Haploseek: a 24-hour all-in-one method for preimplantation genetic diagnosis (PGD) of monogenic disease and aneuploidy

Daniel Backenroth, PhD¹, Fouad Zahdeh, PhD², Yehuda Kling, MSc², Aharon Peretz, BSc³, Tzvia Rosen, BSc², Dina Kort, BSc², Sharon Zeligson, MSc⁴, Tal Dror, BSc⁴, Sophie Kirshberg, MSc⁴, Efrat Burak, MSc⁴, Reeval Segel, MD^{4,5}, Ephrat Levy-Lahad, MD^{4,5}, David Zangen, MD⁶, Gheona Altarescu, MD^{4,5}, Shai Carmi, PhD¹ and David A. Zeevi, PhD²

Purpose: To develop an economical, user-friendly, and accurate all-in-one next-generation sequencing (NGS)-based workflow for single-cell gene variant detection combined with comprehensive chromosome screening in a 24-hour workflow protocol.

Methods: We subjected single lymphoblast cells or blastomere/blastocyst biopsies from four different families to low coverage (0.3×–1.4×) genome sequencing. We combined copy-number variant (CNV) detection and whole-genome haplotype phase prediction via Haploseek, a novel, user-friendly analysis pipeline. We validated haplotype predictions for each sample by comparing with clinical preimplantation genetic diagnosis (PGD) case results or by single-nucleotide polymorphism (SNP) microarray analysis of bulk DNA from each respective lymphoblast culture donor. CNV predictions were validated by established commercial kits for single-cell CNV prediction.

Results: Haplotype phasing of the single lymphoblast/embryo biopsy sequencing data was highly concordant with relevant ground

truth haplotypes in all samples/biopsies from all four families. In addition, whole-genome copy-number assessments were concordant with the results of a commercial kit.

Conclusion: Our results demonstrate the establishment of a reliable method for all-in-one molecular and chromosomal diagnosis of single cells. Important features of the Haploseek pipeline include rapid sample processing, rapid sequencing, streamlined analysis, and user-friendly reporting, so as to expedite clinical PGD implementation.

Genetics in Medicine (2019) 21:1390–1399; <https://doi.org/10.1038/s41436-018-0351-7>

Keywords: combined molecular PGD and comprehensive chromosome screening; low coverage sequencing; rapid diagnostic method; PGT

INTRODUCTION

From a technical standpoint, it is possible to perform combined single-gene testing with comprehensive chromosome screening by means of microarray (using karyomapping, for example) or high coverage next-generation sequencing (NGS) technologies.^{1,2} Nonetheless, these methods are still time consuming and quite costly, thereby hindering widespread clinical application.³ Less costly methods require parallel processing of either two different methods on one sample^{4–6} or two different methods on two biopsies from the same embryo.⁷ However, taking two biopsies reduces the chances of implantation.⁸ Furthermore, running two parallel methods on the same embryo samples, in a two-step

process, complicates the practical application of such endeavors for mainstream preimplantation genetic diagnosis (PGD) couples. Indeed, many PGD labs performing molecular diagnosis still construct family-specific haplotypes for each gene, based on the analysis of polymorphic markers of the couple and several family members—a practice that causes significant delay from the first meeting of the couple with the PGD team until the time of the first cycle. Therefore, a universal, cost-effective, all-in-one solution for comprehensive single-biopsy embryo testing is desired.

To address these issues, we describe here a novel universal NGS-based workflow, which allows high resolution preimplantation aneuploidy screening and single-nucleotide variant

¹Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; ²Translational Genomics Lab, Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; ³IVF Unit, Division of Obstetrics and Gynecology, Shaare Zedek Medical Center, Jerusalem, Israel; ⁴Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; ⁵Hebrew University Faculty of Medicine, Jerusalem, Israel; ⁶Division of Pediatric Endocrinology and Diabetes, Hadassah Hebrew University Medical Center, Jerusalem, Israel. Correspondence: Shai Carmi (shai.carmi@mail.huji.ac.il) or David A. Zeevi (zeevidavid@szmc.org.il)

Equal contributing authors: Gheona Altarescu, Shai Carmi, and David A. Zeevi

Submitted 31 May 2018; accepted 25 October 2018

Published online: 19 November 2018

diagnosis with reduced cost relative to competing array-based and high throughput sequencing-based technologies. This development is enabled by bioinformatic and statistical methodology tailored to low coverage sequencing, implemented via hidden Markov models. Our method is also rapid, permitting comprehensive test results within 24 hours of embryo biopsy. Furthermore, we have developed a user-friendly interface that enables expedient analysis of the data. Based on these improvements to existing techniques, we propose our rapid, low-cost workflow and high resolution test output as an attractive alternative to existing single-cell preimplantation diagnostic techniques.

MATERIALS AND METHODS

Ethics

DNA and tissue culture samples in this study were donated for research with informed consent according to Shaare Zedek institutional review board guidelines.

