healthy babies.

METHODS: We developed a strategy to obtain the full embryonic genome for a β-thalassemia–carrier couple to have a healthy second baby. We carried out sequencing for single blastomere cells and the family trio and further developed the analysis pipeline, including recovery of the missing alleles, removal of the majority of errors, and phasing of the embryonic genome.

RESULTS: The final accuracy for homozygous and heterozygous single-nucleotide polymorphisms reached 99.62% and 98.39%, respectively. The aneuploidies of embryos were detected as well. Based on the comprehensive embryonic genome, we effectively performed whole-genome mendelian disorder diagnosis and human leukocyte antigen matching tests.

CONCLUSIONS: This retrospective study in a β-thalassemia family demonstrates a method for embryo genome recovery through single-cell sequencing, which permits detection of genetic variations in preimplantation genetic diagnosis. It shows the potential of single-cell sequencing technology in preimplantation genetic diagnosis clinical practices.

© 2015 American Association for Clinical Chemistry

In the past 2 decades, in vitro fertilization (IVF)13 with preimplantation genetic diagnosis (PGD) has played a key role in reproductive clinics. Couples carrying chromosomal abnormalities or mendelian disorders would be advised to choose IVF with PGD (1, 2), especially those with an affected child. The affected child provides important information for PGD, revealing mutations related to disease as pathogenic variants for mendelian disease diagnosis (3). During the IVF process, genetic materials can be obtained from embryos at 3 stages: polar body biopsy (4), cleavage biopsy at day 3 (5), and blastocyst biopsy at day 5/6 (6). Cleavage biopsy for a single blastomere is the most popular approach for PGD indications.

β-Thalassemia (OMIM 613985) is a group of autosomal recessive monogenic disorders caused by mutations in the β-globin genes (7). It is currently the most common group of monogenic disorders. To date, the available definitive cure for β-thalassemia major is hematopoietic stem cell transplantation from an HLA-identical donor. The HLA system is the name of the MHC in humans. It contains a large number of genes that encode antigen-presenting proteins that play a major role in the immune system.

* Address correspondence to: X.Z. at: BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China 518083. Fax +86-755-25273884; e-mail zhangxq@genomics.org.cn. C.Z. at: First Affiliated Hospital of Sun Yat-sen University, 58 Zhongshan Rd. II Guangzhou, Gu, China 510080. Fax +86-20-87750632; e-mail zhoucanquan@hotmail.com. J.W. at: BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China 518083. Fax +86-755-25273884; e-mail wangj@genomics.org.cn.

Received June 6, 2014; accepted January 13, 2015. Previously published online at DOI: 10.1373/clinchem.2014.228569

© 2015 American Association for Clinical Chemistry

† Nonstandard abbreviations: IVF, in vitro fertilization; PGD, preimplantation genetic diagnosis; WGA, whole-genome amplification; ADO, allele dropout; PA, preferential amplification; MDA, multiple displacement amplification; SNP, single-nucleotide polymorphism; NGS, next-generation sequencing; WGS, whole-genome sequencing; AF, amniotic fluid; MAR, missing allele rate; HMM, hidden Markov model; CR, copy ratio; PED, pedigree; Ch0, X chromosome.
role in immune system function. PGD of single-gene disorders combined with HLA matching has emerged as a tool for couples at risk of transmitting a genetic disease to select unaffected embryos of an HLA tissue type compatible with that of an existing affected child (8). At delivery, hematopoietic stem cells from the newborn umbilical cord blood can be used to treat the affected sibling. During the PGD procedure, aneuploidy testing for preimplantation embryos is necessary owing to the high incidence of chromosomal abnormalities in embryos cultured in vitro. The sex information of embryos can be an indicator for PGD as well. Hence, genome-wide profiling of embryos with diagnosis of single gene disorders, HLA matching, sex, and aneuploidy provides an approach to PGD for mendelian disorder carriers, such as the carriers of β-thalassemia.

