Title

Validation of the first PGT-whole genome sequencing approach including mitochondrial variants

Yuntao Xia[1]*,Dhruva Chandramohan[1], Willy Chertman[1], Elan Bechor[1], Robert Maier[1], Noor Siddiqui[1], Barry Behr[2], Justin Schleede[1]

Orchid Health, Palo Alto, CA, 94301[1]

Department of Obstetrics & Gynecology - Reproductive Endocrinology and Infertility, Stanford University, Sunnyvale, CA, 94087[2]

*First and corresponding author

Abstract

Whole genome screening is currently not available in preimplantation genetic testing (PGT) due to poor performance of whole genome amplification methods. Here, we present the first validation study using our latest whole genome amplification (WGA) protocol, which yields amplified DNA with comparable quality to genomic DNA when perfoming whole genome sequencing assays. We started by validating PGT for aneuploidy (PGT-A) using our WGA protocol on cell lines and donated human embryos, where amplification success rates were >99.9% and 96.2% respectively. Accuracy on both was >99.9% on aneuploidy status. We next validated whole genome sequencing using amplified and genomic DNA from the Genome in the Bottle reference sample NA12878 to demonstrate accuracy, specificity, sensitivity, and precision of our WGA methods. Amplified DNA and genomic DNA are comparable in this case, with 99.99% accuracy, 99.99% specificity, 98.0% sensitivity and 98.1% precision with our WGA protocol. We additionally examined variant calls between biopsies and the remaining embryos from which they were derived. Again, 99.99% accuracy, 99.99% specificity, 98.1% sensitivity and 98.0% precision were observed. Mitochondrial heteroplasmy between biopsies and embryos was validated, and 99.9% accuracy, 100% specificity, 99.1% sensitivity and 100% precision were observed. Finally, we demonstrated the ability to improve the accuracy of certain types of mendelian variants in our data to close to 100% by leveraging parental and a born child's genome to perform linkage analysis.

Introduction

Preimplantation genetic testing (PGT) has shortened time to a successful pregnancy by improving fetal viability rates per embryo transfer (Vermeesch et al., 2016). However, current methods can only assess ploidy status or screen for specific mendelian variants for which parents are known to be carriers. Whole genome screening cannot be performed at high quality with current methods due to poor coverage and high allelic dropout rates (Zheng et al., 2011). In this whitepaper, we present a validation study on our latest whole-genome amplification (WGA) method, which yields amplified DNA comparable to genomic DNA in ploidy analysis and whole genome sequencing (WGS).

We validated PGT-A first including mosaicism using our WGA protocol. Then, we validated our WGS results by comparing amplified DNA and genomic DNA from NA12878 to its benchmark calls of the Genome in a Bottle from NIST (National Institute of Standards and Technology). We focused on critical quality parameters: genomic coverage, accuracy, specificity, sensitivity, and precision. Next, using donated embryos that would not undergo implantation, we compared results from WGS of embryo biopsies versus the whole embryo in terms of the same quality parameters, demonstrating high concordance. If genomes of parents and a born child were obtained at the time of screening, our accuracy of genome-wide mendelian screening would be further improved through linkage analysis. Mitochondrial validation was performed in parallel by assessing concordance between mitochondrial variants or heteroplasmy from embryo biopsies versus the whole embryos.

Results

Validation on aneuploidy screening via cell lines and human embryos

Ploidy analysis was initially validated by assessing the accuracy of sample-level aneuploidy detection.

A total of 73 samples from 8 cell lines were used first to validate our aneuploidy screening (Figure 1A, 1B). These cell lines covered common karyotypes seen in PGT-A, including euploid males and females and autosomal and sex chromosome aneuploidy. We diluted the DNA to ~50 pg (picogram), which is equivalent to the amount of DNA in embryo trophectoderm biopsies. Whole-genome amplification was then performed, followed by low pass sequencing (~0.01 - 0.05X) for aneuploidy evaluation with a bin size of 500kbp on the genome in the analysis. Overall, we achieved 100% successful amplification and 100% accuracy with the provided sample binary aneuploidy status (Figure 1B), as well as the chromosome level calls.

Next, we extended our PGT-A validation on donated human embryos with a larger variety of karyotypes (Figure2, Table S1). We obtained 2-3 samples per embryo, on which we performed WGA and ploidy analysis. We first evaluated the consistency of samples from the same embryo and then compared the CNV results to the clinical PGT-A reports. So far, we observed >99.9% concordance among samples from the same embryo (Figure 2B, Table S1). However, when we compare our CNV results to the clinical PGT-A reports, there was an inconsistency on embryo #20. We observed a loss of Chr5 in our samples whereas the clinical PGT-A report also showed a gain of Chr18, in addition to the loss on Chr5 (Table S1, Figure S1). To rule out mosaicism, we sequenced the rest of the embryo after taking biopsies and still observed no indication of trisomy on Chr18 (Figure S1). Considering clinical PGT-A is not 100% accurate and inconsistency has been previously reported in multiple case studies (Gleicher et al., 2021), we conclude that this inconsistency was likely not a failure of our WGA method. However, we agreed that embryo #20 is aneuploidy, which is consistent with the ploidy status on that PGT-A report. Therefore, we achieved >99.9% accuracy on aneuploidy status and 96.2% successful amplification, and 99.5% accuracy when we perform chromosome-to-chromosome comparison, though we believe our call on Chr18, which differs from the clinical result, may actually be correct.

