DOPlify™ Advanced DOP-PCR Whole Genome Amplification Kit

Evaluation using euploid and aneuploid cells on MiSeq and NextSeq sequencing platforms

Aneuploidy screening using Next Generation Sequencing (NGS) technology is becoming increasingly popular in the IVF setting. When limited source material for genome wide evaluation is available, whole genome amplification (WGA) is often used to generate sufficient DNA for downstream analysis. A variety of kits are available which make use of different WGA technologies.

The total aligned sequencing data generated from WGA NGS runs is potentially sufficient for subsequent sample screening for a range of genetic anomalies, including preimplantation genetic screening (PGS), mitochondrial DNA (mtDNA) load assessment and preimplantation genetic diagnosis (PGD).

The aim of this study was to determine the suitability of WGA products in a PGS NGS workflow for the concurrent screening of PGD for monogenic disorders. A list of genes of clinical significance was selected, as detailed by the International Standards for Cytogenomic Arrays Consortium (ISCA) for bioinformatic evaluation, as a model of genes of clinical significance.

Material and Methods

Single cells were sorted from a karyotypically defined aneuploid fibroblast cell line (Coriell Institute) and lymphocytes from a female of proven fertility. The cells were subjected to WGA using the RHS DOPlify™ kit. WGA DNA was quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific). Nextera libraries were prepared from 50ng of WGA products from single cell and also from unamplified gDNA from the same cell sources following DNA fragment by tagmentation. DNA barcoding was performed with the use of the Nextera 96 Index Kit (Illumina). The quality of the libraries was assessed with the use of the Agilent High-Sensitivity DNA Kit (Agilent Technologies).

The average size of the library fragments prepared for the MiSeq NGS run was 529.43bp. Paired-end dual index 300bp sequencing was performed on a Miseq (Illumina) using the 600 cycles v3 kit according to standard protocol. Also, paired-end dual index 150bp sequencing was performed on a Nextseq (Illumina) using 300 cycles v3 kit according to standard protocol.

Bioinformatic tools were used to remove unmapped reads, duplicate reads and reads with low mapping scores. The sequencing data was then bioinformatically aligned to hg19 with bwa 0.7.5 (default settings).

In addition, aliquots of RHS WGA products were labelled and hybridized to the EmbryoCellect™ array to determine aneuploidy status according to manufacturer's instructions (EmbryoCellect™ TDS version 3.2.

Next Generation Sequencing Metrics

On completion of the sequencing runs, approximately 4.5 fold more reads were generated using the NextSeq compared to MiSeq platform (see Table 1 below). Comparisons to alternate commercially available NGS technology are also made.

Table 1; Comparison of NGS workflow and outputs from the RHS protocols used in this study and commercially available Pre-implantation Genetic Screening (PGS) products.

	RHS MiSeq	RHS NextSeq	VeriSeq	Ion ReproSeq 318 chip
	600 cycles	300 cycles	Illumina MiSeq	ThermoFisher Scientific Ion PGM
WGA	RHS DOP-PCR	RHS DOP-PCR	SurePlex LA PCR	Isothermal
Library	Nextera	Nextera	Nextera	Ion SingleSeq™ Kit
Read length (bp)	300bp Paired-end	150bp Paired-end	36bp	230-240bp
Total capacity reads	30 million	130 Million	30 million	4-5 million
Number of samples	49	23	24	2-24
Total reads/sample			1,000,000	166,000
Total reads PF/sample	1.3 million	5.8 million	500,000	100,000
Q30 Score	>89	>82		

Results

The DOPlify™ WGA generates DNA fragments ranging in size from 200-2kb (Figure 1).

EmbryoCellect™ results

The ploidy of all single cells was confirmed using the RHS EmbryoCellect™ aCGH platform prior to preparation of the NGS libraries to ensure that the cell had maintained its aneuploidy during culture.

An NGS model for PGS

RHS DOPlify™ whole genome amplified single cell mapped reads per chromosome were normalized to DOPlify™ amplified euploid 46,XX reference single cells. The test single cells derived from the aneuploid cell line were confirmed to have elevated mapped read ratios for chromosomes 2, 21 and Y and a reduced mapped read ratio for the X chromosome compared to the female reference following MiSeq NGS analysis (Figure 2), confirming the chromosomal content of the cell line (48,XY,+2,+21) and providing full concordance with the RHS EmbryoCellect™ aCGH results.