Traditionally, multiplex PCR has been used to detect the pathogenic variants of an embryo with short tandem repeat markers in close proximity to the causative gene as a diagnosis backup. With the introduction of whole-genome amplification (WGA) to amplify the biopsied embryonic single cell, the complete embryonic genome can be obtained to facilitate the diagnosis of genetic variants in embryos. However, the allele dropout (ADO) and preferential amplification (PA) in WGA still restrict the diagnostic accuracy of WGA-based PGD. The ADO in multiple displacement amplification (MDA) is reported to range from 10.0% to 38.9% (9). Karyomapping based on single-nucleotide polymorphism (SNP) array has been reported to overcome the ADO obstacles in single-cell analysis (10). For the embryonic aneuploidy testing, detection of chromosomal abnormalities based on array comparative genomic hybridization (11), SNP array (12), or next-generation sequencing (NGS) (13–17) in biopsied embryonic cells have been published. In this study, genome profiling of human embryos using single blastomere whole-genome sequencing (WGS) was assessed in a β-thalassemia family. The ADO in WGA was addressed and the haplotype of the embryo was established.

Materials and Methods

SAMPLE RECRUITMENT
We recruited a couple who had a daughter suffering from β-thalassemia. This couple with a maternal age of 37 years underwent IVF with PGD to obtain a healthy baby. Single-cell biopsy was performed on day 3 after fertilization and amplified using the MDA approach. Seven blastomeres were biopsied. A single cell was aspirated with a sampling micropipette going through the split in the zona pellucida. After biopsy, embryos were cultured in blastocyst medium (SAGE Biopharma) until vitrification. PCR–reverse dot blot analysis was used to diagnose β-thalassemia, and short tandem repeat linkage analysis was used for HLA matching. The pregnant mother received amniocentesis for prenatal diagnosis of β-thalassemia at 16 weeks of gestation. Both the PGD and prenatal diagnosis indicated a female embryo that was negative for β-thalassemia (not a carrier of or affected with β-thalassemia) and that the HLA types matched those of the affected child. A healthy female baby was born at 38 weeks of gestation in July 2012. We collected a portion of the WGA production of the 7 blastomeres with the amniotic fluid (AF) of the corresponding IVF fetus, as well as 5 mL of whole blood from this couple and the proband daughter. Each participant provided informed consent under the protocol approved by the Ethics Committee of the first affiliated hospital of Sun Yat-sen University.

WGA, LIBRARY CONSTRUCTION, AND SEQUENCING
Each blastomere cell was transferred into a PCR tube containing 3.5 μL of PBS. The wash buffer during cell sorting was used as the negative control. MDA was performed using the REPLI-g Midi kit (Qiagen) according to the instructions. The amplified DNA was stored at −20 °C for further processing.

Approximately 1 μg MDA product from each blastomere was used for library construction and sequencing. For the transferred blastomere, we performed WGS using a pair-end strategy of 90 bp with 30-fold depth.

The peripheral blood samples of parents and proband daughter as well as the AF were used for DNA extraction, library construction, All-in-One target region capture, and sequencing. The MDA products of the other 6 blastomeres nontransferred were also used for All-in-One target region sequencing. We extracted 1 μg DNA from each blood sample and AF, and constructed libraries with 200-bp insert sizes. Libraries were used for All-in-One target capture with a NimbleGen EZ sequence capture array containing 181.37-Mb target regions of the whole exome, one million tag-SNPs, the whole MHC region, and several mendelian disorder–related genes. After chip hybridization and elution, the enriched DNA libraries were quality controlled. We performed NGS for the qualified libraries with pair-end 90-bp reads. The sequence reads of this project have been deposited at NCBI SRA (SRA065291).