Low-level mosaicism in embryos has been reported to be compatible with a successful pregnancy and live birth, especially when the aneuploidy mosaicism level is less than 40% (Abhari and Kawwass, 2021). Therefore, our algorithm was designed to capture mosaicism that is above 30%. We performed a dilution

series from 50% to 10% mosaicism, generated by mixing aneuploid (47,XY, +18) and euploid (46,XX) DNA. As seen in Fig3B-Batch1, mosaic samples above 30% could be visualized and were captured successfully by algorithm. To confirm the results, we performed another mosaicism dilution using a mix of NA09288 (47,XY,+9) and euploid (46,XX) DNA and found the same result (Figure S2A, 3B-Batch2). Additionally, 24 cell line mixtures with 30-35% aneuploidy mosaicism were tested, and 100% of them were successfully marked during analysis (Figure 3B-Batch3).

Validation of whole genome sequencing using cell lines and human embryos

Although multiple clinical labs have demonstrated good performance in PGT-A (Treff et al., 2019), none have attempted validation on PGT- whole genome sequencing, to our knowledge. DNA derived from WGA to date has been lower quality compared to genomic DNA, and so has not been used for embryo screening under clinical conditions (Huang et al., 2015). Our results show that using our WGA protocol, WGS using amplified DNA can achieve comparable results to genomic DNA.

We started with the genome in the bottle – NA12878, and the variants of NA12878 from NIST were used as the reference in data analysis. Both genomic DNA and amplified DNA from NA12878 were analyzed and compared to the benchmark calls from GIAB (NISTv4.2.1). As shown in Table 1 and Table S2, accuracy and specificity at whole genome level are 99.99% for both amplified DNA and genomic DNA. We observed an average of 99.8% genomic coverage in amplified DNA and the same number in gDNA (Table 1). The precision of amplified DNA and gDNA are 98.1% and 97.8%, respectively, suggesting our amplified DNA does not generate more false positive calls than gDNA (Table 1). Moreover, sensitivity reaches 98.0% in amplified DNA compared to 98.9% in gDNA, indicating false negative calls in data should be similar in magnitude as well (Table 1). Therefore, our WGS on amplified DNA was validated against the variant list from NIST and its sequencing quality is comparable to gDNA.

To validate performance on biopsies of clinical embryos, we compared the results of the biopsies to their corresponding embryos. Using the whole embryo as the reference, we observed an average of 99.6% genomic coverage, 99.9% accuracy, 99.9% specificity, 98.0% precision, 98.1% sensitivity (Table 2 & Table S3). Such concordance allows us to use WGS of an embryo biopsy to reflect the genome in the embryo.

We also performed WGS on biopsies from embryos that previously underwent PGT-M. Two variants were screened during PGT-M, TGM1 24728950C>T and 24718694C>T as one parent carries 24728950C>T and the other carries 24718694C>T. Based on our WGS data, both positions, with an average of reading depth of 28, exhibit correct variant calls compared to the clinical PGT-M report (Table 3).

Validation on inherited variants using linkage analysis

If genomes of the parents and a born sibling are obtained, linkage analysis can be applied to further improve accuracy. Amplified DNA from the son (NA24385, HG002) of the Ashkenazim trio dataset from GIAB was used for this validation, along with the benchmark callset for the entire trio (NISTv4.2.1) (Wagner et al., 2020). Linkage analysis focuses on large chunks of inherited DNA segments rather than single variants. Therefore, even if a specific variant is not captured in sequencing data, it can be inferred through the surrounding genome. A total of 2.82 million SNVs was validated under 6 most common

categories: Parents are heterozygous carriers and born sibling is either homozygous REF or homozygous ALT; One of the parents is heterozygous carrier, the other parent is homozygous REF, and the born sibling is either homozygous REF or heterozygous carrier; One of the parents is homozygous ALT or affected, the other parent is heterozygous carrier, and the born sibling is either homozygous ALT or heterozygous carrier. For each variant, we focused on a 2 Mbp window on WGS data around the variant and compared direct WGS calling to the combination of WGA and linkage analysis. While WGS alone generates 99.4% accuracy on the 2.82 million SNPs, combining it with linkage analysis improves the accuracy to 99.9% (Table 4).

Validation on mitochondrial DNA

High fraction of pathogenic heteroplasmy has been correlated to mitochondrial diseases including early onset metabolic and degenerative diseases. Our validation on mitochondrial DNA focused on mitochondrial genome coverage, sensitivity, and precision by comparing biopsies to their embryos. Setting 15% to be our reportable heteroplasmy threshold, we observed an average of 100% mitochondria DNA coverage, >99.9% accuracy, >99.9% specificity, 99.0% sensitivity and >99.9% precision (Table 5). Mitochondrial DNA is known to be exclusively inherited from the mother. We have already confirmed it using our amplified DNA in previous studies (Xia et al., 2021).

Discussion

Here, our aim is to provide the first combined validation including PGT-A, PGT-whole genome sequencing, linkage analysis, and mitochondrial DNA using our amplified DNA technology. This analysis demonstrates that we are able to achieve equivalent DNA sequencing coverage, accuracy, sensitivity, specificity, and precision as those obtained from unamplified genomic DNA. In addition, we have shown our analysis provides superior mitochondrial coverage compared to what has been reported with other WGS assays (Sallevelt et al., 2013). Consequently this technique will have significant impact on the ability to screen embryos preimplantation for aneuploidy, familial disease, and deleterious mutations increasing the likelihood of a successful pregnancy while simultaneously reducing the likelihood of abnormal live born fetus. Polygenic risk scores are also achievable with our method due to high genomic coverage and accuracy in our data. Some polygenic risk scores utilize 6 - 7 million SNPs (Khera et al., 2018), which is difficult to achieve on SNP array chips without heavy imputation.

