Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by next-generation sequencing with linkage analyses

Live Births after Simultaneous Avoidance of Monogenic Diseases and Chromosome Abnormality by Next Generation Sequencing with Linkage Analyses

Liying Yan[a,b,1], Lei Huang[a,1], Liya Xu[a,1],Jin Huang[a,b], Fei Ma[a], Xiaohui Zhu[a], Yaqiong Tang[a], Mingshan Liu[a], , Ying Lian[a], Ping Liu[a], Rong Li[a], Sijia Lu[c], Fuchou Tang[a, d,f,2], Jie Qiao[a,b,f,2], X. Sunney Xie[a,e,2]

[a]Biodynamic Optical Imaging Center, College of Life Sciences and Department of Obstetrics and Gynecology, Third Hospital, Peking University, Beijing 100871, China
[b]Key Laboratory of Assisted Reproduction, Ministry of Education and Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing 100191, China
[c]Yikon Genomics, 1 China Medical City Avenue, TQB Building, 5th floor, Taizhou, Jiangsu, 225300, China.
[d]Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China
[e]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
[f]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, P. R. China

[1]These authors contributed equally to the work.
[2]To whom correspondence should be addressed. E-mail: tangfuchou@pku.edu.cn (F.T.), jie.qiao@263.net (J.Q.), xie@chemistry.harvard.edu (X.S.X.)

Abstract

In *in vitro* fertilization (IVF), pre-implantation genetic diagnosis (PGD) and pre-implantation genetic screening (PGS) help patients to select embryos free of monogenic diseases and aneuploidy (chromosome abnormality). Next generation sequencing (NGS) methods, while experiencing a rapid cost reduction, have improved the precision for PGD/PGS. However, the PGD precision has been limited by the false positive and false negative single nucleotide variations (SNVs), which are not acceptable in IVF and can be circumvented by linkage analyses, such as short tandem repeats (STR) or karyomapping. Noteworthy, the existing SNV/CNV (copy number variations) detection and linkage analysis often require separate procedures for the same embryo. Here we report a NGS based PGD/PGS procedure that can simultaneously detect single gene disorder and aneuploidy, and is capable of linkage analysis in a cost-effective way. This method, called Mutated Allele Revealed by Sequencing with Aneuploidy and Linkage Analyses (MARSALA) involves Multiple Annealing and Looping Based Amplification Cycles (MALBAC) for single-cell whole genome amplification (WGA). Aneuploidy is determined by CNVs, whereas SNVs associated with the monogenic diseases are detected by PCR amplification of the MALBAC product. The false positive and negative SNVs are avoided by an NGS-based linkage analysis. Two healthy babies were born after such embryo selection, free of the monogenic diseases of their parents. The monogenic diseases originated from a single base mutation on the autosome and the X-chromosome of the disease-carrying father and mother, respectively.

Key words: Monogenic Diseases, Chromosome abnormality, IVF, Aneuploidy, NGS, PGD, PGS, MALBAC, WGA

Significance

One missing or wrong nucleotide out of six billion in a human genome can cause a genetic disease. In IVF, it has been a daunting challenge to detect such a point mutation in a single human germ cell, yet one cannot afford to make any mistakes in selecting a viable embryo for transfer. MARSALA combines next generation sequencing and single cell whole genome amplification methodologies, allowing embryo diagnosis with a single molecule precision, reducing false positive or false negative errors significantly. MARSALA can benefit couples who desire to stop transmission of their genetic diseases to their offspring.

Introduction

There are about 7,000 known monogenic diseases, the genes for more than half of which have been identified (1). Most of them can either cause death, disability or congenital malformation, bringing heavy burdens to both the affected families and the society's health care system. In

addition, chromosome abnormality, i.e. CNV at particular chromosome locations, is a major cause of miscarriages as well as genetic disorders such as Down syndrome. The probability of aneuploidy rises drastically with maternal age, resulting in a decreased live birth rate as women age.

PGD or PGS allows selecting embryos free of either single gene disorders or aneuploidy (2), respectively. Previous PGD techniques include fluorescence in situ hybridization (FISH) (3) and Sanger sequencing after polymerase chain reaction (PCR) (4), which have been used to detect either specific point mutations (by Sanger sequencing after PCR) or chromosome abnormalities (by FISH) but not both at the same time. Genome-wide aneuploidy screening, such as comparative genomic hybridization (array-CGH) (5, 6), single-nucleotide polymorphism (SNP) array (7), multiplex quantitative PCR (8), and next-generation sequencing (NGS) have been used for selecting embryos free of aneuploidy, and increasingly applied in the PGS field (5, 9). NGS offers many advantages, including reduced costs, increased precision and higher base resolution (9-11). Recently NGS has also been used to detect monogenic diseases or de novo mutations (12, 13).

Due to the small amount of genetic material, genomic analyses of a single or a few cells require whole genome amplification (WGA). There are three commercially available WGA methods, which have been recently compared (14), they include Degenerate Oligomer Primer PCR (DOP-PCR) (15), Multiple Displacement Amplification (MDA) (16), and Multiple Annealing and Looping-based Amplification Cycles (MALBAC) (17). Unfortunately, any WGA method exhibits false positive and/or false negative SNVs, which could lead to wrong selection of embryos for IVF. A major challenge for PGD is how to eliminate these errors. Linkage analysis has become a standard method to circumvent this problem (18, 19) by either detecting short tandem repeats (STR) or karyomapping with an SNP array (20-24) to determine the disease allele.