Lymphoblast tissue culture derivation, single cell isolation, and embryo biopsy

Lymphocytes derived from fresh whole blood were transformed by Epstein-Barr virus (EBV) infection as described previously.⁹ Resulting lymphoblast cell cultures were grown in suspension in Roswell Park Memorial Institute (RPMI) medium supplemented with L-glutamine, penicillin-streptomycin, and 10% fetal calf serum (Biological Industries). Prior to single cell isolation, each cell culture was diluted 1:1000 in RPMI medium before aliquoting into 6-well plates. Single cells from each culture were identified on an inverted microscope and transferred into a 0.2-ml polymerase chain reaction (PCR) tube containing 2 μ l of phosphate buffered saline (PBS). Embryo blastomere/blastocyst biopsies from clinical PGD cycles were also transferred into a 0.2-ml PCR tube containing 2 μ l of PBS.

Whole-genome amplification and low coverage genome sequencing

Single culture cell/blastomere/blastocyst DNA was whole-genome amplified (WGA) using the PicoPLEX WGA kit (Rubicon Genomics) according to the manufacturer's protocol. Amplified DNA from each cell was then converted into genome sequencing libraries using the Nextera XT library prep kit (Illumina). Resultant libraries were normalized and pooled for rapid 1 $\times$ 50 single read sequencing on a NextSeq 500 (Illumina) instrument to generate 0.3 $\times$ –1.4 $\times$ genomic sequence coverage per sample.

Single-nucleotide polymorphism (SNP) microarray analysis

Genomic DNA was extracted from peripheral blood using the FlexiGene DNA kit (Qiagen). Subsequently, DNA samples were subject to CytoScan[®] 750K or CytoScan[®] HD single-nucleotide polymorphism (SNP) array (Affymetrix) and then scanned on an Affymetrix GeneChip (GCS3000) Scanner. Chromosome Analysis Suite (Chas; Affymetrix) was used to extract genotype calls and copy-number predictions (where relevant) from each array.

NGS variant calling and ground truth haplotypes

Low coverage high throughput sequencing reads were aligned to the reference human genome (hg19) using Burrows-Wheeler alignment (BWA)¹⁰ followed by crude variant calling using SAMtools mpileup at SNP positions matching those on the Affymetrix arrays. To obtain ground truth haplotypes, we inferred paternal and maternal haplotypes throughout the genomes of each sequenced sample using corresponding bulk microarray data. The ground truth haplotypes were then compared with haplotypes inferred using the low coverage sequencing data alone, given microarray genotypes for the parents and one child of each family.

Haplotype prediction using low coverage sequencing data

Details regarding haplotype prediction using low coverage sequencing data are described in the Supplementary Materials and Methods.

Copy-number variant prediction

Details regarding copy-number variant (CNV) prediction using low coverage sequencing data are described in the Supplementary Materials and Methods.

Output visualization

A custom user-friendly interface was created to visualize data output on a web browser in two plots depicting the haplotype prediction and the marginal probabilities, respectively (see Results for further explanation of this output). Detected CNVs were plotted with ggplot2.

Validation

All single-cell haplotype predictions from tissue cultures were validated by Affymetrix 750k SNP microarray of bulk genomic peripheral blood DNA from the corresponding tissue culture donor. Embryo biopsy haplotype predictions were validated by PCR-based clinical PGD case results. For all WGA samples in the study, CNV predictions were validated by testing the same WGA material with established commercial kits for single-cell CNV detection (BlueGnome array comparative genome hybridization or VeriSeq-PGS; Illumina).

RESULTS

Whole-genome low coverage sequencing provides accurate haplotype information in single-cell samples

Three families, of 3–4 children each, were recruited into the pilot study. Lymphoblast tissue cultures were prepared from whole blood of each child and then DNA from single-cell isolates was subjected to low coverage (0.5 $\times$ –0.9 $\times$) genome sequencing.

High coverage sequencing generally mitigates genotyping errors that are traced to technical limitations of NGS systems. Low coverage sequencing, on the other hand, is highly susceptible to technical error because detected variants cannot be confirmed by more than one or two sequencing reads. Performance of low coverage sequencing on whole-genome

amplified (WGA) samples is especially challenging given the additional inherent limitation of allelic dropout in WGA material. Nevertheless, the economic savings of low coverage versus high coverage sequencing suggest that there is merit to establish a reliable haplotyping method that is resistant to haphazard errors in WGA samples and the low coverage sequencing data sets derived from them. We thus attempted to reconstruct whole-genome haplotypes from the low coverage sequencing data of the aforementioned families.

The first step in this endeavor involved derivation of high quality whole-genome genotypes for each of the parents. To this end, SNP microarray genotyping was performed, in each family, on whole-blood genomic DNA from both parents and from one child, whom we designate as child 1. Then, variants at matching array positions were called in the low coverage whole-genome data sets of the other children, whom we designate as children 2, 3, and (whenever available) 4. We integrated the information from the SNP arrays of the parents and child 1 together with the sequencing data for the other children using a hidden Markov model (HMM; see Methods). The HMM predicted whether or not the parental haplotypes transmitted to child 1 were also inherited by each of the sequenced children across all chromosomes.