ALIGNMENT, SNP/INDEL CALLING
The reads with low quality and adapter contamination were filtered out before aligning. The cleaned reads were mapped to the reference genome (Hg19, NCBI release GRCh37) using Burrows-Wheeler alignment [BWA (http://bio-bwa.sourceforge.net/)] (18) with the default parameters. We removed PCR duplications using SAMtools (19). Afterward, we recalibrated the base qualities and performed the SNP/Indel calling using the Ge-
OME Analysis Toolkit (http://www.broadinstitute.org/gatk/) (20) (indels are insertions and deletion in a genome with a length of approximately 15 nucleotides).

**ORIGINAL GENOTYPE CALLING**

We performed the original genotype calling on the candidate sites, where one of the parents carried an SNP. The ADO and PA of WGA were evaluated as well. Traditionally, ADO is defined as the random amplification failure of 1 of the 2 heterozygous alleles, whereas PA is defined as a bias that favors the amplification of one allele over the other, leading to failure to reach the threshold of detection (21). For better understanding, herein we denoted the missing allele rate (MAR) as the overall percentage of ADO and PA. In this section, the expected homozygous (parents carried same homozygous SNP) and heterozygous (parents carried different homozygous SNP) would be used to estimate the sequence error rate and MAR respectively. The L statistic was defined as the odds ratio between heterozygous and homozygous genotypes:

\[
L = \log_{\text{max}} \left[ \frac{\text{Pr(genotype } = AB)}{\text{Pr(genotype } = AA, \text{ Pr(genotype } = BB)} \right],
\]

where Pr is the probability of each genotype. Pr was calculated through binomial distribution (Bin) in these formulas:

\[
\text{Pr(genotype } = AB) = \text{Bin}(N, 0.5),
\]

\[
\text{Pr(genotype } = AA) = \text{Bin}(N, 1 - e), \text{ and}
\]

\[
\text{Pr(genotype } = BB) = \text{Bin}(N, e),
\]

where the sequence error (e) was estimated as the mismatch rate at expected homozygous sites and N represents the number of expected recombinational SNPs.

**EMBRYO HAPLOTYPING PHASING**

We used heterozygous callings of original genotypes to predict the parental transmission through a hidden Markov model (HMM). An HMM can be considered a generalization of a mixture model, for which the hidden variables, which control the mixture component to be selected for each observation, are related through a Markov process. During this process, the maternal and paternal transmissions were predicted separately: (a) The hidden states of the model, \(S = \{0, 1\}\), defined which parental allele was transmitted to the embryo. In particular, 0 represented the allele transmitted to the proband sister. (b) The distinct observation symbols per state, \(V = \{0, 1\}\), expressed the consistency between the transmitted allele and the draft genotype; 1 stood for consistency between the transmitted allele and the draft genotype and 0 stood for inconsistency. (c) The initial state distribution \(\pi = \{\pi_i\}, (i \in S, \pi_i = 0.5)\) was given the prior probability of the different hidden state at the first site. (d) The state transition probability distribution, \(A = [a_{ij}], (i, j \in S)\), represented the probability of recombination or not, which could be calculated according to former studies (22). If \(i = j, a_{ij} = 1 - \frac{N_i}{N_p};\) otherwise, \(a_{ij} = \frac{N_i}{N_p}\), where \(N_i\) and \(N_p\) represent the number of expected recombinational SNPs. (e) The observation symbols probability distribution in each state, \(B = \{b(k), (i \in S, k \in V)\}\), were estimated using the expected homozygous and expected heterozygous sites:

\[
b(0) = \frac{\text{no. sites } (L > 0, \text{expected homozygous})}{\text{no. sites } (L > 0, \text{expected heterozygous})},
\]

\[
b(0) = 1 - b(0).
\]

Finally, we deciphered the optimized hidden states using HMM and the Viterbi algorithm.

**ANEUPLOIDY DETECTION**

Aneuploidies of the blastomeres were analyzed according to the depth of the heterozygous SNP alleles. The signals were filtered with 2 criteria: with size larger than 30 Mb and with ratios between the depth of the 2 alleles >2 or <0.5. After filtering, the signals with a copy ratio (CR) <0.75 were considered to be deletions, and the signals with CR >1.25 were considered to be duplications.