Several groups have reported the combined use of two or three of these methods in the same embryo, to increase precision in the selection of embryos (25-27). Konstantinidis et al. reported a combined procedure of detection of chromosomal abnormality and linkage analysis using the karyomapping method (23). These reports proved that simultaneous detection for single gene disorders and chromosome abnormalities can be performed. However, separate procedures had to be used for each embryo (27). Moreover, the targeted mutations still could not be observed directly in these non-NGS methods (23).

Here we present for the first time an integrated NGS-based PGD/PGS procedure, with simultaneous and direct detection of disease-causing mutations, chromosome abnormalities and linkage analyses, which have led to live births of two baby girls free of their parents' monogenic diseases.

Results and Discussion

MARSALA

We describe a novel NGS based method, Mutated Allele Revealed by Sequencing with Aneuploidy and Linkage Analyses (MARSALA) which, in a single procedure can carry out Pre-implantation Genetic Diagnosis (PGD), Pre-implantation Genetic Screening (PGS) and linkage analyses of individual embryos. As shown in Fig. 1, a few trophectoderm cells are biopsied from a blastocyst on Day 5 and amplified by WGA with MALBAC. Then WGA product is re-amplified by PCR primers at the targeted SNVs and the product is mixed with the WGA product. This mixture is sequenced by NGS at a low sequencing depth (0.1X-2X), which is sufficient for measuring CNVs accurately, and for call targeted SNVs in the same run by virtue of the pre-amplification (>1,000X).

In order to circumvent the false positive and false negative SNV errors, MARSALA relies on a large number of (rather than one uncertain) SNP markers for the disease allele in order to increase the certainty of allele identification, in a similar principle to those of STR and karyomapping analyses. This can be done if the genetic disorder derives from either parent. Alternatively, we can also analyze the two polar bodies of each zygote on Day 1 in order to deduce whether the embryo is viable for transfer if the genetic disorder derives primarily from the female (10, 28).

By NGS analyses of a single cell (or a few cells) of SNVs adjacent to a targeted mutation, MASARLA is able to avoid false positive and false negative errors. Hence MARSALA allows for simultaneous detection of both the disease-causing point mutation and chromosome abnormality via a single low-depth NGS sequencing procedure.

Family history and blastocyst biopsy

The first case involves a couple of maternal age 34 and paternal age 32. The husband has a family history of an autosomal dominant disorder, and suffers from Hereditary Multiple Exostoses (HME), which is characterized by multiple bony spurs or lumps in the bones from an early age. He underwent a series of operations previously to remove the exostoses. More recently, an exostose was developed on his spinal canal. His father, two paternal aunts and one of his cousins are HME patients with the same symptoms (Fig. 2A). Genetic diagnosis of the patient showed a frame-shift point mutation c.233delC at the EXT2 gene, already known for causing this disease (29).

This couple underwent IVF and PGD treatment. Thirty-two metaphase-II oocytes were collected and fertilized by intra-cytoplasmic sperm injection respectively (ICSI). Eighteen embryos

developed to the blastocyst stage and a cluster of trophectoderm (TE) cells was biopsied from each embryo for PGD. The DNA of all biopsies was successfully amplified by MALBAC.

In the second case, the wife carries an X-linked chromosome recessive heredity disorder. The couple of maternal age 33 and paternal age 32 had already an affected son, who suffers from X-linked hypohidrotic ectodermal dysplasia (XLHED) characterized by hair, sweat gland and teeth abnormalities (30). Genetic diagnosis of the son showed a point mutation c.T1085G at the EDA1 gene, inherited from his mother (Fig. 3A). The SNV on the EDA gene has been known to be associated with this disorder. Five embryos developed to the blastocyst stage on Day 6 and their TE cells were also biopsied. One of the embryos' TE cells (E05) was not successfully biopsied or amplified.

Simultaneous SNV and CNV detection by MARSALA

The mutated region of a target gene was re-amplified with a pair of specific primers and then the PCR products were mixed with the whole genome amplification product for NGS. In this way, the existence of the point mutation and aneuploidy can be detected in one NGS run, as shown in Figure 2. Figure 2B and Figure 2C show that the region of interest can be sequenced to ultra-high coverage (>1,000X) while still maintaining accurate CNV measurement throughout the whole genome. In doing so, a significant portion of the allele dropout, due to allele amplification imbalance, can be overcome. This method is applicable for almost all monogenic diseases with known mutations, even when there are multiple mutated sites in the alleles from one or both of the parents.

For the first case, the NGS data is shown in Table S2 with a 0.1 X genome coverage for each embryo. The number of total reads covering the disease mutation site among the 18 TE samples varied from 4,421 to 134, 373 (Table S2). Seven embryos were identified to be carriers of the mutated SNV (Fig. 2B). The CNV data of all chromosomes for each embryo are presented in Figure 2C, showing seven of them having aneuploidy.