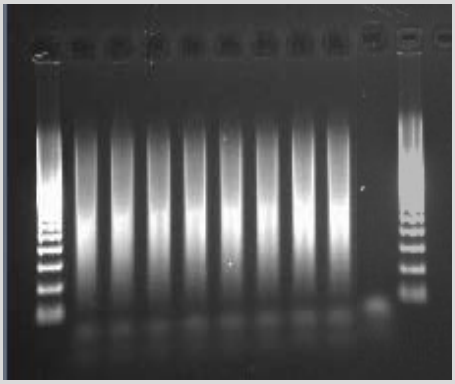

Figure 1. Example of DOPlify™ WGA PCR products.

Lane 1 contains a DNA size marker ladder (DMW-100M, GeneWorks, Australia), lanes 2-5 are WGA amplified single cells, lanes 6-9 are 30pg reference 46,XY genomic DNA amplified templates and lane 10 is a no template control where no DNA was added to the reaction.

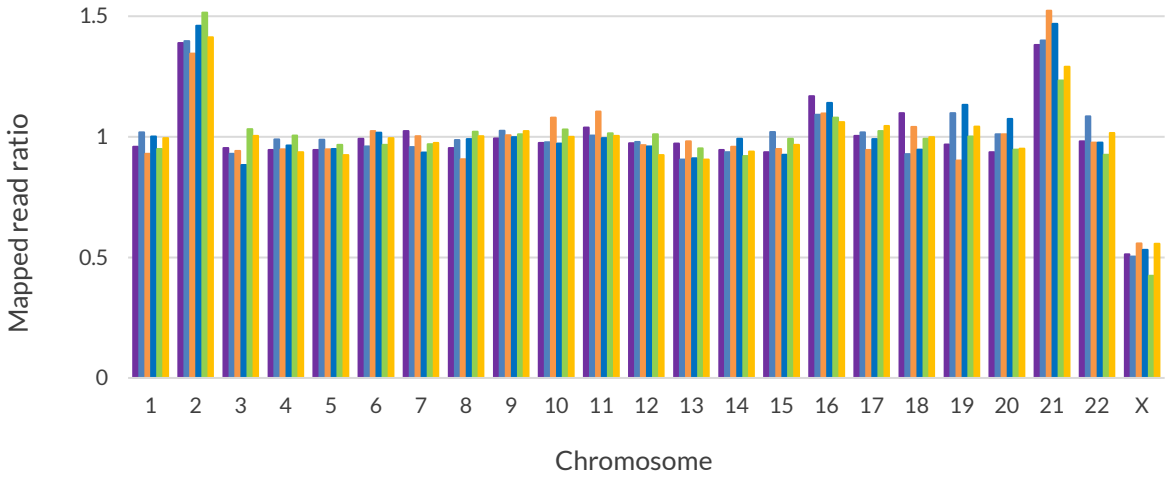

Figure 2; 48,XY,+2,+21 single cell ratios of mapped reads per chromosome normalised to euploid single cells (46,XX) following MiSeq sequencing. Data for the Y chromosome is not shown due to the aneuploid cell line being male and the euploid reference being female causing the ratio to be in excess of 3.

Single cell Mitochondrial DNA content

Cellular mitochondrial DNA (mtDNA) load has been suggested as an additional biomarker of embryo viability. An elevation of mtDNA has been observed in aneuploid embryos and also in 30% of non-implanting euploid embryos (Fragouli E et al 2015). The content of mtDNA a cell contains can be quantitatively assessed using qPCR or NGS technology.

The MiSeq data generated following the sequencing of single cell RHS DOPlify™ DNA libraries was observed to contain DNA reads that mapped to the 16,569bp mitochondrial genome. The percent of total reads mapping to the mtDNA genome were significantly higher in aneuploid single cells compared to the euploid single cells. Some cell to cell variability in the percentage of mapped reads was observed. The single cells were manually sorted from single culture flasks.