We demonstrate the haplotype predictions for family 1, in which the father and mother were both carriers of different variants in the *BRCA2* gene that were transmitted to their children in various combinations (see inset of Fig. 1). Child 1 in that family had inherited a different *BRCA2* variant from each parent. For each child, we plot along each chromosome the maternal and paternal haplotypes that were either shared (dark shading) or not shared (light shading) with child 1 (Figs. 1 and 2). In families 2 and 3, the single-cell children samples were also genome sequenced at low coverage ($0.7\times$ – $1.4\times$) (Table 1). As in family 1, we inferred the sharing of child 1's haplotypes by the sequenced children in each of these families. The results are shown in Figures S1 and S2.

In the long run, clinical application of low coverage genome sequencing for PGD haplotype phasing requires quality control measurements of the test results. Accordingly, we used the “posterior probabilities” reported by our HMM as a measure of confidence in the haplotype prediction. This statistic, which we also call the “marginal probability,” is an estimate of the degree of confidence for a given maternal/paternal haplotype prediction (at any given microarray SNP) based on the sequencing reads and the array-determined parental haplotypes. A marginal probability score near 1.0 (for a specific parental haplotype prediction) indicates high predictive confidence that a given sibling has inherited the same haplotype as child 1 (for that parent). Conversely, a marginal probability score near 0.0 indicates high predictive confidence for the transmission of the haplotype not shared with child 1. See Fig. 2 for example marginal plots.

We generally consider marginal scores between 0.01 and 0.99 as low confidence. Typically, these low confidence sites appear in close proximity to recombination site predictions. However, they may also appear at sites where chromosome

copy-number aberrations are present, or in regions of consanguinity, as we discuss below. Hence, we used marginal probability scores to quality control all haplotype predictions output by our HMM.

To assess performance of the haplotype predictions in all sequenced samples (from all three families in the pilot study), we compared the predictions with the ground truth SNP array haplotypes, as determined using bulk DNA of each respective child. When we assessed phasing accuracy of SNPs passing quality control (marginal probability >0.99 or <0.01), the overall accuracy was remarkably high ($>98\%$), even with sequence coverage as low as $0.5\times$ (mean at array SNPs), and with only 24.7–34.6% of the SNPs covered by any read (Table 1; for phasing accuracies at other marginal probability thresholds see Figure S3). The proportion of SNPs passing quality control ranged from 89.8 to 98.6% (Figure S4). This suggests that our prediction model has high sensitivity for accurate haplotype prediction despite low coverage sequencing of single-cell samples.

Interestingly, the phasing accuracies in families 2 and 3 were slightly lower than that in family 1 (Table 1). We hypothesized that this effect is due to long genomic stretches in which one or both parents are homozygous, owing to consanguineous unions. Alternatively, parents from consanguineous families could also have identical diploid genotypes (i.e., they are so-called IBD2 in a region), which would also lead to low phasing accuracy. Indeed, such regions of consanguinity (ROC) accounted for 6.7–10.4% of the SNPs in families 2 and 3 as opposed to only 0.2–0.3% in family 1 (Table 1). In these regions, the single-cell child genotype is trivially determined but the assignment of transmission pattern is more difficult because more than one pattern is consistent with the observed genotypes. Indeed, when excluding ROC, we found that phasing accuracy was near perfect in all single-cell samples ($>99.8\%$) whereas high error rates were identified in ROC-specific regions, even for SNPs passing quality control (Table S1). This suggests that for couples from consanguineous families such as the parents in families 2 and 3, it is important to identify ROC in the array data (ahead of sequencing) so as to prevent misdiagnosis of parental transmission patterns in these regions.

Whole-genome low coverage sequencing detects chromosome copy-number variants in single-cell samples

Inference of chromosomes' copy number from low coverage sequencing data is well established in the field of reproductive medicine. Briefly, the common practice is to divide the human genome into “windows,” or bins, inside which NGS read counts are summed for comparison with a reference of multiple samples sequenced to similar coverage. The single cell from the family 1 child 3 lymphoblast culture presented with two very large CNVs in the sequencing data (Fig. 3); one segmental copy-number gain of ~ 70 Mb length on chromosome 3 and another ~ 25 Mb loss on chromosome 21. The chr21 loss also led to low confidence haplotype prediction along the same segment, as can be seen in the paternal

Fig. 1 Haplotype phasing using low coverage genome sequencing in single-cell samples. Single lymphoblasts from children 2, 3, and 4 (from family 1, depicted in the inset) were subjected to low coverage ($\sim 0.5\times$ – $0.9\times$) genome sequencing analysis. In parallel, bulk genomic DNA of the parents and one other sibling (child 1) were genotyped on single-nucleotide polymorphism (SNP) microarrays to phase whole-genome parental haplotypes. Array SNPs that were covered in each of the single-cell samples from children 2, 3, and 4 were phased using a custom hidden Markov model and haplotype blocks were derived on all chromosomes. Parental haplotype block predictions for each single lymphoblast are depicted by chromosome and child according to the legend. X markings indicate centromeres. V1 and V2, in the inset, indicate two different variants in the *BRCA2* gene. Similarly, N1 and N2 indicate two different wild type (“normal”) alleles. All haplotype predictions in the genome plot were tested against SNP microarray data from corresponding bulk DNA samples of each child, demonstrating $>99.4\%$ phasing accuracy for all samples.