For the second case, the number of the total reads covering the disease mutation among the polar body and TE samples varied from 316,944 to 753,555 (Table S3). Two embryos (Embryo E01, E03) were identified to have the mutated SNV (Fig. 3B). In addition embryo E03 was found to have a chromosome abnormality as well (Fig. 3C).

Linkage Analyses by MARSALA

In the first case, we sequenced the blood samples of the couple as well as the husband's disease-carrying father. We called all SNPs with sequencing depth >10x within a 1 Mb vicinity of the mutated site (X/C) and focused on the heterozygous SNPs of the husband (for example, the first row, A/G in Fig. 2D), homozygous SNPs of the wife (A/A), and homozygous SNPs of his

disease-carrying father (A/A). We know the husband's allele with base G must come from his (normal) mother since his disease-carrying father is A/A homozygous, hence, the husband's disease allele must be the allele with the base A. We used a similar strategy to deduce the inherited allele of the embryos under screening. For example, at the first SNP position (the first row, Fig. 2E), Embryo E05 is identified as homozygous A/A, then one of its alleles (A) must be derived from the husband's disease-carrying father (A), and therefore Embryo E05 should not be transferred.

The MARSALA results at the ten selected genomic positions are listed in Fig. 2E, which were highly concordant in all 18 embryos, consistent with the above PCR results of the MALBAC products (Fig. 2B). Multiple SNPs close to the target mutation can significantly increase the accuracy of identifying the mutated allele, even if the direct mutation calling is unsuccessful. Accordingly, we applied this strategy to all 18 embryos, seven were identified as affected and the remaining 11 unaffected (Table S1). Among all the embryos, neither affected nor aneuploid, Embryo E04 was selected for transfer.

For the second case, to carry out the linkage analysis of each embryo, we sequenced the genomes of the blood samples of the wife, husband and disease-carrying son. We called all SNPs of the three individuals with sequencing depth >10x within 3 Mb vicinity of the EDA1 gene and focused on the heterozygous SNPs in the X chromosome of the wife and SNPs in the X chromosomes of the affected son and unaffected husband (Fig. 3D). Analogous to the Case 1, we can deduce for each embryo whether the disease-carrying allele from the mother is present (Fig. 3D). The MARSALA results at the ten selected genomic positions, listed in Fig. 3E, are highly concordant in all four embryos, consistent with the PCR results of the MALBAC amplified DNA (Fig. 3B).

A summary of results for all embryos of Case 2 is shown in Table S1. After excluding all affected and aneuploid embryos, we selected Embryo E02 for transfer.

When using MALBAC for single cell WGA, we were able to achieve linkage analyses with only 2X sequencing depth for each embryo as well as 2X depth for the parents and a disease-carrying relative. This is because the sequence-dependent bias of MALBAC is highly reproducible from cell to cell, allowing for some genomic regions covered ~ 10X depth at the 2X average depth. In contrast, the 2X sequencing depth is not enough for MARSALA using MDA for WGA.

Comparison of PGD/PGS results of blastocyst and polar body biopsies

We have previously demonstrated the proof of principle that when a single gene disorder and aneuploidy is maternal, it is possible to deduce whether an embryo is viable for transfer by sequencing the polar bodies, PB1 and PB2 (10). Polar body biopsy is less invasive than blastocyst biopsy and can offer more time for diagnosis when fresh embryo transfer is desired,

but the cost of its genetic analysis is higher. As the monogenic disease of the second case is of maternal origin, we performed MARSALA in biopsied polar bodies in order to compare it with that of blastocyst biopsy.

Each pair of PBs (PB1 and PB2) from the five embryos was biopsied for single cell sequencing by NGS, allowing us to deduce the SNVs and aneuploidy of each female pronucleus (10, 28) If there are two copies of disease-associated alleles in PB1 and PB2, the corresponding embryo is healthy, no matter the gender, male or female. As an example, the detection of targeted mutation of Embryo E03 and E05 is shown in Fig. 4A.

The results obtained for the analyses of PB1 and PB2 are consistent with those obtained from the blastocyst biopsy, with the exception that blastocyst WGA of Embryo E05 contained errors. In this case, polar body biopsy provided an accurate alternative for deducing the female pronucleus. Fig. 4B shows the deduced CNV profile of the female pronucleus of Embryo E05, showing that this embryo is normal.

We note that polar body analysis can only detect maternal but not paternal disorders or embryo development errors. Surprisingly, in Embryo E03 chromosomal abnormalities were identified in the blastocyst biopsy analysis (shown in Fig. 3C), the copy number of Chr4 was one instead of two. However, the polar body biopsy sequencing results showed a normal copy number of Chr4 in this same embryo (Fig. 4B). We hypothesized that this discrepancy is due to the aneuploidy of the sperm cell with ~5% probability (31). Indeed this was verified by the fact that Chr4 of this blastocyst only exhibited the maternal rather than the paternal SNPs (Table S4).

Linkage analysis in principle can also be carried out to avoid false positives and false negatives of SNV detection with the genomes of the wife and her disease-carrying son (Fig. 4C).