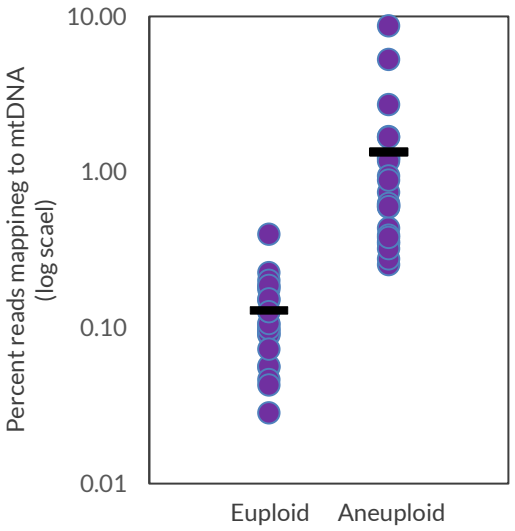

Figure 3. Single cell Mitochondrial DNA content determined using MiSeq NGS of RHS WGA single cell libraries. (euploid cells 0.13±0.08 n=23, aneuploid cells 1.34±1.91 n=24 ;Student's T-test; p<0.01. Black bar=mean).

Detection of Genes of Clinical Interest

Whilst the low pass sequencing performed for the assessment of aneuploidy (involving multiplexing of 49 samples on the MiSeq and 23 samples on the NextSeq) is not recommended for the parallel detection of genes of clinical interest, RHS investigated the probability of this NGS data being used for gene detection.

Both MiSeq and NextSeq data sets were screened for sequence reads that aligned to genes of clinical interest listed in 1) the ISCA Panel, which contains 377 genes (International Standards for Cytogenomic Arrays Consortium) and 2) the Newborn Screening Panel (2013), which contains 72 genes (American College of Medical Genetics; Genetic Med 2006).

Of all the genes investigated, not all were detected in the single cell or unamplified gDNA sample sequencing data confirming that low pass NGS requires modification for the combined use of PGS and PGD. This may involve subsequent specific PCR on the WGA products, concurrent amplification of the gene of interest within the WGA program or targeted specific enrichment of the WGA products during NGS library preparation.

For the purposes of this study, genes were determined to be present in the sequencing data set if at least 1 read was identified to align with the hg19 code for the gene of interest (Table 2). It is presumed that this read indicates that the gene is at least partially represented in the DOPlify™ WGA products and that there may be sufficient representation to enable the use of subsequent enrichment strategies to at the very least provide phasing data to detect the presence or absence of the disease carrying gene.

ISCA Panel: There were 324/377 (85.9%) genes identified in the unamplified gDNA sequencing data generated on the MiSeq compared to 374/377 (99.2%) genes identified in sequencing data generated on the NextSeq sequencer. This demonstrates that additional NGS data per sample obtained from the NextSeq (1.3 million reads per sample vs 5.8 million reads per sample) will increase the probability of detecting genes of clinical interest. It also demonstrates that a long read length (300bp) on the MiSeq will provide some gene specific sequencing data even when 49 samples are multiplexed suggesting that it is not implausible that incidental findings will occur with NGS analysis.

For the single cell DOPlify™ products, 338 ± 3.4 genes (mean ± SD; 89.7%) were identified in single cell samples (n=5) on the NextSeq and 277 ± 29 on the MiSeq. Considering that there was no enrichment for genes of clinical significance, there is a high proportion of genes where at least some sequence data is available. There was also a greater standard deviation for the replicates from the MiSeq (Table 3), suggesting that sample to sample variation is higher with lower resolution NGS and that there is a higher likelihood that chance plays a part in whether a gene is present or not within the NGS library and sequencing data.

Newborn Screening Panel: There were 70/72 (97%) genes in the Newborn Screening Panel identified in the unamplified gDNA sequencing data generated on a MiSeq compared to 72/72 (100%) genes identified in sequencing data generated on the NextSeq sequencer. All 72 genes (100%) were identified in the DOPlify™ single cell samples from the NextSeq (n=5) but only 49 ± 3 were detected from the MiSeq.