marginal probability plot of child 3 (Figure S5). This result was verified by a commercial CNV detection kit (Veriseq-PGS) analysis of child 3 single-cell DNA (Figure S6; Table 1). Our prediction that children 2 and 4 are euploid was also verified by Veriseq-PGS (Table 1). In addition, we ran similar tests on single cells from the children in families 2 and 3 with whole-genome CNV predictions that were proven highly accurate after Veriseq-PGS validation (Table 1). Therefore, we conclude that our new integrated method for single-cell genetic testing may also be robust enough for clinical comprehensive chromosome copy-number screening (CCS).

Reproducibility and sequencing coverage limits of Haploseek

In the clinical setting, it is essential to establish the reliability of a new method by identifying technical limitations and assessing reproducibility. We started by investigating the lower limit of the per-sample sequencing coverage that would not overcompromise on test accuracy. The answer to this question would help determine how many embryo samples could be sequenced and diagnosed in one sequencing run to maximally reduce the cost per embryo of the assay. We took

all non-child 1 sequencing data sets from families 1–3 and randomly downsampled the sequencing reads of each sample to various lower fractions of coverage relative to the initial experiment. To measure assay performance, we computed the proportion of SNPs with highly accurate haplotype prediction (marginal scores >0.99 or <0.01 , for which phasing accuracy $>99\%$ was demonstrated above) at different lower coverage milestones. We limited this analysis to genomic regions for which ROC were not present. Our *in silico* experiment indicated that accurate haplotype predictions could be obtained for all samples at $0.4\times$ the original coverage of experiments in Table 1 (Figures S7–S9). In terms of microarray SNP coverage, these statistics suggest that Haploseek could accurately predict haplotypes for samples with coverage as low as $0.2\times$ genome coverage. Hence, these experiments indicate that Haploseek supports multiplex sequencing with uncompromised haplotype prediction accuracy, a prospect that markedly reduces the cost per embryo (see Discussion).

To validate the aforementioned computational predictions, as well as assess the reproducibility of Haploseek results, we prepared new sequencing libraries, in duplicate, from all child

Fig. 2 Gene-proximal haplotype predictions in low coverage single-cell sequencing data. Pictured is a screen capture of the haplotype data in Fig. 1 zoomed in to the *BRCA2* gene locus (± 2 Mb) on chromosome 13. Maternal and paternal haplotypes are each depicted in “prediction” and “marginal” plots for each child, with alternating clear and gray background shading distinguishing the plots of one child from that of neighboring children above and below. The “prediction” plot indicates the hidden Markov model (HMM) haplotype prediction in each child relative to child 1. Dark red and dark blue shaded segments indicate child 1 matching maternal and paternal haplotypes (containing V1 and V2 *BRCA2* variants), respectively. Lighter shading indicates wild type haplotypes that are not present in child 1. The “marginal” plot indicates the degree of confidence with which the HMM is reporting a certain haplotype prediction. A marginal value of 1 means a high confidence prediction of a child 1 matching haplotype. A marginal value of 0 means high confidence prediction of a child 1 mismatching haplotype. We call sites with marginal values between 0.01 and 0.99 low confidence. These sites often appear near meiotic recombination sites. The marginal plots also depict the SNP sites (red/blue dots) on the arrays that were used to phase the parental haplotypes and were also successfully sequenced. Note that children 2, 3, and 4 exhibit an identical paternal recombination event ~ 1 Mb downstream of the *BRCA2* gene (demarcated by purple arrow and dashed vertical line). This strongly suggests that child 1 is recombinant at this site. Array data from children 2, 3, and 4 confirmed these high resolution findings (see also the inset of Fig. 1).

WGA samples in families 1–3. This amounted to seven pairs of libraries (three from family 1; two from family 2; and two from family 3), or 14 indexed samples altogether, which were all loaded onto a single 1 $\times$ 50 sequencing run. At this level of sequencing multiplexing, the mean genome coverage per sample ranged from 0.3 $\times$ to 0.6 $\times$ (Table S2) and, in most cases, resulted in lower sequencing coverage relative to that of the original experiment (Figures S7–S9). Nonetheless, haplotype prediction accuracy remained high in non-ROCs for all samples and their replicates, as predicted by the aforementioned computational downsampling experiments (Figures S7–S9; Table S2). Moreover, despite new sets of randomly sequenced SNPs in each sample replicate, nearly equivalent results and accuracy measurements were generated for all samples without exception (Table S2). Furthermore, the CNV predictions of Haploseek in the lower coverage data were completely equivalent to that of the original pilot

experiments (Table S2). Therefore, these results indicate that Haploseek haplotype and CNV predictions are highly reproducible, as well as accurate even at 0.3 $\times$ –0.6 $\times$ coverage.