**SEX DETERMINATION**

Ideally, the CR of the sex chromosomes can reveal the sex of the embryo. The CR was defined as the ratio between the observed and expected reads:

\[
CR = \frac{R}{R_{\text{total}}},
\]

where subscript c is the chromosomal name and \(R\) and \(L\) represent the number of mapped reads and chromosome length, respectively. In the case of heterozygosity examination, we evaluated the difference in the \(L\) distribution on the X chromosome (ChrX) between expected heterozygous and homozygous sites by the rank-sum test.

**Results**

**NEXT-GENERATION SEQUENCING**

The couple received IVF treatment with a PGD test for both β-thalassemia and HLA. A single-blastomere WGA of the transferred embryo was collected for WGS [pedigree (PED)-E-A]. A total of 118.51 Gbp data was generated using NGS, which covered 96.8% of the reference genome [HG19, NCBI (National Center for Biotechnology Information) built 37] to 38.25× depth. To construct the parental haplotype, we performed All-in-One target sequencing using genomic DNA extracted from the peripheral blood of parents (PED-Mother & PED-Father).
and the proband daughter (PED-Sister) as well as WGA products from the 7 blastomeres, including the transferred embryo (PED-E-B) and the other 6 embryos nontransferred (PED-E-1,2,4,5,6,7). In addition, amniotic fluid of the corresponding IVF baby (PED-AF) was collected to evaluate the accuracy of method through All-in-One target sequencing (Fig. 1A). The summary of data production for NGS is displayed in Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue4.

WGA BIAS ASSESSMENT AND ORIGINAL SNP PROFILING

According to the AF genotypes on autosomes, we found that 34004 of 158643 embryo heterozygous SNPs had dropped at least 1 allele, corresponding to an ADO rate of 21.43% (Fig. 2A). Approximately 56% of the ADO occurred in the paternal allele on the autosomes (Fig. 2B). As was reported previously (23, 24), we observed a GC content–related bias, namely the ADO rate was increased in GC-poor (<30%) and GC-rich (>30%) genomic regions (see online Supplemental Fig. 1).

To eliminate this bias, we performed original genotype calling of the embryo genome using the parameter L to determine the original genotype (see online Supplemental Fig. 2). It indicated that L could isolate heterozygous sites from homozygote and ADO backgrounds. Besides the expected homozygous and heterozygous sites, 330600 original genotypes were determined successfully. According to the AF genotype on autosomes, 75.39% and 99.32% of the homozygous and heterozygous genotypes were accurately predicted (Fig. 1B). Regarding ChrX, 71.36% and 99.66% of the homozygous and heterozygous genotypes were accurately determined (Fig. 1C). This indicated an MAR of 40.02% on autosomes and 37.99% on sex chromosomes.

EMBRYO HAPLOTYPE PHASING

We phased the embryonic autosomes with the parental haplotypes, which were inferred from the genotype of the proband daughter (Fig. 1A). We successfully phased 131785 and 133601 of the maternal and paternal heterozygous sites, including mutations in the hemoglo-
bin, beta \textit{(HBB)}\textsuperscript{14} gene. Subsequently, we used an HMM to determine the parental transmission, identifying recombination breakpoints and filtering out sequencing error at the same time. The approach led to successful phasing of 98.57% of the autosomal SNPs in the embryonic genome. According to the AF genotype, the accuracy of homozygous callings was significantly improved from 75.39% to 99.73%. Using the haplotype information, 62,110 homozygous callings were identified as PA or ADO and corrected back to heterozygous with an accuracy of 95.44% (Fig. 1B), making the overall accuracy of the heterozygous callings 97.85%. Also, the autosomal MAR was markedly decreased from 40.02% to 0.50% (Fig. 1D).

