PG-Seq™ A novel complete NGS solution for PGS and PGD

Maximising clinical NGS data:

The effect of WGA & sequencing read length

The amount of Next Generation Sequencing (NGS) data generated from single or small cell number DNA samples is highly influenced by the choice of Whole Genome Amplification (WGA) kit and sequencing kit. The mitochondrial genome (mtDNA) represents an ideal model to evaluate the effect of WGA and NGS methods on sequencing results due to its small size, high copy number and the consequent read depth obtainable from low pass sequencing. The mtDNA results provide insight into genome wide effects expected with more expensive deeper sequencing of the entire human genome.

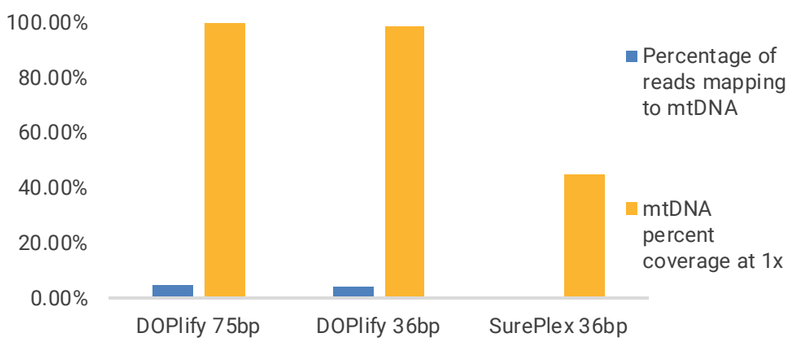
Aim – To compare the effect of two WGA and NGS methodologies on mtDNA sequence data.

Methods – Trophectoderm biopsies were amplified with either SurePlex (Illumina) or DOPlify™ (RHS) WGA kits and sequenced using standard VeriSeq (Illumina: 1x36bp, 24 sample multiplex) and PG-Seq™ (RHS: 1x75bp, 48 sample multiplex) protocols. Bioinformatics was used to truncate the 75bp read lengths from PG-Seq™ to 36bp. All samples were aligned to hg19.

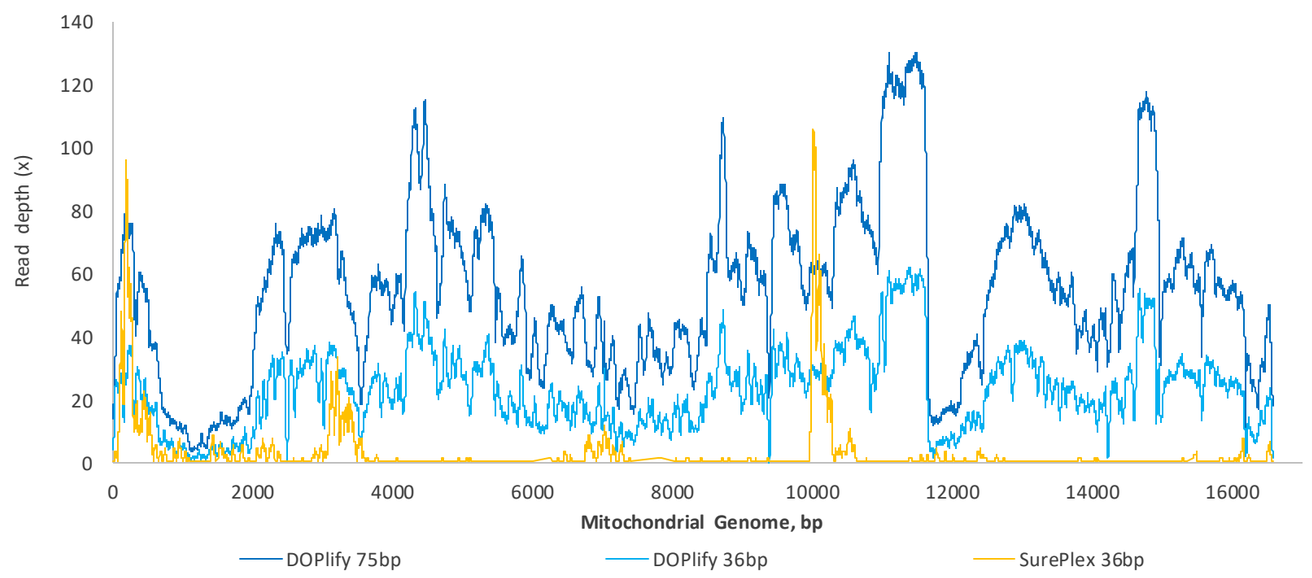
Results – As expected, the average reads per sample reduced with PG-Seq™ due to double the number of samples being multiplexed. Reducing the read length to 36bp also decreased the number of mapped reads due to the increase in multiple mapping locations with the shorter read length. Despite the difference in number of mapped reads with each protocol, the ploidy of the samples was still readily determined.