Since our approach uses next-generation sequencing, all the same limitations apply. These include the ~1% of the genome that goes uncovered and known challenging regions such as centromeres. Although sensitivity and specificity are high there remains a small chance for false positives and false negatives calls. In addition, there are many regions of the genome where obtaining accurate sequence data is difficult which include segmental duplications, LINEs, SINEs, trinucleotide and other repeat expansions, pseudogenes, etc. Accurate copy number analysis is dependent on the depth of sequencing, regions associated with highly variable or absent coverage may not be suitable for copy number analysis. Finally, our WGA protocol is optimized for intact DNA, so performance is expected to suffer if the biopsied cells are dying and undergoing DNA fragmentation.

Methods

Sample collection

Cell lines were obtained from Coriell. Embryo samples were all donated to research and provided from the HRC Fertility Center (wIRB #20215134). All embryos were created using Intracytoplasmic sperm injection (ICSI). Embryos from day 5/6 were biopsied following standard SOP. Each biopsy contained approximately 5 cells and was collected in a 200ul PCR tube with 3ul of cell buffer from BioSkryb Genomics.

Library preparation and sequencing

DNA sizes after WGA were first confirmed by running 1-2% Agarose E-Gel (Invitrogen). Then, 500ng of DNA was used for library preparation with KAPA HyperPlus kit per manufacturer's instructions. Dual Index UMI adapters (Integrated DNA Technologies) were used in the ligation. Library concentration was quantified using the Qubit 4 dsDNA HS. Library sizes were measured through Agilent 4150 Tapestation D1000 ScreenTape assay (Agilent Technologies). Sequencing runs were performed on a MiniSeq for aneuploidy screening and NovaSeq6000 for WGS per manufacturer's instructions.

Genomic Data Analyses

We followed GATK4 (v4.2.0.0, java v1.8.0_231) best practices with genome assembly of GRCh37. In brief, fastq files were aligned with bwa (0.7.17), and then the corresponding bam files were processed with base quality score recalibration (BQSR) and duplicate marking. Gvcfs were generated for each sample and jointly called as part of a larger cohort. GATK VariantRecalibrator (VQSR) was run on the resulting callset. Mitochondrial DNA was analyzed using GATK without VQSR. CollectWgsMetrics and CollectVariantCallingMetrics were run on the final bams and vcfs respectively to generate the reported QC metrics over the NA12878 benchmark regions to enable cross-sample comparison. Sensitivity, precision, etc were calculated through RTG packages. Reproducibility and PGT-M results were obtained from joined-called vcf files.

Aneuploidy Screening

Ginkgo was used for CNV calling using default parameters as the package is designed for amplified DNA. All data were aligned to GRCh37 with a bin size of 500kbp.

Linkage analysis

The individual benchmark regions (v4.2.1) for the Ashkenazim GIAB trio were intersected to establish a consensus set of regions for linkage assessment. The benchmark calls for the trio were then merged using bcftools (v1.14) with the '-missing-to-ref' option, then merged with calls made on an amplified sample of the offspring NA24385 (HG002). For each of the site for which a call was made on the amplified sample, flanking informative SNPs in a 1 megabase window were analyzed on each side to establish the parental haplotypes where possible. Accuracy of the initial variant calls and the linkage corrected calls was assessed against the benchmark HG002 calls over the consensus.

Author Contributions

YX, JS designed experiments. BB, NS, RM contributed key materials, methods, and discussion. YX, DC, EB performed and analyzed experiments. YX, WC, JS wrote the manuscript.

Conflict of Interests

DC, WC, EB, JS, RM, NS and YX are members of Orchid Health and BB is a scientific advisor to Orchid.

Acknowledgements

We would like to acknowledge HRC Fertility for their assistance in the study. We would like to thank Psomagen (CAP & CLIA) for the sequencing service.

Legends

Figure 1. A) Visualized copy number results of Coriell cell line samples. Expected karyotypes were observed with our WGA protocol. B) Statistics of 73 tests using Coriell cell line samples. Aneuploidy status accuracy indicates binary normal vs aneuploidy results. Chromosome-to-Chromosome accuracy measures whether we achieved accurate calling on every chromosome.

Figure 2. A) Visualized copy number results of embryo samples. Consistency was observed on samples from the same embryo. B) Statistics of all tests using samples from donated embryos. We not only measured consistency on samples from the same embryo, but also compared our copy number results to clinical PGT-A reports.

Figure 3. A) Mosaicism was captured at 30% level using a mix of 47, XY+18 and euploid 46, XX. B) Percent of mosaicism samples captured by algorithm. When mosaicism was <30%, no samples were captured. When mosaicism was >= 30%, >99.9% samples were captured. Batch 1 has 10 samples from 10% to 50%. Batch 2 has 10 samples from 10% to 50%. Batch 3 has 24 samples from 30% to 35%.

Figure S1. Both biopsy samples from embryo #20 exhibited no trisomy signal on Chr18. The results on the embryo confirmed the result.

Figure S2. Mosaicism was captured at 30% level on mixtures of 47,XY+9 and 46,XX.

Table 1. Comparison of gDNA and amplified DNA from NA12878 in terms of genomic coverage, total SNV called, accuracy, specificity, precision, and sensitivity. Variants list of NA12878 from NIST was used as the reference.

Table 2. Comparison of biopsies and their embryos in terms of genomic coverage, total SNV called, accuracy, specificity, precision, and sensitivity. The whole embryos were used as references.

Table 3. Variants in PGT-M reports were captured in PGT-whole genome sequencing data.

Table 4. Linkage analysis improved overall accuracy of detection of mendalian variants in whole genome sequencing.

Table 5. Analysis of mitochondrial variants from PGT-whole genome sequencing data. Mitochondrial variants from the whole embryos were used as references.

Table S1. List of embryo samples tested in PGT-A validation.