We further performed a simulation by subsampling of PED-E-A to characterize how the accuracy of the embryo genotype depends on sequencing depth (see online Supplemental Fig. 3). As little as 25% of the WGS data was able to decrease the MAR to \(<1\%\) (mean depth \(= 9.5 \times\)). Concomitantly, the accuracy of the final genotypes stabilized above 97%. This information makes the overall accuracy of the heterozygous callings 97.85%.

\textsuperscript{14} Human genes: HBB, hemoglobin, beta; HLA-DRB1, SLC25A22, solute carrier family 25 (mitochondrial carrier: glutamate), member 22; INS, insulin; SMPD1, sphingomyelin phosphodiesterase 1, acid lysosomal; TH, tyrosine hydroxylase.
dicated that our model was robust for lower depth single-cell sequencing.

The accuracy of homologous recombination site pinpointing using our method was demonstrated. We compared the recombination breakpoints in parental alleles between amniotic fluid PED-AF and the embryo transferred PED-E-A (see online Supplemental Fig. 4). We also evaluated the accuracy statistics of SNP calling close to the recombination breakpoints (see online Supplemental Fig. 5), which illustrated the reliability of the embryonic haplotype phased by our method.

SEX DETECTION AND SEX CHROMOSOME ANALYSIS

Genomic information on sex chromosomes is central to the diagnoses of mendelian diseases such as Duchenne muscular dystrophy (MIM number 310200). We performed analysis on sex determination and haplotype on sex chromosomes.

The CR of sex chromosomes would have revealed the sex information (see online Supplemental Fig. 6). The relative ratios of ChrX and ChrY for the transferred embryo were 0.75 and 0.07, respectively. The lower ratio of ChrY indicated that the signals of ChrY might randomly originate from an experimental or analytical process. To reach an unequivocal conclusion, we determined the sex of the embryo by examining the heterozygosity of ChrX. Theoretically, the $L$ distribution between expected heterozygous and homozygous sites would be similar for male embryos but different for female embryos. The rank–sum test ($P$ value $= 2.2 \times 10^{-16}$) indicated a significant difference in the $L$ distribution for the transferred embryo (PED-E-A), implying a female embryo. This finding was consistent with the birth. The $L$ distribution between heterozygous and homozygous sites was significantly different for PED-E-2, PED-E-5, PED-E-6, PED-E-7, indicating the 4 embryos were female. The other 2 embryos PED-E-1 and PED-E-4 were determined as male with the same $L$ distributions. The sex information of all 7 embryos was validated by SNP array (Table 1; also see online Supplemental Fig. 8).

We next determined the haplotype of ChrX for the female embryo transferred. Of the 4887 original genotypes, over 99% of the transmitted allele was successfully predicted. The general accuracies of the homozygous and heterozygous callings were 97.18% and 99.52%, respectively (Fig. 1C). Meanwhile, the MAR was decreased from 39.73% to 0.12% (Fig. 1D).

ANEUPLOIDY DETECTION

Aneuploidies of the 7 blastomeres were detected using the depth ratio of the parental alleles in heterozygous sites. Online Supplemental Fig. 7 shows the sequencing depth of the parental alleles with CR $>0.75$ (deletion) or CR $<1.25$ (duplication) after filtering for partial chromosomes in the 7 embryos. The transferred embryo was confirmed as euploidy because the majority of filtered alleles were within 0.75 < CR < 1.25 in all the 22 autosomes and sex chromosomes. The aneuploidy results of all 7 embryos are listed in Table 1, and the NGS results were all validated by SNP array (see online Supplemental Fig. 8). The embryonic aneuploidies by NGS in this study were all consistent with the SNP array results.

WHOLE-GENOME MENDELIAN DISORDER DIAGNOSIS AND HLA MATCHING

The embryonic genome allowed us to perform a genome-wide diagnosis of mendelian disorders, including 83 common mendelian disorders in the clinic involving 384 genes (Fig. 3A). The genotypes indicated that the transferred embryo was totally negative for these 83 disorders. Again, the results were confirmed by the healthy phenotypes after birth. A detailed diagnosis on $\beta$-thalassemia and HLA matching was further performed.