Since there is a higher likelihood that a read will be detected in larger genes than in smaller ones, the number of nucleotide bases detected in single cell (SC) and unamplified gDNA NGS data for each gene was plotted against overall gene size as annotated in hg19 (Figure 4). The number of mapped nucleotides in both the unamplified gDNA and in the PCR products from single cells following WGA correlated to the gene size; the larger the gene, the more nucleotides mapped to it.

Further investigation of the representation of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) and Breast Cancer 1, Early Onset (BRCA1) was undertaken (Figure 5).

The combination of WGA of low starting template and multiplexing of libraries for NGS results in a low coverage of the genes. Cell to cell variability was also observed, predominantly influenced by the number of mapped reads varying between individual replicates.

Table 2. Number of genes where at least one read mapped to a gene listed in the panel.

Sample	# samples multiplexed	ISCA genes detected per sample (n=377)	Newborn Screen genes detected per sample (n=72)
Single cells MiSeq (mean, n=4)	49	277 ± 29 (73%)	49 ± 3 (68%)
gDNA MiSeq	49	324 (86%)	70 (97%)
Single cells NextSeq (n=5)	24	338 ± 3.4 (90%)	72 (100%)
gDNA NextSeq	24	374 (99%)	72 (100%)

Table 3. Summary of incidental presence of genes.

	NextSeq (N=4)		MiSeq (N=5)	
	ISCA (377)	Newborn (72)	ISCA (377)	Newborn (72)
# of genes excluded in 1 single cell	57	12	23	2
# of genes excluded in 2 single cells	30	12	9	5
# of genes excluded in 3 single cells	35	7	10	0
# of genes excluded in 4 single cells	45	9	11	0
# of genes excluded in all 5 single cells			16	0