Table S2. Detailed numbers for true positive, false positive, true negative and false negative SNVs from NA12878 samples. These numbers were used to compute Table 1.

Table S3. Detailed numbers for true positive, false positive, true negative and false negative SNVs from embryo biopsies. These numbers were used to compute Table 2.

Reference

Abhari, S., and J.F. Kawwass. 2021. Pregnancy and Neonatal Outcomes after Transfer of Mosaic Embryos: A Review. *J Clin Med*. 10.

Gleicher, N., P. Patrizio, and A. Brivanlou. 2021. Preimplantation Genetic Testing for Aneuploidy - a Castle Built on Sand. *Trends Mol Med*. 27:731-742.

Huang, L., F. Ma, A. Chapman, S. Lu, and X.S. Xie. 2015. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annu Rev Genomics Hum Genet*. 16:79-102.

Khera, A.V., M. Chaffin, K.G. Aragam, M.E. Haas, C. Roselli, S.H. Choi, P. Natarajan, E.S. Lander, S.A. Lubitz, P.T. Ellinor, and S. Kathiresan. 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 50:1219-1224.

Sallevelt, S.C., J.C. Dreesen, M. Drusedau, S. Spierts, E. Coonen, F.H. van Tienen, R.J. van Golde, I.F. de Coo, J.P. Geraedts, C.E. de Die-Smulders, and H.J. Smeets. 2013. Preimplantation genetic diagnosis in mitochondrial DNA disorders: challenge and success. *J Med Genet*. 50:125-132.

Treff, N.R., R. Zimmerman, E. Bechor, J. Hsu, B. Rana, J. Jensen, J. Li, A. Samoilenko, W. Mowrey, J. Van Alstine, M. Leondires, K. Miller, E. Paganetti, L. Lello, S. Avery, S. Hsu, and L.C.A. Melchior Tellier. 2019. Validation of concurrent preimplantation genetic testing for polygenic and monogenic disorders, structural rearrangements, and whole and segmental chromosome aneuploidy with a single universal platform. *Eur J Med Genet*. 62:103647.

Vermeesch, J.R., T. Voet, and K. Devriendt. 2016. Prenatal and pre-implantation genetic diagnosis. *Nat Rev Genet*. 17:643-656.

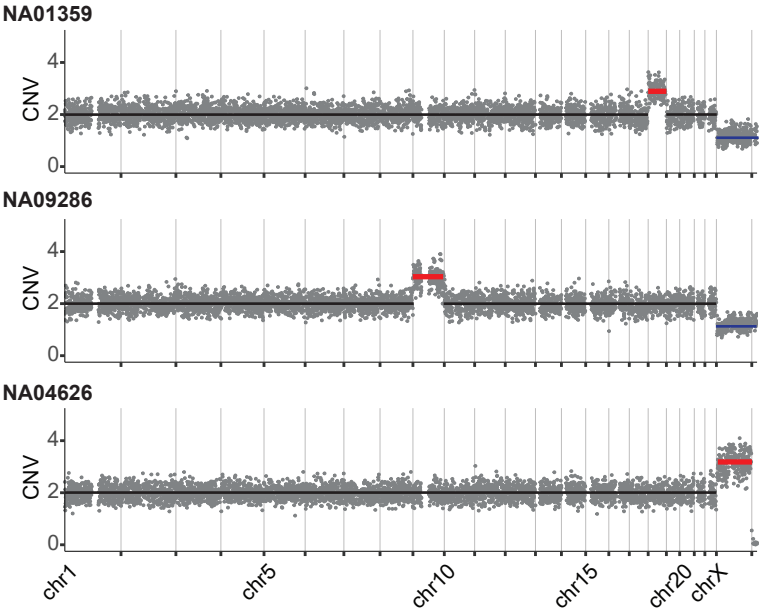
Wagner, J., N.D. Olson, L. Harris, Z. Khan, J. Farek, M. Mahmoud, A. Stankovic, V. Kovacevic, A.M. Wenger, and W.J. Rowell. 2020. Benchmarking challenging small variants with linked and long reads. *BioRxiv*.

Xia, Y., V. Gonzales-Pena, D.J. Klein, J.J. Luquette, L. Puzon, N. Siddiqui, V. Reddy, P. Park, B.R. Behr, and C. Gawad. 2021. Genome-wide Disease Screening in Early Human Embryos with Primary Template-Directed Amplification. *bioRxiv*.

Zheng, Y.M., N. Wang, L. Li, and F. Jin. 2011. Whole genome amplification in preimplantation genetic diagnosis. *J Zhejiang Univ Sci B*. 12:1-11.