Mutations in the $HBB$ gene cause $\beta$-thalassemia. The All-in-One sequencing indicated that the proband sister carried a homozygous A→C point mutation at rs7480526 (NM_000518.4:c.315 + 74T>G), a known pathogenic variant for $\beta$-thalassemia (25) in the intron region of the $HBB$ gene. Both parents carried a single copy of this mutation and exhibited a normal phenotype.

Table 1. Aneuploidies detection for the 7 embryos by NGS and SNP array.

<table>
<thead>
<tr>
<th>Embryos</th>
<th>NGS results</th>
<th>SNP array results</th>
</tr>
</thead>
<tbody>
<tr>
<td>PED-E-A</td>
<td>46,XX</td>
<td>46,XX</td>
</tr>
<tr>
<td>PED-E-1</td>
<td>46,XY</td>
<td>46,XY</td>
</tr>
<tr>
<td>PED-E-2</td>
<td>46,XX</td>
<td>46,XX</td>
</tr>
<tr>
<td>PED-E-4</td>
<td>41,XY, −1(M), −6(P), −12(P), −14(M), −15(P)</td>
<td>41,XY, −1, −6, −12, −14, −15</td>
</tr>
<tr>
<td>PED-E-5</td>
<td>47,XX, +7(P), +9(M), −11(P)</td>
<td>47,XX, +7, +9, −11</td>
</tr>
<tr>
<td>PED-E-6</td>
<td>44,XX, −12(P), −19(P)</td>
<td>44,XX, −12, −19</td>
</tr>
<tr>
<td>PED-E-7</td>
<td>45,XX, +6(P), −8(P), −20(P)</td>
<td>45,XX, +6, −8, −20</td>
</tr>
</tbody>
</table>

M, maternal alleles originated; P, paternal alleles originated.
Fig. 3. Genome-wide clinical screening.

(A) Information of the genome for the transferred embryo PED-E-A. The outer to inner circles formed represent the genome-wide mendelian disorder diagnosis, chromosomal structure, sequence depth, paternal allele transmission, and maternal allele transmission. (B) Details of the β-thalassemia diagnosis for PED-E-A. (C) Details of the HLA-matching test for PED-E-A.

SLC25A22, solute carrier family 25 (mitochondrial carrier: glutamate); INS, insulin; SMPD1, sphingomyelin phosphodiesterase 1, acid lysosomal; TH, tyrosine hydroxylase.
We performed an HBB gene diagnosis based on the genotype and haplotype for all 7 embryos. The transferred embryo was homozygous A in rs7480526. Haplotypes of the proband sister were defined as Hap0/Hap0. The predicted transmission showed that the haplotype was Hap1/Hap1 for PED-E-A at rs7480526 (Fig. 3B). We diagnosed this transferred embryo as β-thalassemia negative (not a carrier of or affected with β-thalassemia), which was consistent with the AF genotypes and diagnosis after birth. Genotype at rs7480526 and predicted transmission of the haplotype for all 7 embryos are shown in Table 2 and online Supplemental Fig. 9. The genotype at rs7480526 was consistent with the predicted transmission of haplotype for all 7 embryos and was further validated by Sanger sequencing (Table 2; also see online Supplemental Fig. 10).