A clinical application of Haploseek for identifying embryos of normal karyotype in translocation PGD

Given the success of the pilot experiments involving tissue culture cells from families 1–3, we further validated Haploseek using blastomere or blastocyst biopsy WGA material from clinical PGD cases. Family 4 is nonconsanguineous, consisting of a father with a normal karyotype, a mother with a balanced reciprocal translocation (karyotype 46,XX,t[3;4][p26;p16.1]), a diseased child of an unbalanced karyotype (child 0 with karyotype: 46,XY,der[3]t[3;4][p26;p16.1]mat), and another healthy male child (child 1) who shares the mother’s balanced translocation karyotype (Fig. 4a). In 2014, the family 4 parents sought PGD in our clinic for selection of only

Table 1 Summary of Haploseek results for combined copy-number variation (CNV) and haplotype prediction of the indicated single-cell isolates

Family	Child	Parent	Mean coverage of array SNPs ^a	Percent of array SNPs sequenced	No. of SNPs with phasing prediction pass filter ^b	Phasing accuracy on SNPs pass filter ^c	No. of SNPs in ROC ^d	Percent SNPs in ROC ^d	No. of SNPs not in ROC ^d	Phasing accuracy not in ROC ^d	CNV prediction	Veriseq-PGS CNV validation ^e
1	2	Father	0.87	34.6	191,190	100.000%	643	0.3	190,547	100.000%	Euploid female	Euploid female
		Mother			190,052	99.999%	643	0.3	189,409	99.999%		
	3	Father	0.64	28.9	187,644	100.000%	414	0.2	187,230	100.000%	chr3q gain;	chr3q gain;
		Mother			186,972	99.987%	414	0.2	186,558	99.987%	chr21q loss;	chr21q loss;
	4	Father	0.52	26.4	188,933	100.000%	643	0.3	188,290	100.000%	aneuploid female	aneuploid female
		Mother			186,695	99.995%	643	0.3	186,052	99.995%	Euploid female	Euploid female
2	2	Father	1.41	33.3	180,910	98.427%	13321	7.4	167,589	99.973%	chr4q gain;	chr4q gain;
		Mother			184,033	98.720%	18419	10.0	165,614	99.968%	chr12q loss;	chr12q loss;
	3	Father	1.39	31.0	179,930	99.582%	13568	7.5	166,362	99.980%	aneuploid male	aneuploid male
		Mother			185,606	99.324%	19310	10.4	166,296	99.966%	Euploid female	Euploid female
3	2	Father	0.72	24.7	177,066	98.747%	11945	6.7	165,121	99.926%	Euploid female	Euploid female
		Mother			174,677	99.868%	12552	7.2	165,125	99.920%		
	3	Father	1.16	31.4	179,761	98.624%	13883	7.7	165,878	99.814%	Euploid female	Euploid female
		Mother			175,753	98.824%	12480	7.1	163,273	99.940%		

Haplotype/genotype predictions were validated by single-nucleotide polymorphism (SNP) microarray of bulk genomic DNA and CNV predictions were validated with Veriseq-PGS. ROC denotes regions of consanguinity (long genomic stretches [95% of 200 consecutive SNPs] in which both or one of the parents in a given family were homozygous; or long genomic stretches [95% of 200 consecutive SNPs] where both parents shared both haplotypes).

^aThe per-sample mean sequencing coverage of all SNP sites on the Affymetrix Cytoscan 750k SNP microarray platform.

^bNumber of array SNPs, out of 200,484 SNPs on the Affymetrix Cytoscan 750k SNP microarray, for which haplotype prediction passed marginal probability filters of >0.99 or <0.01 (see Fig. 2 for example marginal probability plots).

^cPhasing accuracy was defined as the proportion of SNPs where the inferred inheritance state with respect to both of child 1's maternal and paternal haplotypes (given marginal probability >0.99 or <0.01) did not coincide with a long segmental deletion (as in chr21q in family 1 child 3) and was the same as the inferred inheritance state determined when running the hidden Markov model using array genotypes rather than sequencing data.

^dOnly autosomal SNPs passing marginal probability filters (>0.99 or <0.01) were included in these analyses. This accounts for varying SNP counts in the analysis of each child in each family.

^eVeriseq-PGS was performed on the same single-cell isolate as that in Haploseek.

Fig. 3 Low coverage genome sequencing reveals aneuploidy in the family 1 child 3 single-lymphoblast sample. We analyzed single-cell low coverage sequencing data from the experiment in Fig. 1 for copy-number variation using a custom hidden Markov model (Methods). Results are shown for children 1–4 of family 1. The y-axis shows GC-content normalized read counts across 1-Mb sized windows, divided by the median of normalized counts from euploid reference samples. Final scores (y-axis) of 0.0, 0.5, 1.0, and 1.5 indicate 0, 1, 2, and 3 chromosome copies, respectively. Chromosomes are indicated on the x-axis below the plots and copy-number variant (CNV) predictions are indicated by shaded vertical bars as indicated in the legend. Note the CNV segmental gain of chr3q and the loss of chr21q in child 3. All children aside from child 3 were euploid. On sex chromosomes, the 1.0 score on chrX and the 0.0 score on chrY are consistent with female sex for all four children.