Figure 1

A) **Aneuploidy screening on cell lines**
- Copy number binned every 500 kb

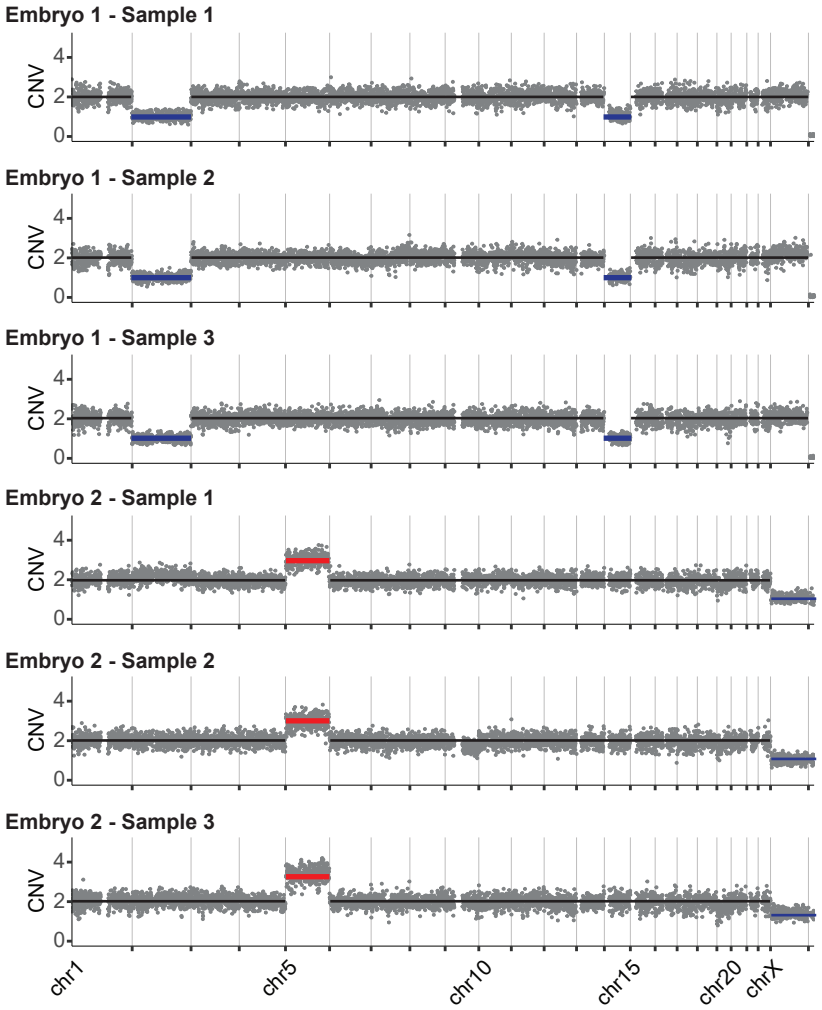


B)

Coriell Cell line ID	# tested	# with result	Coriell Karyotype	Aneuploidy Status Accuracy	Chr-to-Chr Accuracy
NA12878	15	15	46XX	100%	100%
NA24385	5	5	46XY	100%	100%
NA19240	7	7	46XX	100%	100%
NA01359	13	13	47XY,+18	100%	100%
NA09286	13	13	47XY,+9	100%	100%
NA04626	6	6	47XXX	100%	100%
NA24149	10	10	46XY	100%	100%
NA24146	10	10	46XX	100%	100%
Total	73	73		100%	100%

Figure 2

A) **Aneuploidy screening on donated embryo samples**
- Copy number binned every 500 kb

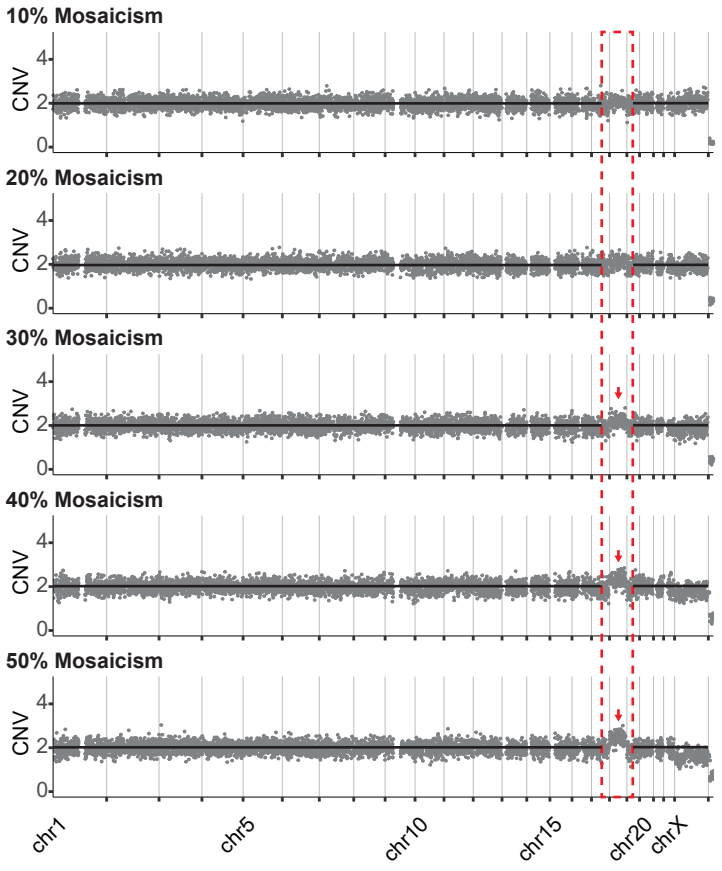


B)

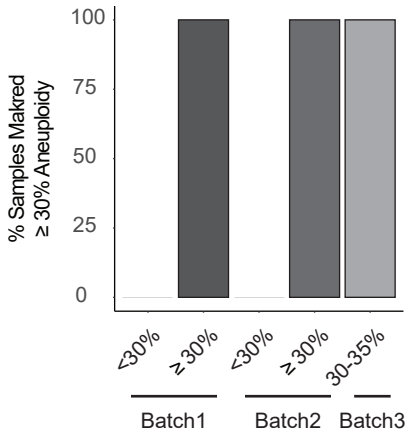
Embryo ID	# Tested	# with Result	Clinical PGT Report	Orchid Report	Sample Concordance from the Same Embryo	Aneuploidy Status Accuracy	Chr-to-Chr Accuracy
1	3	3	44XX,-2,-14	44XX,-2,-14	100%	100%	100%
2	3	3	47XY,+5	47XY,+5	100%	100%	100%
3	3	3	46XX	46XX	100%	100%	100%
4	3	3	45XX,-16	45XX,-16	100%	100%	100%
5	2	2	46XY,-18,+22	46XY,-18,+22	100%	100%	100%
...	...	...	...	...	...	...	...
(See. Suppl.)	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
	Total Cases	With Results	Amplification Success	Concordant with clinical	Sample Concordance from the Same Embryo	Aneuploidy Status Accuracy	Chr-to-Chr Accuracy
	104	100	96.2%	98	>99.9%	>99,9%	99.86%