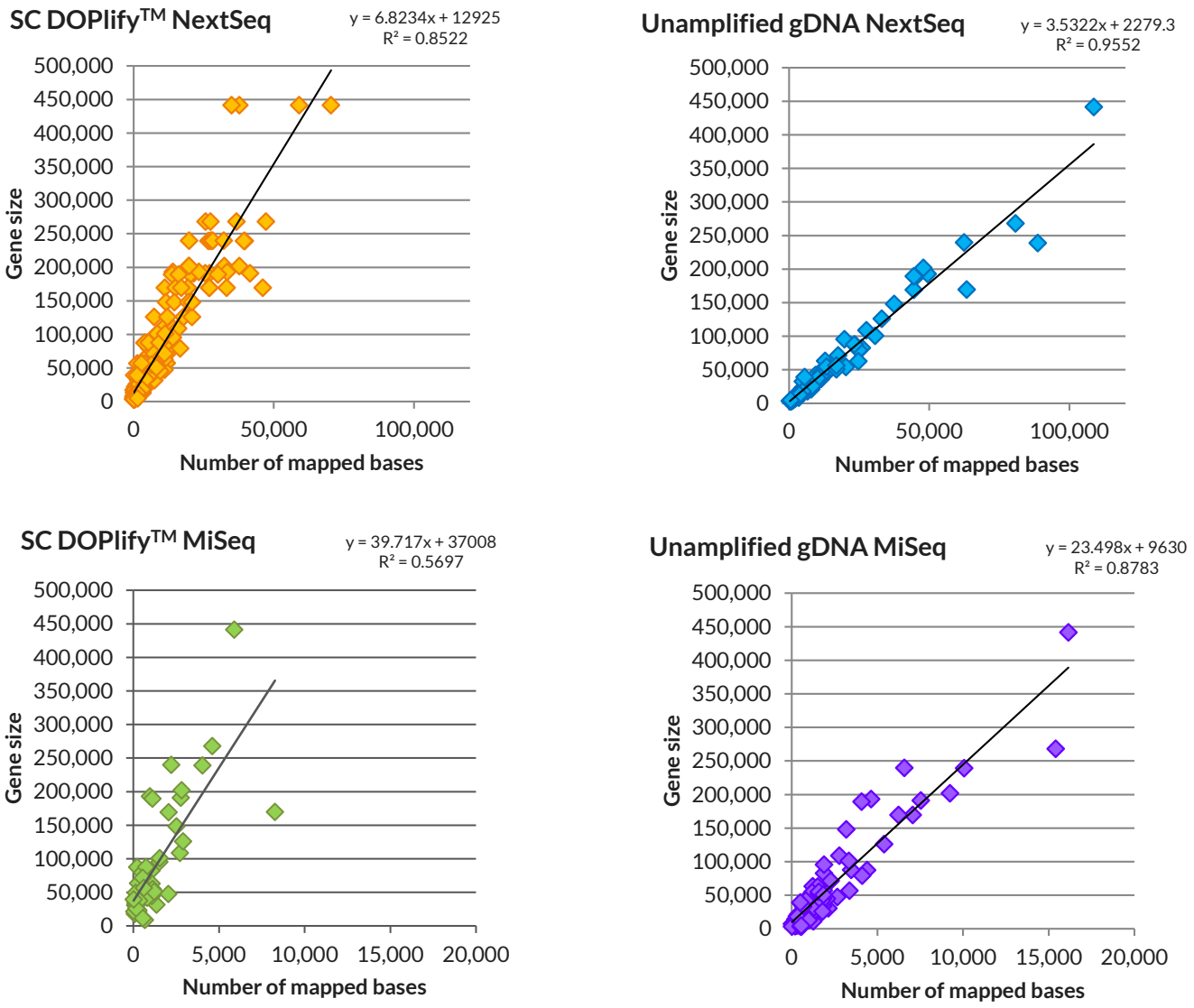

Figure 4; Number of mapped bases compared to total gene size for the Newborn Screening panel genes.

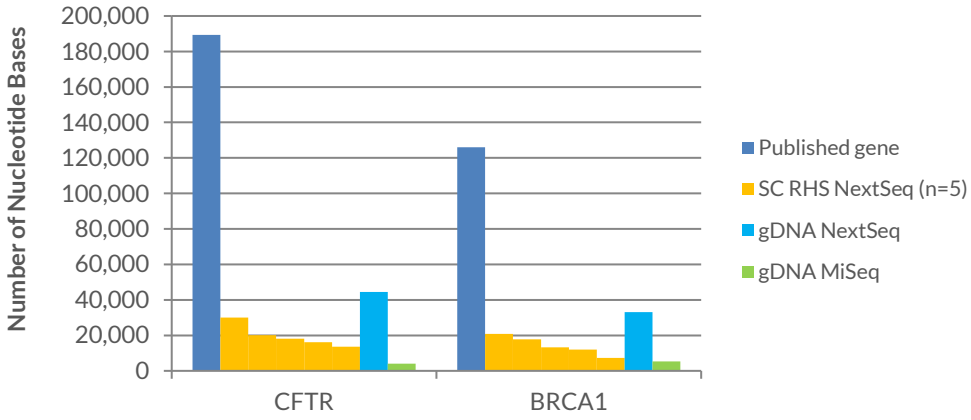

Figure 5. Number of nucleotide bases in the sequencing data that mapped to the gene compared to the total length of the gene (purple). Cystic fibrosis transmembrane conductance regulator (CFTR) and breast cancer 1, early onset (BRCA1).

Conclusions

DOPlify™ provides cell ploidy status of single cells and simultaneously cell mitochondrial DNA load using a range of different NGS platforms and workflows. The amount of sequencing data available for bioinformatic interrogation is highly dependent on the number of samples multiplexed in a library, the read length used and the platform configuration. It is also likely to be influenced by the library preparation kit. Each of these parameters impact on the DNA library complexity.

While specific gene sequence is present in single cell NGS data, the breadth and depth of coverage required for PGD is not achievable using current PGS workflows. However, there is the possibility that low pass NGS will uncover incidental finding if the data is sufficiently interrogated. The ability to enrich for genes of clinical significance either during WGA, after WGA on the amplified product or during NGS library preparation will be a key indicator for PGS + PGD success.

References

Fragouli E et al. Altered levels of mitochondrial DNA are associated with female age, aneuploidy, and provide an independent measure of embryonic implantation potential. PLoS Genet. 2015 Jun 3;11(6):e1005241.
Illumina Website
ThermoFisher Scientific Website

Reproductive Health Science Ltd

40-46 West Thebarton Road Thebarton South Australia 5031 AUSTRALIA ASX:RHS
info@rhsc.com.au | www.rhsc.com.au