“normal” and not “balanced” embryo karyotypes. For typical translocation cases, standard practice is to offer a chromosome copy-number assay that would distinguish embryos of balanced/normal karyotype from unbalanced embryos without discriminating between the two “healthy” (balanced versus normal) outcomes. However, to distinguish between the two latter scenarios and simultaneously fulfill the request of the family 4 parents, embryo haplotyping would also be required, in addition to CCS. Hence, our integrated CNV and haplotype measuring Haploseek platform is well equipped to handle this situation. Nonetheless, because Haploseek did not exist at the time of the original genetic counseling, we instead offered the couple a two-step diagnostic approach, whereby each day 3 and/or day 5 embryo from their PGD cycles would be assessed for chromosome copy number and haplotype by two independent methods. For CCS, we used either array comparative genomic hybridization (aCGH) or low pass genome sequencing (using Veriseq-PGS). For haplotype determination, we used a PCR-based assay, which measured the size of multiple polymorphic short tandem repeats (STRs) flanking each known maternal translocation breakpoint on chromosomes 3 and 4.

Given that CCS is a generic application, no precase workup was required. However, STR-based typing of embryo biopsies requires extensive preparation. First, we harnessed the genetic information from the family 4 child of unbalanced karyotype (child 0) by probing his genomic DNA by microarray to

identify his maternal translocation breakpoints (Fig. 4b). Then, we designed multiple PCR assays for STR typing of the immediate (± 2 Mb) translocation breakpoint-flanking genomic regions on chr3:5,012,809 and chr4:7,927,680 in the family 4 parents and child 1 (who shares the same balanced karyotype as his mother). Treating the precase workup as that of two autosomal dominant diseases, we determined the balanced translocation-linked haplotypes of the mother by identifying both the chr3 and chr4 breakpoint-spanning haplotypes that she shared with child 1. Using this phased information, we identified polymorphic markers for PGD case application. We proceeded to perform 10 PGD cycles for family 4 over the course of three years, before pregnancy and birth of a healthy child with normal karyotype was achieved.

During these cycles, 86 blastomeres/blastocysts were tested by two-step CCS and PCR haplotyping to identify (and distinguish between) embryos of aneuploid, unbalanced, balanced, and normal karyotype. In parallel, the residual WGA material from these biopsies was stored and donated to the diagnostic lab for research. To clinically validate Haploseek and further confirm the low sequencing coverage predictions of our previous Haploseek experiments, we selected 16 blastomere/blastocyst biopsies from this cohort for retesting. Among these embryos, the original testing identified a relatively even distribution of normal, balanced, and/or aneuploid karyotypes. Genome sequencing libraries were prepared from each of the 16 WGA samples and

Fig. 4 Clinical utility of Haploseek for normal karyotype embryo selection in translocation preimplantation genetic diagnosis (PGD) testing. (a) A pedigree of family 4 in which the mother and child 1 are both carriers of the balanced translocation $t(3;4)(p26;p16.1)$. Child 0 is a diseased offspring with an unbalanced karyotype $46,XY,der(3)t(3;4)(p26;p16.1)mat$. For haplotype phasing of the parents, child 1 DNA was used to flag the maternal translocation chromosomes in clinical PGD haplotype testing. (b) High resolution DNA microarray analysis demarcates translocation breakpoints in child 0 DNA. The left panel shows the microarray analysis suite (Chas) screen capture of the chr3p deletion in child 0 and the right panel shows the Chas screen capture of the chr4p duplication in child 0. (c) Screen captures of Haploseek haplotype prediction plots for embryo biopsies of children 8 and 9 in family 4. The ± 2 Mb flanking region of the familial chr3 and chr4 translocation breakpoints are shown in the left and right panels, respectively. Breakpoints are marked by dashed vertical lines. For family 4 (in which the mother is the balanced translocation carrier), only the maternal haplotype predictions are relevant for balanced versus normal karyotype diagnosis. Both children 8 and 9 were euploid on chromosomes 3 and 4 according to Haploseek copy-number variant (CNV) analysis (see Table S3 and Figure S10). However, only child 8 inherited the normal maternal haplotypes in chromosomes 3 and 4 according to Haploseek (given that the maternal haplotypes flanking the translocation breakpoints did not match that of child 1, the reference child with the balanced karyotype). Conversely, the child 9 karyotype was balanced. These and all other haplotype predictions tested by Haploseek were completely concordant with polymerase chain reaction (PCR)-based clinical PGD test results (see Table S3).

sequenced in a single 1×50 sequencing run according to the standard Haploseek protocol. The sequencing coverage per sample ranged from 0.35 to 0.54. Haploseek CNV analysis determined that seven embryos were unbalanced for the familial translocation (Figure S10; Table S3). Among the remaining nine embryos that were euploid at the chr3 and chr4 translocation breakpoints, five were of balanced and four were of normal karyotype, as determined by Haploseek haplotype analysis (Fig. 4c and Table S3). Seven of 16 embryos were aneuploid for chromosomes other than chr3 or chr4 (including embryos with either aneuploid [2], balanced [3], or normal [2] karyotype at the translocation). Importantly, all CNVs and haplotypes determined by the two-step clinical PGD testing were concordant with Haploseek's predictions. Thus, collectively these results indicate that Haploseek is appropriate for standard and challenging clinical PGD applications alike.