The genotypes of HLA genes (see online Supplemental Table 2) have been used before transplantation (26). For this family, the umbilical cord blood stem cells of the transferred embryo could be used to treat the proband sister’s thalassemia (27). We selected the heterozygous genotypes of all of the HLA genes (see online Supplemental Table 2) in NGS data to predict the transmission haplotypes of each embryo. According to the predicted transmission, the haplotypes of the transferred embryo (PED-E-A) were Hap0/Hap0 in HLA genes (Fig. 3C), consistent with the proband sister. Thus, we determined that their HLA types were matched. The predicted transmission of haplotype in MHC regions for all 7 embryos is shown in Table 2 and online Supplemental Fig. 9. The haplotype in HLA genes predicted from NGS data were further validated by PCR with Sanger sequencing of an oligonucleotide sequence in the major histocompatibility complex, class II, DR beta 1 (HLA-DRB1) gene with both paternal and maternal heterozygous alleles (see online Supplemental Fig. 11). The haplotype of HLA-DRB1 in each embryo by Sanger sequencing was consistent with the result predicted from NGS.

**Table 2.** The genotype at rs7480526, the haplotype in the HBB gene, and MHC regions for the 7 embryos.

<table>
<thead>
<tr>
<th>Embryos</th>
<th>Genotype at rs7480526</th>
<th>Haplotype in HBB gene</th>
<th>Haplotype in MHC regions</th>
<th>Genotype at rs7480526</th>
<th>Haplotype in HLA-DRB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PED-Sister</td>
<td>CC</td>
<td>Hap0/Hap0</td>
<td>Hap0/Hap0</td>
<td>CC</td>
<td>Hap0/Hap0</td>
</tr>
<tr>
<td>PED-E-A</td>
<td>AA</td>
<td>Hap1/Hap1</td>
<td>Hap0/Hap0</td>
<td>AA</td>
<td>Hap0/Hap0</td>
</tr>
<tr>
<td>PED-E-1</td>
<td>AC</td>
<td>Hap0/Hap1</td>
<td>Hap0/Hap1</td>
<td>AC</td>
<td>Hap0/Hap1</td>
</tr>
<tr>
<td>PED-E-2</td>
<td>AC</td>
<td>Hap0/Hap1</td>
<td>Hap1/Hap0</td>
<td>AC</td>
<td>Hap1/Hap0</td>
</tr>
<tr>
<td>PED-E-4</td>
<td>AC</td>
<td>Hap1/Hap0</td>
<td>Hap0/Hap0</td>
<td>AC</td>
<td>Hap0/Hap0</td>
</tr>
<tr>
<td>PED-E-5</td>
<td>AA</td>
<td>Hap1/Hap1</td>
<td>Hap0/Hap1</td>
<td>AA</td>
<td>Hap0/Hap1</td>
</tr>
<tr>
<td>PED-E-6</td>
<td>AC</td>
<td>Hap1/Hap0</td>
<td>Hap0/Hap0</td>
<td>AC</td>
<td>Hap0/Hap0</td>
</tr>
<tr>
<td>PED-E-7</td>
<td>CC</td>
<td>Hap0/Hap0</td>
<td>Hap1/Hap0</td>
<td>CC</td>
<td>Hap1/Hap0</td>
</tr>
</tbody>
</table>

**Discussion**

In a retrospective study, we demonstrated a model for retrieval of the embryo genome using single-cell sequencing assisted by parental haplotypes in a β-thalassemia family. Over 97% of the embryo autosomal genotypes and 99% of the genotypes on ChrX were reconstructed accurately. In addition, the diagnosis of β-thalassemia, HLA matching, sex, and aneuploidy were analyzed and proved to be accurate. This established the technical foundation of embryo genome recovery using single-cell sequencing in clinics.

Although we successfully recovered the embryo SNPs and made an accurate diagnosis, there are 2 major directions for further development. First, the parental haplotype can be inferred using a simple trio strategy. We can infer the parental haplotypes using these “offspring” and phase these embryo genomes reciprocally. There may be several experimental approaches to obtain the parental haplotype directly, such as a fosmid library (28), long fragment read technology (29), and microfluidic technology (30). The single-cell sequencing of sperm or second polar body can help to obtain the parental haplotypes as well. Second, in this study we focused exclusively on SNPs and aneuploidies. However, clinical applications of embryonic genomes require information for other forms of variations, e.g., short indels and de novo mutations. Indels play an important role in mendelian disorders (31) and are potential markers of complex traits (32). In this study, we also tried to detect embryo indels using our strategy. The accuracy of autosomal homozygous and heterozygous callings of indels in embryo was improved from 69.00% and 83.10% to 89.14% and 88.13%, respectively (see online Supplemental Table 3). Because of the false positives in WGA and sequencing, we were unable to confirm de novo mutations.