Confirmation by Sanger sequencing, CGH array and STR analyses

To further validate our MARSALA method, we subjected our targeted PCR fragment from the MALBAC products of the 18 trophectoderm biopsies to Sanger sequencing using the primers of EXT2 c.233delC listed in Table S2. Six out of the 18 embryos (E05, E06, E10, E11, E12, and E16) carrying the paternal mutation were confirmed by Sanger sequencing (Fig. S1A and S2). Notably, the high throughput sequencing results using MARSALA were more accurate than Sanger sequencing, since no deletion on EXT2 c.233 was found in embryo E09 by the latter method, likely due to ADO (Fig. S2).

DNA from the same MALBAC products of embryos E02, E03, E04, E12, E14, E17, and E18 were analyzed by array CGH. The results obtained were exactly the same as the results obtained from the low coverage MALBAC-NGS (Fig S1B).

The two D11S1993 alleles from the father were distinguished by two STRs of different sizes. A 229bp STR was linked with the normal *EXT2* allele, whereas a 235bp STR was linked with the

mutated allele associated with HME. As for the D11S4103 allele, the 195bp STR was linked with the normal allele and the 197 bp STR was linked with the mutated allele. Thus, based on STR genotyping, seven embryos (E05, E06, E09, E10, E11, E12, and E16) were identified as carrying the mutated allele of *EXT2*, consistent with our NGS results (Figure S1C).

Fetus validation

Ultrasound examination on day 30 after embryo transfer revealed a single intrauterine gestational sac with normal fetal heart beat for the two IVF cases reported here. In the first case, to verify the lack of the mutated *EXT2* allele as well as chromosomal normality, a prenatal diagnosis was conducted with Sanger sequencing, SNP array and karyotype analysis using amniotic fluid cells at 17 weeks' gestation. For the second case, to verify the lacking of the mutated *EDA1* allele as well as chromosomal normality, a prenatal diagnosis was conducted with Sanger sequencing, SNP array and karyotype analysis using amniotic fluid cells at 19 weeks' gestation. A baby girl from the Case 1 couple was born on September 19th, 2014. And a baby girl from the Case 2 couple was born on November 30th, 2014. Both of them were confirmed to be healthy through physical examination by a senior pediatrician. The umbilical cord blood of each baby was collected for further validation and the babies were again confirmed free of the *EXT2* paternal mutation and *EDA1* maternal mutation, respectively (Fig. S3).

Cost of MARSALA

A point mutation identification normally requires 30x-sequencing depth, which is costly. However the detection of aneuploidy requires only 0.1x sequencing depth data (0.3G data), which currently costs ~$30 on Illumina HiSeq 2500 platform. By adding specific primers for the targeted genes, MARSALA avoids the need for high sequencing depths and allows simultaneous detection of aneuploidy and specific point mutations at 0.1X sequencing depth, substantially decreasing the cost. In our experience, the linkage analyses using MALBAC only requires 2x sequencing depth. Overall MARSALA is therefore cost effective.

Limitations of MARSALA

MARSALA is not capable of detecting *de novo* mutations. To use MARSALA, prior knowledge of a single gene associated to the disorder from either parent is needed. In fact, the genome sequences of the parents are prerequisites for MARSALA. which are available from carrier testing. The use of specific primers for MARSALA might be considered as a limitation. Although not costly, they could be time consuming. However, the primer sequences can be determined even before the PGD procedure based on the parents' genomes and therefore should not add extra time to the overall procedure.

We note that the primed amplification of MALBAC products can be omitted since it only serves as an additional validation for the detectable SNVs, which were not possible to read directly in the previous non-NGS based STR analyses and karyomapping.

Conclusions

First and for most, MARSALA allows for the simultaneous direct observation of aneuploidy, targeted mutation sites, as well as their linked SNPs. Compared to previous point mutation detection methods such as FISH and PCR, it significantly reduces the false positive and negative SNVs by virtue of linkage analyses. Compared to previous aneuploidy detection methods such as array CGH and SNP array, the NGS-based MARSALA offers more accurate CNV measurements. Compared to previous linkage analyses such as STR and karyomapping methods, it provides direct visualization of the mutation sites, avoiding miss-identification due to loss of linkage because of homologous recombination. MARSALA has markedly improved the precision of PGD and PGS, and streamlined the PGD/PGS procedure.

Secondly, MARSALA is cost effective compared to microarray-based methods. In principle, MARSALA can be carried out with any WGA methods. We chose MALBAC because it has the lowest ADO (false negative rate for SNV) and the highest precision for CNV detection (14). 2X sequencing depth is enough when MALBAC is used for linkage analyses.

We have demonstrated that normal embryos can be selected by a one-step procedure of next generation sequencing, MARSALA, to circumvent both monogenic diseases and chromosomal abnormalities with high precision. MARSALA works for both autosome and X-linked diseases, male and female carriers. The first two neonates resulting from such embryo selection for a male and a female patient each carrying a monogenic disease were successfully born.