Figure 3

A) Aneuploidy screening on mosaic samples
- Copy number binned every 500 kb



B)



Sample	Genomic Coverage	Total SNV	Accuracy	Specificity	Precision	Sensitivity
NA12878 gDNA Sample 1	99.8%	3286979	99.996%	99.997%	97.8%	98.8%
NA12878 gDNA Sample 2	99.9%	3290927	99.996%	99.997%	97.8%	99.0%
NA12878 gDNA Sample 3	99.8%	3286497	99.996%	99.997%	97.8%	98.8%
NA12878 gDNA Sample 4	99.9%	3287953	99.996%	99.997%	97.9%	99.0%
NA12878 gDNA Sample 5	99.8%	3277286	99.996%	99.997%	98.0%	98.8%
Average	**99.8%**	**3285928**	**99.996%**	**99.997%**	**97.8%**	**98.9%**
NA12878 Amplified DNA Sample 1	99.7%	3238978	99.994%	99.997%	97.9%	97.6%
NA12878 Amplified DNA Sample 2	99.8%	3263126	99.995%	99.997%	98.0%	98.4%
NA12878 Amplified DNA Sample 3	99.8%	3252094	99.995%	99.998%	98.2%	98.3%
NA12878 Amplified DNA Sample 4	99.8%	3258943	99.996%	99.998%	98.2%	98.4%
NA12878 Amplified DNA Sample 5	99.7%	3229761	99.994%	99.997%	97.9%	97.3%
Average	**99.8%**	**3248580**	**99.995%**	**99.997%**	**98.1%**	**98.0%**

Sample	Genomic Coverage	Total SNV	Accuracy	Specificity	Precision	Sensitivity
1	99.7%	3343804	99.995%	99.997%	97.8%	98.4%
2	99.6%	3327631	99.995%	99.997%	98.0%	98.1%
3	99.6%	3312139	99.995%	99.997%	97.9%	98.1%
4	99.4%	3279468	99.993%	99.997%	97.8%	97.1%
5	99.6%	3404155	99.995%	99.998%	98.4%	98.1%
6	99.7%	3414277	99.996%	99.998%	98.4%	98.4%
7	99.6%	3360682	99.995%	99.998%	98.4%	98.1%
8	99.6%	3163591	99.995%	99.998%	98.2%	98.2%
9	99.6%	3160113	99.995%	99.998%	98.3%	98.2%
10	99.6%	3304310	99.994%	99.996%	97.2%	98.0%
11	99.7%	3301716	99.994%	99.997%	97.5%	98.2%
12	99.7%	3368834	99.996%	99.998%	98.6%	98.2%
13	99.7%	3340483	99.995%	99.998%	98.8%	97.5%
14	99.6%	3270893	99.995%	99.997%	97.5%	98.4%
15	99.5%	3262566	99.995%	99.997%	97.6%	98.3%
16	99.6%	3274334	99.995%	99.997%	97.8%	98.2%
17	99.6%	3275227	99.995%	99.997%	97.7%	98.2%
18	99.6%	3254447	99.995%	99.998%	98.1%	98.4%
19	99.4%	3234717	99.995%	99.998%	98.2%	97.9%
20	99.7%	3274002	99.995%	99.997%	97.9%	98.5%
21	99.5%	3257361	99.995%	99.998%	98.1%	98.2%
22	99.6%	3270572	99.995%	99.998%	98.2%	97.9%
23	99.7%	3284833	99.995%	99.998%	98.2%	98.3%
24	99.6%	3237713	99.993%	99.997%	97.9%	97.0%
25	99.7%	3382545	99.996%	99.998%	98.3%	98.6%
Average	**99.6%**	**3294417**	**99.995%**	**99.997%**	**98.0%**	**98.1%**

Sample	Clinical PGT-M Report		WGS results - Genotypes	
	Chr14:24728950C>T (c.944G>A)	Chr14:24718694C>T (c.2279G>A)	Chr14:24728950	Chr14:24718694
Biopsy sample 1	Negative	Positive	0/0, C/C	1/0, C/T
Biopsy sample 2	Negative	Positive	0/0, C/C	1/0, C/T
Biopsy sample 3	Negative	Negative	0/0, C/C	0/0, C/C
Biopsy sample 4	Negative	Negative	0/0, C/C	0/0, C/C

Category	Parents' genotype		Born Child Genotype	Total SNVs	Accuracy in WGS	Accuracy in Linkage
Both parents are carriers	1/0	1/0	1/1	165036	99.5%	99.9%
Both parents are carriers	1/0	1/0	0/0	168443	99.9%	99.8%
One parent is carrier	1/0	0/0	1/0	943057	98.6%	99.9%
One parent is carrier	1/0	0/0	0/0	811420	99.9%	99.9%
One parent affected. The other is carrier.	1/1	1/0	1/1	366706	99.6%	99.9%
One parent affected. The other is carrier.	1/1	1/0	1/0	368507	98.6%	99.8%
				Average	99.4%	99.9%