	PG-Seq™ 75bp	PG-Seq™ 36bp	VeriSeq 36bp
# samples multiplexed	48	48	24
# reads mapped to hg19	637,000	553,000	1,162,000

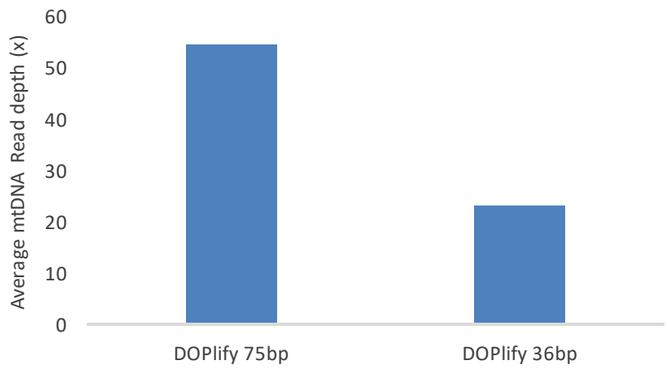
Interestingly, DOPlify™ samples sequenced with 36bp read lengths generated 28x more mtDNA mapped reads compared with SurePlex at 36bp.



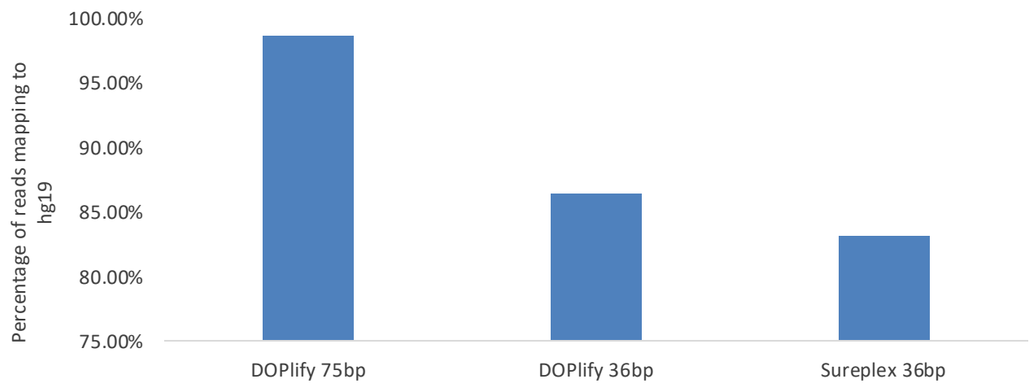
Not only were there more mapped reads, but there was greater than 50% broader coverage of the mtDNA genome using PG-Seq™ when compared to VeriSeq, even when the PG-Seq™ read length was reduced to the same 36bp as VeriSeq. This suggests that the difference is due to the Whole Genome Amplification method, not the sequencing workflow.



Reducing the read length from 75bp to 36bp in the PG-Seq™ samples caused a significant reduction in average read depth across the mtDNA genome with more than 50% of read depth data being lost.



The percentage of reads mapping to the human genome using PG-Seq™ was significantly higher when analysed with 75bp samples vs 36bp. Only ~8,400 reads were not mapped when multiplexing 48 samples in a 75bp run compared to ~88,000 unmapped reads for a 36bp run. The shorter read length reduced the uniqueness of the reads allowing more than 10x the number of reads to not map uniquely.



Conclusions

Both the WGA methodology and sequencing read length significantly impacts the amount of data generated by NGS. Using DOPlify™ at 75bp in a 48 sample run provides significantly broader mitochondrial genome coverage and increased average read depth compared to either DOPlify™ at 36bp (48 sample run) or SurePlex at 36bp (24 sample run).

Kimberly Warren[1], Volkan Baltaci[2], Evrim Unsal[2], Suleyman Aktuna[2], Merve Aydin[2], Michael Barry[3], Melinda Jasper[1]

1 RHS Ltd, Australia | 2 Mikrogen Diagnosis Laboratory, Turkey | 3 FertilitySA, Australia