Materials and Methods

Blastocyst biopsy. We adapted the widely used approach of embryo biopsy at the blastocyst stage(10), collecting a few TE cells from each hatching blastocyst on Day 5. All embryos were obtained from patients who chose to be subjected to an IVF procedure and voluntarily gave their consents for providing the samples for these studies. This study was carefully reviewed and approved by the Reproductive Study Ethics Committee at Peking University Third Hospital (research license 2014SZ001) and Harvard University Committee of the Use of Human Subjects (CUHS).

Single cell whole genome amplification with MALBAC. Whole genome amplification of lysed cells was performed using the method Multiple Annealing and Looping Based Amplification

Cycles (MALBAC)(17) following the standard protocol provided by the commercial MALBAC amplification kit (Yikon Genomics Inc., Taizhou, China). Briefly, the cell was lysed by heating (3 hour at 50°C, and 10 min at 80°C) in 5 μL lysis buffer. Then 30 μl freshly-prepared pre-amplification Mix was added to each tube and incubated at 94°C for 3 min. Then DNA was amplified using 8 cycles of 40s at 20°C, 40s at 30°C, 30s at 40°C, 30s at 50°C, 30s at 60°C, 4 min. at 70°C, 20s at 95°C, 10s at 58°C, and immediately put on ice. We then added 30 uL Amplification Reaction Mix to each tube and incubated at 94°C for 30s, then 17 cycles of 20 seconds at 94°C, 30 seconds at 58°C and 3min at 72°C. For blood samples, we extracted the gDNA by using the QIAamp DNA Blood Mini Kit (Qiagen Inc.) and amplified 1ng gDNA using the MALBAC kit (Yikon Genomics Inc, Taizhou, China). For comparison with the MDA method, we used REPLI-g Single Cell kit (Qiagen, Inc.) to amplify the two embryos in Case I. According to a recent review, MALBAC offers the highest accuracy for CNV detection and the lowest allele dropout rate, therefore less false negatives (14).

MARSALA simultaneously detects aneuploidy and a targeted mutation. Whole genome sequencing of MALBAC products can identify chromosome abnormalities efficiently at low sequencing depths (<0.1X). However, precise detection of inherited point mutations would require higher sequencing depth (>10X), which significantly increases the cost. We developed a simple and inexpensive method to detect single mutations by amplifying the MALBAC products using PCR primers specific to the mutated allele (see Table S5). To do so, we incubated 8 ng of the MALBAC product from a single cell, using PCR primers specific to the mutated allele, at 98°C for 30s, 32 X cycles of 15s at 98°C, 30s at 60°C, 30s at 72°C, and additional 2 min. at 72°C. This PCR product was mixed with the MALBAC product (1-5% of MALBAC product), and used to construct a sequencing library using the NEBNext Ultra DNA library Prep kit (New England Biolabs, Inc). The sequencing was done by Illumina HiSeq 2500. Following this procedure, targeted point mutations and aneuploidy can be simultaneously detected in one NGS run with only 0.1x sequencing depth, although 2x was used for linkage analysis (see below).

Linkage analyses with MARSALA. To identify homozygous SNP positions adjacent to the target heterozygous point mutation of the patient, we sequenced the genomes of the parents, as well as the genome from a relative carrier of the monogenic disease allele. Then, from the sequences of each embryo (at blastocyst stage or the first and second polar bodies), the SNP readouts (heterozygous or homozygous) at these adjacent positions allowed the identification of the diseased allele in the embryo (Figure 2D, 3D, 4C). We noted that this method worked well and reproducibly with just 2X sequencing depth for each MALBAC amplified DNA sample involved, reducing the high cost of multiple sequencing rounds. Samples that had two rounds of

MALBAC decreased potential background noises to <5% (Fig.3B).

Comparison of blastocyst and polar body biopsies. In our second case, that involves a mutation in the EDA1 gene, the monogenic disease is maternal. For this female patient, in addition to analyzing biopsied cells at the blastocyst stage, we also biopsied the first (PB1) and second polar bodies (PB2) at 1 and 9 h after ICSI (10) in order to compare these results with those obtained from the blastocyst biopsies. The biopsied PBs were transferred to the lysis buffer as described in Hou et al. (10) for whole genome amplification and NGS, followed by MARSALA analyses using the genomes of the couple and their affected son.

Acknowledgments

Supported by grants from the National Natural Science of China (31230047, 31322037, 31440063, 31522034 to Drs. Qiao, Tang, Yan), National High Technology Research and Development Program (2015AA020407 to Drs. Yan and Huang), the Beijing Municipal Science and Technology Commission (Z131100005213006, CBXM2015-036 to Drs. Qiao, Tang, Xie), Peking University grant for translational research (to Drs. Xie, Qiao, Tang), and the National Basic Research Program of China (2011CB944500, 2012CB966704 to Drs. Qiao and Tang). The authors would like to acknowledge Dr. P. Purcell (Harvard University) and Dr. C. Racowsky (Brigham & Women's Hospital) for their critical reading and editing of the manuscript, the staff at the Department of Obstetrics and Gynecology in Peking University Third Hospital for assistance during the patient's pregnancy, the staff of the BIOPIC sequencing facility at Peking University, particularly Y. Zhang, for assistance.