Sample	Total Bases	True Positive	True Negative	False Positive	False Negative	Accuracy	Specificity	Sensitivity	Precision
1	16569	19	16550	0	0	100.000%	100.0%	100.0%	100.0%
2	16569	19	16550	0	0	100.000%	100.0%	100.0%	100.0%
3	16569	21	16548	0	0	100.000%	100.0%	100.0%	100.0%
4	16569	21	16548	0	0	100.000%	100.0%	100.0%	100.0%
5	16569	32	16537	0	0	100.000%	100.0%	100.0%	100.0%
6	16569	32	16537	0	0	100.000%	100.0%	100.0%	100.0%
7	16569	32	16537	0	0	100.000%	100.0%	100.0%	100.0%
8	16569	25	16544	0	0	100.000%	100.0%	100.0%	100.0%
9	16569	21	16548	0	4	99.976%	100.0%	84.0%	100.0%
10	16569	25	16544	0	0	100.000%	100.0%	100.0%	100.0%
11	16569	25	16544	0	0	100.000%	100.0%	100.0%	100.0%
12	16569	51	16518	0	0	100.000%	100.0%	100.0%	100.0%
13	16569	50	16519	0	1	99.994%	100.0%	98.0%	100.0%
14	16569	6	16563	0	0	100.000%	100.0%	100.0%	100.0%
15	16569	6	16563	0	0	100.000%	100.0%	100.0%	100.0%
16	16569	6	16563	0	0	100.000%	100.0%	100.0%	100.0%
17	16569	6	16563	0	0	100.000%	100.0%	100.0%	100.0%
18	16569	6	16563	0	0	100.000%	100.0%	100.0%	100.0%
19	16569	6	16563	0	0	100.000%	100.0%	100.0%	100.0%
20	16569	25	16544	0	0	100.000%	100.0%	100.0%	100.0%
21	16569	25	16544	0	0	100.000%	100.0%	100.0%	100.0%
22	16569	24	16545	0	1	99.994%	100.0%	96.0%	100.0%
23	16569	25	16544	0	0	100.000%	100.0%	100.0%	100.0%
24	16569	29	16540	0	1	99.994%	100.0%	96.7%	100.0%
25	16569	32	16537	0	0	100.000%	100.0%	100.0%	100.0%
Average	16569	22.09	16,546.91	0.00	0.26	99.998%	>99.99%	99.0%	>99.99%

Figure S1

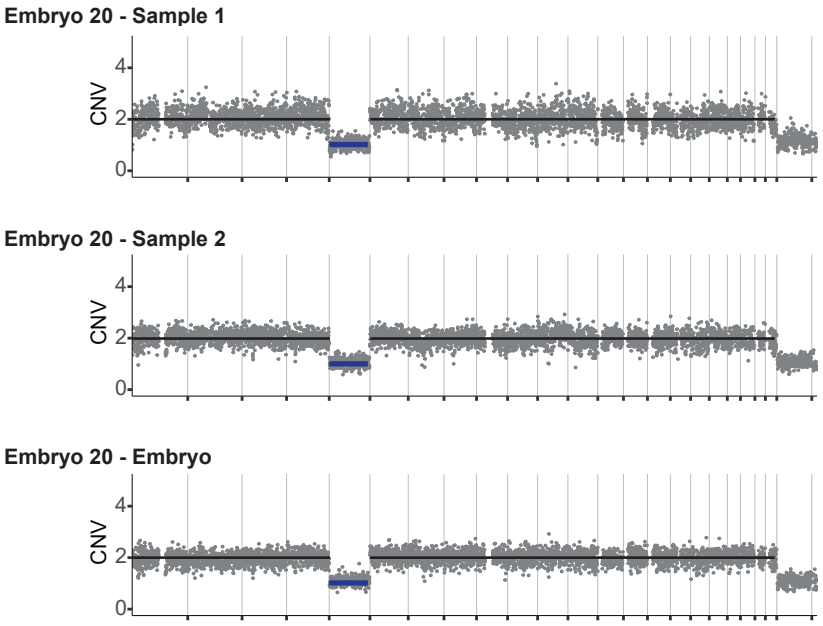
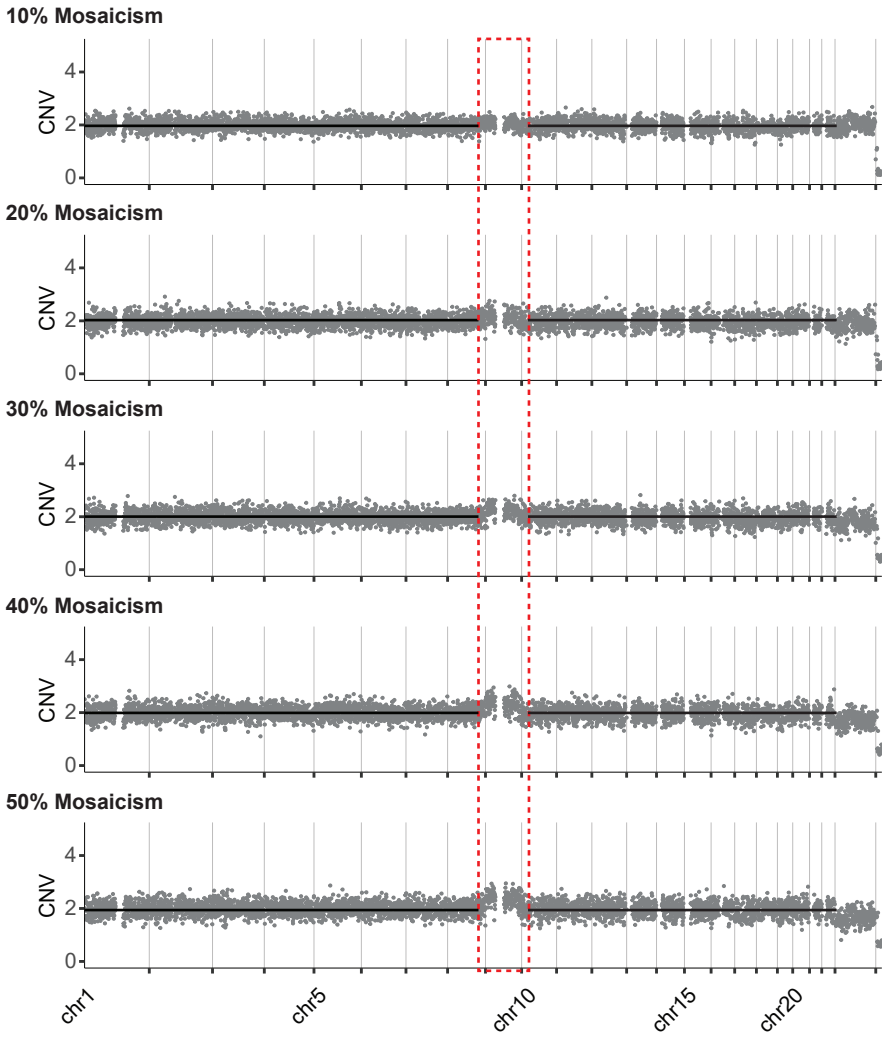