In traditional PGD, multiplex PCR and array-based technology have been used to identify pathogenic mu-
tations (33). However, multiplex PCR has a lower throughput and a limited number of known sites can be analyzed. Array-based technology provides a high-throughput approach such as aCGH (array comparative genomic hybridization) and SNP array, but it can identify only known common markers in human genome. Although karyomapping based on SNP array provides an option for the detection of the majority of mutations, the resolution and accuracy are restricted. NGS provides a high-throughput and single-nucleotide resolution technology to analyze the embryonic genome, enabling PGD. Compared to array-based karyomapping, NGS mainly shows technological advantages. The strength of NGS is in the detection of family-specific mutations or rare mutations, which would not be included in the probe targets of SNP array. A large number of pathogenic mutations would not be directly detected in most SNP array assays, such as the rs7480526 of the HBB gene in this study. Furthermore, detection of embryo indels was evaluated, indicating the feasibility of NGS for analysis in embryos of complicated variations that cannot be analyzed in SNP array assay.

The cost of NGS with 30-fold coverage of the embryo genome would be about $1000. However, sequencing with lower depth is also possible because, in our model, the MAR decreased <1% with a depth of 9.5×. The method theoretically can be modified so that a condensed target region for sequencing could be used and the cost reduced accordingly. The turnaround time of sequencing is about 20 working days, so that embryo vitrification is needed. Several studies have indicated that, with the introduction of vitrification, the overall efficiency during IVF–PGD might improve dramatically (34–36). Furthermore, with continuous technical improvements the cost-effectiveness of sequencing is rapidly increasing as is the accuracy.

In this study, we introduced an NGS-based method to access the embryo genome. Numerical chromosome anomalies, which often occur during human cleavage-stage embryogenesis, have been accurately detected. The ability to recover an embryo genome at high resolution using single-cell sequencing should bring new insights into the diagnoses of mendelian disorders and personal genomic analysis for PGD. In the case of parental carriage of mendelian disorders, the embryo genome provides all of the necessary evidence for diagnosis. It paves the way for PGD toward personal medicine, such as diagnosis of allergies (37). The additional information available to parents will raise ethical questions, especially regarding the prediction of novel parental pathological mutations. Another major problem in testing embryos without a phenotype for diseases that are not in the family is in interpreting the variants. Variants of uncertain significance will often be discovered, and many of these variants are difficult to classify even in affected individuals. The variants of uncertain significance must be considered in embryonic personal genome interpretation during clinical practice. The pretest and posttest genetic counseling is also substantial and should be standardized. The ethical questions raised need to be thoroughly discussed within the scientific community and at a societal level.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: C. Zhou, First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honoraria: None declared.

Research Funding: J. Wang, W. Wang, S. Chen, X. Yin, X. Pan, F. Chen, and H. Jiang, this study was funded by the Laboratory of Shenzhen Birth Defect Screening Project (JZF no. [2011] 861 and JZF no. [2011] 862) and was approved by the Shenzhen Municipal Commission for Development and Reform and Key Laboratory Project in Shenzhen (CXB2011108250096A) and Key Laboratory of Cooperation Project in Guangdong Province (2011A060906007).

Expert Testimony: None declared.


Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or final approval of manuscript.

Acknowledgments: We thank Laurie Goodman and Huijue Jia for writing advice. We thank Yingrui Li, Xu Yang, and Fengping Xu for support and thank Chunlai Zhang, Xuchao Li, and Chun Gong for assistance in data analysis.

References