Author Contributions

X.S.X., J.Q., F.T., S.L, L.Y., L. H. designed research; J.Q., L.P., R.L. were responsible for clinical work, including IVF treatment, oocyte collection,embryo transfer, embryo and fetus development monitoring, the baby follow-up; Y.L., J.H., and L.Y. performed blastocyst and polar bodies biopsy, ICSI, embryo culture, freezing and thawing; L.H., L. X. , Y. T., and L. Y amplified the whole genome and prepared sequencing libraries; L. X., L.Y., L.H., J.H., M.L. confirmed PGD results by Sanger sequencing, CGH array and STR analyses; X.Z., L.Y., L.X., and J.H.validated the family genetic information, amniotic fluid cell and umbilical cord blood analysis. F.M., L.Y., L.H., L.X., S.L., X. Z., and J.H., analyzed data; L.Y., L.H., L.X., S.L., F.T., J.Q., and X.S.X. wrote the paper.

Conflict of Interest

S.L. and X.S.X. are cofounders and shareholders of Yikon Genomics. Others declare no conflict of interest.

References

1. Boycott KM, Vanstone MR, Bulman DE, & MacKenzie AE (2013) Rare-Disease Genetics in the Era of Next-Generation Sequencing: Discovery to Translation.*Nature Reviews Genetics*. **14,** 681-91.
2. Yan L, Wei Y, Huang J, Zhu X, & Shi X*, et al.* (2014) Advances in Preimplantation Genetic Diagnosis/Screening.*Science China-Life Sciences*. **57,** 665-71.
3. Rubio C, Rodrigo L, Mir P, Mateu E, Peinado V, Milan M, Al-Asmar N, Campos-Galindo I, Garcia S, & Simon C (2013) Use of Array Comparative Genomic Hybridization (array-CGH) for Embryo Assessment: Clinical Results.*Fertility and Sterility*. **99,** 1044-8.
4. Handyside AH, Kontogianni EH, Hardy K, & Winston RM (1990) Pregnancies From Biopsied Human Preimplantation Embryos Sexed by Y-specific DNA Amplification.*Nature*. **344,** 768-70.
5. Yin X, Tan K, Vajta G, Jiang H, & Tan Y*, et al.* (2013) Massively Parallel Sequencing for Chromosomal Abnormality Testing in Trophectoderm Cells of Human Blastocysts.*Biology of Reproduction*. **88,** 69.
6. Wang L, Cram DS, Shen J, Wang X, & Zhang J*, et al.* (2014) Validation of Copy Number Variation Sequencing for Detecting Chromosome Imbalances in Human Preimplantation Embryos.*Biology of Reproduction*. **91,** 37.
7. Tobler KJ, Brezina PR, Benner AT, Du L, Xu X, & Kearns WG (2014) Two Different Microarray Technologies for Preimplantation Genetic Diagnosis and Screening, Due to Reciprocal Translocation Imbalances, Demonstrate Equivalent Euploidy and Clinical Pregnancy Rates.*J Assist Reprod Genet*. **31,** 843-50.
8. Treff NR, Tao X, Ferry KM, Su J, Taylor D, & Scott RJ (2012) Development and Validation of an Accurate Quantitative Real-Time Polymerase Chain Reaction-Based Assay for Human Blastocyst Comprehensive Chromosomal Aneuploidy Screening.*Fertility and Sterility*. **97,** 819-24.
9. Wells D, Kaur K, Grifo J, Glassner M, Taylor JC, Fragouli E, & Munne S (2014) Clinical Utilisation of a Rapid Low-Pass Whole Genome Sequencing Technique for the Diagnosis of Aneuploidy in Human Embryos Prior to Implantation.*Journal of Medical Genetics*. **51,** 553-62.
10. Hou Y, Fan W, Yan L, Li R, & Lian Y*, et al.* (2013) Genome Analyses of Single Human Oocytes.*Cell*. **155,** 1492-506.
11. Fiorentino F (2012) Molecular Genetic Analysis of Single Cells.*Seminars in Reproductive Medicine*. **30,** 267-82.