Figure S2



Embryo ID	# tested	# with result	Clinical PGT report	Orchid Results	Concordance within samples	Aneuploidy Status Concordance	Chr-to-Chr Concordance
6	3	3	XY, +8	XY, +8	100%	100%	100%
7	3	3	XX, +4, -10, -18	XX, +4, -10, -18	100%	100%	100%
8	3	3	XY, +15	XY, +15	100%	100%	100%
9	3	3	XX, -20	XX, -20	100%	100%	100%
10	3	3	XY, -13, -16	XY, -13, -16	100%	100%	100%
11	3	3	XX, -21	XX, -21	100%	100%	100%
12	3	3	XY, +7, -15	XY, +7, -15	100%	100%	100%
13	3	3	XY, -5	XY, -5	100%	100%	100%
14	3	3	XY	XY	100%	100%	100%
15	3	2	XY	XY	100%	100%	100%
16	3	3	XY	XY	100%	100%	100%
17	3	2	XXX	XXX	100%	100%	100%
18	3	2	XX, +4, -19	XX, +4, -19	100%	100%	100%
19	3	3	XY	XY	100%	100%	100%
20	2	2	XY, -5, +18	XY, -5	100%	100%	95.80%
21	3	3	XY, +2, +22	XY, +2, +22	100%	100%	100%
22	3	3	XX, +9, +13, +21	XX, +9, +13, +21	100%	100%	100%
23	3	3	XY	XY	100%	100%	100%
24	3	3	XX	XX	100%	100%	100%
25	3	2	XY	XY	100%	100%	100%
26	3	3	XX, -3, +15, +20, +21, +22	XX, -3, +15, +20, +21, +22	100%	100%	100%
27	3	3	XX -16	XX -16	100%	100%	100%
28	3	3	XX -2	XX -2	100%	100%	100%
29	3	3	XY	XY	100%	100%	100%
30	3	3	XY	XY	100%	100%	100%
31	3	3	XY -22	XY -22	100%	100%	100%
32	3	3	XY, -21	XY, -21	100%	100%	100%
33	3	3	XY, -16	XY, -16	100%	100%	100%
34	3	3	XY	XY	100%	100%	100%
35	3	3	XX	XX	100%	100%	100%
	Total Cases	Total Cases with Results	Amplification Success Rate	Concordant with clinical		Aneuploidy Status Concordance	Chr-to-Chr Concordance
	104	100	0.962	98		>99.9%	99.86%

Sample	True Positives	False Positives	True Negatives	False Negatives
NA12878 gDNA Sample 1	3214156	73257	2498498635	37667
NA12878 gDNA Sample 2	3218277	73072	2498498820	33546
NA12878 gDNA Sample 3	3214192	72753	2498499139	37631
NA12878 gDNA Sample 4	3218242	70121	2498501771	33581
NA12878 gDNA Sample 5	3212093	65599	2498506293	39730
Average	**3215392**	**70960**	**2498500932**	**36431**
NA12878 Amplified DNA Sample 1	3172496	67132	2498504760	79327
NA12878 Amplified DNA Sample 2	3199885	63958	2498507934	51938
NA12878 Amplified DNA Sample 3	3194879	57880	2498514012	56944
NA12878 Amplified DNA Sample 4	3200575	59098	2498512794	51248
NA12878 Amplified DNA Sample 5	3162468	67937	2498503955	89355
Average	**3186060**	**63201**	**2498508691**	**65763**

Sample	True Positives	False Positives	True Negatives	False Negatives
1	3270826	72978	2498427158	52753
2	3260902	66729	2498433445	62639
3	3241275	70864	2498447239	64337
4	3208229	71239	2498446853	97394
5	3348398	55757	2498354672	64888
6	3358076	56201	2498354258	55180
7	3307540	53142	2498399224	63809
8	3106492	57099	2498602671	57453
9	3105403	54710	2498605123	58479
10	3213236	91074	2498453268	66137
11	3218889	82827	2498461470	60529
12	3322745	46089	2498393320	61561
13	3299943	40540	2498398825	84407
14	3187551	83342	2498499668	53154
15	3184474	78092	2498504866	56283
16	3201676	72658	2498492003	57378
17	3201352	73875	2498490706	57782
18	3192477	61970	2498516295	52973
19	3175772	58945	2498519275	69723
20	3205470	68532	2498499144	50569
21	3196791	60570	2498507126	59228
22	3212393	58179	2498482967	70176
23	3225625	59208	2498481928	56954
24	3169496	68217	2498487962	98040
25	3324367	58178	2498394241	46929
Average	**3229576**	**64841**	**2498466148**	**63150**