DISCUSSION

Classical molecular PGD still requires custom haplotype construction for the family being tested, a practice that is time consuming and only useful for one specific genomic locus at a time. In this regard, whole-genome aneuploidy screening is more easily and rapidly applied, because it does not require any custom family-specific preparation in advance of the clinical test. Existing CNV detection methods (namely, aCGH and low coverage NGS) bypass customization by assaying all human chromosomes in each embryo biopsy with a universal assay. This generic test setup allows couples with relevant aneuploidy-related indications (i.e., advanced maternal age, recurrent miscarriage, recurrent implantation failure, or balanced translocation carriage) to proceed immediately with in vitro fertilization (IVF) treatments and CCS of embryos without delay. Therefore, the time-saving benefits of universal assay development would be evident if applied equally to couples with molecular PGD indications as well.

As mentioned previously, a number of techniques have previously been developed to address the need for generic molecular PGD assays (while still maintaining generic whole-genome CNV testing capability). These techniques, while accurate, are not widely applied because they require either extended time to diagnosis (taxing lab technician and/or molecular reagent resources and excluding the possibility of a fresh embryo transfer) and/or incur high per-embryo costs to fund the procedure. These applications can be grouped by technology into array-based or NGS-based categories. In the array category, SNP genotyping arrays (such as karyomapping and haplarithmisis) have proven to be quite effective at whole-genome haplotyping of embryo biopsies within a time frame that is suitable for fresh embryo transfer (24-hour protocols from biopsy until test result).^{1,2,11–15} However, these methods become quite costly when multiple embryos are tested, because multiple embryo biopsy samples cannot be multiplexed onto one individual microarray chip. Rather, each embryo is assayed on its own (often costly) microarray.

In the NGS category, sample multiplexing is theoretically feasible and therefore the cost per embryo should be lower than that of SNP array methodologies. Nonetheless, sample multiplexing is limited by the capacity of high throughput sequencers. For whole-genome haplotype phasing of embryo biopsies, previous studies have shown that high coverage (30×) genome sequencing is effective.^{16,17} However, these high coverage sequencing methods are lengthy (requiring a few days for processing) and in many instances do not allow multiplexing and thus incur higher costs than array-based methodologies.

In a more practical manner, one group has developed a more economical NGS method, termed MARSALA (mutated allele revealed by sequencing with aneuploidy and linkage analyses), which combines low coverage genome sequencing (0.1× coverage) together with low coverage targeted sequencing (2× coverage) to facilitate combined aneuploidy screening with haplotype phasing.^{18,19} The MARSALA method is indeed rapid (24 hours from biopsy to test result), but because the haplotype phasing is accomplished by targeted sequencing, it is not universal. Thus, each genomic region to be assessed for molecular PGD requires its own custom targeted sequencing assay in the MARSALA scheme (which is precisely one of the problems with classical molecular PGD methods). Therefore, to address all the aforementioned issues with each of the previously described comprehensive PGD and CCS technologies and methods, we developed Haploseek.

Haploseek is economical because it is predicated upon low coverage sequencing of embryo biopsies. Up to 16 embryo biopsies may be tested in one 10-hour 1×50 sequencing run; or up to 48 embryo biopsies may be tested in one 18-hour 2×75 sequencing run (for a mean per-sample genome coverage of 0.4× at full run capacity). Haploseek does incur SNP array investment for molecular PGD application because it requires precase processing of the PGD couple and a first-degree family member by means of array technology. However, this outlay is a one-time investment that carries

over into subsequent PGD testing of all embryos from the couple in all PGD cycles. Moreover, for array-based universal PGD methods such as karyomapping, this initial outlay is also required. Nonetheless, the primary difference between Haploseek and these array-based methods is cost. In Israel, the price savings for Haploseek versus karyomapping in a 16-embryo PGD case (involving a precase array testing of a father, mother, and child trio) is approximately US\$390. The saving increases to US\$971 when the 16 embryos are split between two different families (for each of which a trio of arrays were performed in advance of the PGD cycles). This latter scenario is more likely to occur in fresh PGD cases because, for a single couple, an 8-embryo fresh cycle is a more likely occurrence than a 16-embryo fresh cycle. Regarding frozen cycles, where at least 48 hours are available between embryo biopsy and test result, the longer 2×75 sequencing configuration can facilitate even more profound cost-saving in comparison with karyomapping. Assuming eight embryos per PGD couple, Haploseek can fit 48 embryos from six families (including the cost of performing a trio of microarrays for each family) into a single sequencing run (without compromising on accuracy) for an outstanding US\$5000 cost reduction relative to karyomapping. Thus, for laboratories or clinics seeking to increase PGD testing throughput, Haploseek seems well positioned economically for the future.