12. Treff NR, Fedick A, Tao X, Devkota B, Taylor D, & Scott RJ (2013) Evaluation of Targeted Next-Generation Sequencing-Based Preimplantation Genetic Diagnosis of Monogenic Disease.*Fertility and Sterility*. **99,** 1377-1384.e6.
13. Peters BA, Kermani BG, Alferov O, Agarwal MR, & McElwain MA*, et al.* (2015) Detection and Phasing of Single Base De Novo Mutations in Biopsies From Human in Vitro Fertilized Embryos by Advanced Whole-Genome Sequencing.*Genome Research*. **25,** 426-34.
14. Huang L, Ma F, Chapman A, Lu S, & Xie XS (2015) Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications.*Annu Rev Genomics Hum Genet*. **16,** 79-102.
15. Navin N, Kendall J, Troge J, Andrews P, & Rodgers L*, et al.* (2011) Tumour Evolution Inferred by Single-Cell Sequencing.*Nature*. **472,** 90-4.
16. Dean FB, Hosono S, Fang L, Wu X, & Faruqi AF*, et al.* (2002) Comprehensive Human Genome Amplification Using Multiple Displacement Amplification.*Proc Natl Acad Sci U S A*. **99,** 5261-6.
17. Zong C, Lu S, Chapman AR, & Xie XS (2012) Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell.*Science*. **338,** 1622-6.
18. Dreesen J, Destouni A, Kourlaba G, Degn B, & Mette WC*, et al.* (2014) Evaluation of PCR-based Preimplantation Genetic Diagnosis Applied to Monogenic Diseases: A Collaborative ESHRE PGD Consortium Study.*European Journal of Human Genetics*. **22,** 1012-8.
19. Wilton L, Thornhill A, Traeger-Synodinos J, Sermon KD, & Harper JC (2009) The Causes of Misdiagnosis and Adverse Outcomes in PGD.*Human Reproduction*. **24,** 1221-8.
20. Thornhill AR, Handyside AH, Ottolini C, Natesan SA, & Taylor J*, et al.* (2015) Karyomapping-A Comprehensive Means of Simultaneous Monogenic and Cytogenetic PGD: Comparison with Standard Approaches in Real Time for Marfan Syndrome.*J Assist Reprod Genet*. **32,** 347-56.
21. Natesan SA, Handyside AH, Thornhill AR, Ottolini CS, Sage K, Summers MC, Konstantinidis M, Wells D, & Griffin DK (2014) Live Birth After PGD with Confirmation by a Comprehensive Approach (Karyomapping) for Simultaneous Detection of Monogenic and Chromosomal Disorders.*Reproductive Biomedicine Online*. **29,** 600-5.
22. Natesan SA, Bladon AJ, Coskun S, Qubbaj W, & Prates R*, et al.* (2014) Genome-Wide Karyomapping Accurately Identifies the Inheritance of Single-Gene Defects in Human Preimplantation Embryos in Vitro.*Genetics in Medicine*. **16,** 838-45.
23. Konstantinidis M, Prates R, Goodall NN, Fischer J, & Tecson V*, et al.* (2015) Live Births Following Karyomapping of Human Blastocysts: Experience From Clinical Application of the Method.*Reproductive Biomedicine Online*. **31,** 394-403.
24. Handyside AH (2015) Live Births Following Karyomapping - a "Key" Milestone in the

Development of Preimplantation Genetic Diagnosis.*Reproductive Biomedicine Online*. **31,** 307-8.
25. Shen J, Cram DS, Wu W, Cai L, Yang X, Sun X, Cui Y, & Liu J (2013) Successful PGD for Late Infantile Neuronal Ceroid Lipofuscinosis Achieved by Combined Chromosome and TPP1 Gene Analysis.*Reproductive Biomedicine Online*. **27,** 176-83.
26. Daina G, Ramos L, Obradors A, Rius M, Martinez-Pasarell O, Polo A, Del RJ, Obradors J, Benet J, & Navarro J (2013) First Successful Double-Factor PGD for Lynch Syndrome: Monogenic Analysis and Comprehensive Aneuploidy Screening.*Clinical Genetics*. **84,** 70-3.
27. Rechitsky S, Pakhalchuk T, San RG, Goodman A, Zlatopolsky Z, & Kuliev A (2015) First Systematic Experience of Preimplantation Genetic Diagnosis for Single-Gene Disorders, And/Or Preimplantation Human Leukocyte Antigen Typing, Combined with 24-Chromosome Aneuploidy Testing.*Fertility and Sterility*. **103,** 503-12.
28. Kuliev A & Rechitsky S (2011) Polar Body-Based Preimplantation Genetic Diagnosis for Mendelian Disorders.*Molecular Human Reproduction*. **17,** 275-85.
29. Seki H, Kubota T, Ikegawa S, Haga N, Fujioka F, Ohzeki S, Wakui K, Yoshikawa H, Takaoka K, & Fukushima Y (2001) Mutation Frequencies of EXT1 and EXT2 in 43 Japanese Families with Hereditary Multiple Exostoses.*Am J Med Genet*. **99,** 59-62.
30. Monreal AW, Zonana J & Ferguson B (1998) Identification of a New Splice Form of the EDA1 Gene Permits Detection of Nearly All X-linked Hypohidrotic Ectodermal Dysplasia Mutations.*American Journal of Human Genetics*. **63,** 380-9.
31. Lu S, Zong C, Fan W, Yang M, & Li J*, et al.* (2012) Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing.*Science*. **338,** 1627-30.

Figure Legends

Fig 1. Workflow of MARSALA

Fig 2. MARSALA analyses of blastocyst sequencing for Case 1 with the husband affected by Hereditary Multiple Exostoses (HME)

A. Pedigrees of the couple's family with HME. Blackened symbols represent affected individuals, whereas open symbols represent unaffected individuals. Circles and squares indicate females and males, respectively. The arrow indicates the affected parent. Individuals indicated by asterisks were verified by Sanger sequencing.

B. Sequencing result of the targeted mutation site in the EXT2 gene of the husband (H) and wife (W) as well as their eighteen embryos (E1- E18). Embryos E05, 06, 09, 10, 11, 12, and 16 had the disease-carrying mutated allele (red) to be excluded for transfer. In green, the fraction of reads of the covered region is shown, consistent with the reference genome.

C. CNVs of the 18 embryos at low sequencing depth (0.1X) of NGS. Significant chromosomal abnormality was identified in Embryos E06 (Monosomy 18, deletion in part of chromosome 6), E07 (Monosomy 22), E09 (Trisomy 12), E12 (Monosomy 8), E13 (Trisomy 8, 15, deletion or duplication in part of chromosome 9, 10, 18), E14 (deletion or duplication in part of chromosome 3), E16 (deletion in part of chromosome 4), E18 (Trisomy 22), hence not suitable for transfer.

D. Schematic representation of MARSALA for disease-carrying alleles. We sequenced the amplified genomes from each embryo, from the couple and from the husband's father. The gene loci heterozygous for the husband (e.g. A/G) but homozygous for the husband's father (e.g. A/A) and the wife (e.g A/A), are shown within a range of 1 Mb upstream or downstream from the mutated site. These SNPs are sufficient for deducing whether each embryo carries the disease or normal allele from the husband (see below). Even if the SNV is not detected directly in the low coverage sequencing; the disease-carrying allele can still be identified with certainty. Purple "X*" indicates the deletion of a single base in the mutated allele (frame-shift mutation). Purple "C" indicates the base associated with the monogenic disease in the wild-type allele. "A, G," indicate the SNVs linked to the mutated allele.

E. Linkage analyses results of the 18 embryos. Ten SNV markers were selected for the disease-carrying allele identification for each embryo. Seven embryos (E05, E06, E09, E10, E11, E12, E16) were identified to carry the disease-associated allele by these ten markers with high certainty. Abbreviations: RSID: Reference SNP cluster ID, CHROM: chromosome number, POS: genomic location, REF: the SNPs of reference allele, ALT: the SNPs of alternative allele, FH: husband's father, H: husband, W: wife. "-" represents the alleles that are not covered by single-cell low-depth sequencing.

Fig 3. MARSALA analyses of blastocyst sequencing for Case 2 with the wife carrying a disease causing SNV in the Ectodysplasin A1 (EDA1) gene

A. Pedigrees of the couple's family with EDA1 point mutation. Blackened symbols represent affected individuals, whereas open symbols represent unaffected individuals. Circles and squares indicate females and males, respectively. The arrow indicates the affected proband. Individuals indicated by asterisks were verified by Sanger sequencing.

B. Sequencing result of the targeted mutation site in the EDA1 gene of the husband (H), wife (W) and affected child (C) as well as the four embryos. Embryos E01 and E03 had the disease-carrying mutated allele (red) needed to be excluded for transfer. Green indicates the fraction of reads of the covered region, consistent with the reference genome.

C. CNVs of the four embryos at low sequencing depth (0.1X) of NGS. Significant chromosomal abnormality was identified in Embryo E03 (Monosomy 4), hence not suitable for transfer.

D. Schematic representation of MARSALA for disease-carrying alleles. We sequenced the amplified genomes from each embryo, and the bulk genomes of the couple and their affected son. The gene loci heterozygous for the wife (e.g. T/G) are shown within the range of 3 Mb upstream or downstream from the disease causing mutation site. These SNPs are sufficient for deducing whether each embryo carries the mutated or the normal allele from the wife. Even if the SNV is not detected in the low coverage sequencing; the disease-carrying allele can still be identified with certainty. Purple "G*" indicates the mutated allele (single base substitution). Purple "C" indicates the wild-type allele. "A, G, T" indicate the SNVs linked to the disease-associated allele.

E. Ten SNV markers were selected for the disease-carrying allele identification for each embryo. Two embryos (E01, E03) were identified to carry the disease-associated allele with high certainty. Abbreviations: RSID: Reference SNP cluster ID, CHROM: chromosome number, POS: genomic location, REF: the SNPs of reference allele, ALT: the SNPs of alternative allele, S: affected son, H: husband, W: wife. "-" represents the alleles that are not covered by single-cell low-depth sequencing.

Fig 4. MARSALA analyses of polar body sequencing for Case 2

A. Sequencing result of the targeted mutation site in the EDA1 gene using polar body biopsies of two embryos E03 and E05. The deduction from E03 polar bodies' SNVs is consistent with the affected SNV result obtained from the E03's blastocyst biopsy. However, the E05 polar body result allows for the deduction of no affected SNV in E05, which was not obtainable from the E05 blastocyst as the WGA of the E05 biopsy failed.

B. CNV patterns of PB1 and PB2 of embryo E03 at low sequencing depth (0.1X) of NGS, from which the CNV of the female pronucleus is deduced as normal. In contrast, the direct measured CNV pattern of E03 blastocyst biopsy showed an abnormal Chromosome 4 (compare Fig. 4B vs Fig. 3C). This can be explained by the aneuploidy of the sperm, consistent with the fact that Chromosome 4 of this blastocyst only exhibited the maternal rather than the paternal SNVs.

C. Schematic representation of MARSALA for linkage analysis to identify the existence of disease-carrying alleles in female pronucleus in order to avoid false positive and false negative SNVs in the deduction of the existence of disease allele in the female pronucleus.