In conclusion, we describe a new user-friendly, cost-effective, accurate, and practical method for integrated molecular PGD and CCS testing of single-cell samples. Other advantages as well as limitations of the method are discussed further in the Supplementary Discussion.

ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (<https://doi.org/10.1038/s41436-018-0351-7>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

The authors acknowledge the Shaare Zedek Mirsky intramural grant and thank Rabbi David Fuld for funding this research. S. C. thanks the Israel Science Foundation grant 407/17.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Zamani Esteki M, Dimitriadou E, Mateiu L, et al. Concurrent whole-genome haplotyping and copy-number profiling of single cells. *Am J Hum Genet.* 2015;96:894–912.
2. Thornhill AR, Handyside AH, Ottolini C, et al. Karyomapping—a comprehensive means of simultaneous monogenic and cytogenetic PGD: comparison with standard approaches in real time for Marfan syndrome. *J Assist Reprod Genet.* 2015;32:347–356.
3. Vermeesch JR, Voet T, Devriendt K. Prenatal and pre-implantation genetic diagnosis. *Nat Rev Genet.* 2016;17:643–656.
4. Rechitsky S, Pakhalchuk T, San Ramos G, Goodman A, Zlatopolsky Z, Kuliev A. First systematic experience of preimplantation genetic diagnosis for single-gene disorders, and/or preimplantation human leukocyte

- antigen typing, combined with 24-chromosome aneuploidy testing. *Fertil Steril*. 2015;103:503–512.
5. Minasi MG, Fiorentino F, Ruberti A, et al. Genetic diseases and aneuploidies can be detected with a single blastocyst biopsy: a successful clinical approach. *Hum Reprod*. 2017;12:1–8.
 6. Zimmerman RS, Jalas C, Tao X, et al. Development and validation of concurrent preimplantation genetic diagnosis for single gene disorders and comprehensive chromosomal aneuploidy screening without whole genome amplification. *Fertil Steril*. 2016;105:286–294.
 7. Van der Aa N, Zamani Esteki M, Vermeesch JR, Voet T. Preimplantation genetic diagnosis guided by single-cell genomics. *Genome Med*. 2013;5:71.
 8. Haapaniemi Kouru K, Malmgren H, Nordenskjold M, Fridstrom M, Csemiczky G, Blennow E. One-cell biopsy significantly improves the outcome of preimplantation genetic diagnosis (PGD) treatment: retrospective analysis of 569 PGD cycles at the Stockholm PGD centre. *Hum Reprod*. 2012;27:2843–2849.
 9. Hui-Yuen J, McAllister S, Koganti S, Hill E, Bhaduri-McIntosh S. Establishment of Epstein-Barr virus growth-transformed lymphoblastoid cell lines. *J Vis Exp*. 2011;57:e3321.
 10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.
 11. Dimitriadou E, Melotte C, Debrock S, et al. Principles guiding embryo selection following genome-wide haplotyping of preimplantation embryos. *Hum Reprod*. 2017;32:687–697.
 12. Kumar A, Ryan A, Kitzman JO, et al. Whole genome prediction for preimplantation genetic diagnosis. *Genome Med*. 2015;7:35.
 13. Handyside AH, Harton GL, Mariani B, et al. Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes. *J Med Genet*. 2010;47:651–658.
 14. Natesan SA, Handyside AH, Thornhill AR, et al. Live birth after PGD with confirmation by a comprehensive approach (karyomapping) for simultaneous detection of monogenic and chromosomal disorders. *Reprod Biomed Online*. 2014;29:600–605.
 15. Natesan SA, Bladon AJ, Coskun S, et al. Genome-wide karyomapping accurately identifies the inheritance of single-gene defects in human preimplantation embryos in vitro. *Genet Med*. 2014;16:838–845.
 16. Peters BA, Kermani BG, Alferov O, et al. Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing. *Genome Res*. 2015;25:426–434.
 17. Xu Y, Chen S, Yin X, et al. Embryo genome profiling by single-cell sequencing for preimplantation genetic diagnosis in a beta-thalassemia family. *Clin Chem*. 2015;61:617–626.
 18. Yan L, Huang L, Xu L, et al. Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by next-generation sequencing with linkage analyses. *Proc Natl Acad Sci USA*. 2015;112:15964–15969.
 19. Liu W, Zhang H, Hu D, Lu S, Sun X. The performance of MALBAC and MDA methods in the identification of concurrent mutations and aneuploidy screening to diagnose beta-thalassaemia disorders at the single- and multiple-cell levels. *J Clin Lab Anal*. 2018;32:e22267.