The application of a hybridisation-based next-generation sequencing enrichment panel for the analysis of somatic variants in tumour samples and reference standards

Jacqueline Chan[1], Lyudmila Georgieva[1], Sabine Eckert[1], Faidra Partheniou[1] and Graham Speight[1]

[1]Oxford Gene Technology (OGT), Oxford, UK

Introduction

- Breast and Ovarian cancers are some of the most common cancers in women.

- Next-generation sequencing (NGS) has enabled the simultaneous study of mutations in high penetrance breast cancer predisposition genes.

- These include *BRCA1*, *BRCA2*, *TP53*, *PTEN*, and *PIK3CA*, as well as more moderate risk genes such as *PALB2*, *BRIP1*, *RAD51C* and *RAD51D*.

Using OGT's extensive background in bait design we have developed a range of fully tested and optimised baits targeting all coding exons of a range of key cancer-related genes (Table 1).

APC	BRCA2	CHEK2	GATA3	PIK3CA	RB1
ATM	BRIP1	EGFR	MSH6	PTEN	SF3B1
BARD1	CDH1	ERBB2	NBN	RAD51C	STK11
BRCA1	CDK12	ESR1	PALB2	RAD51D	TP53

Table 1: Key breast and ovarian cancer-related genes with empirically tested bait sets available in the SureSeq myPanel™ range.
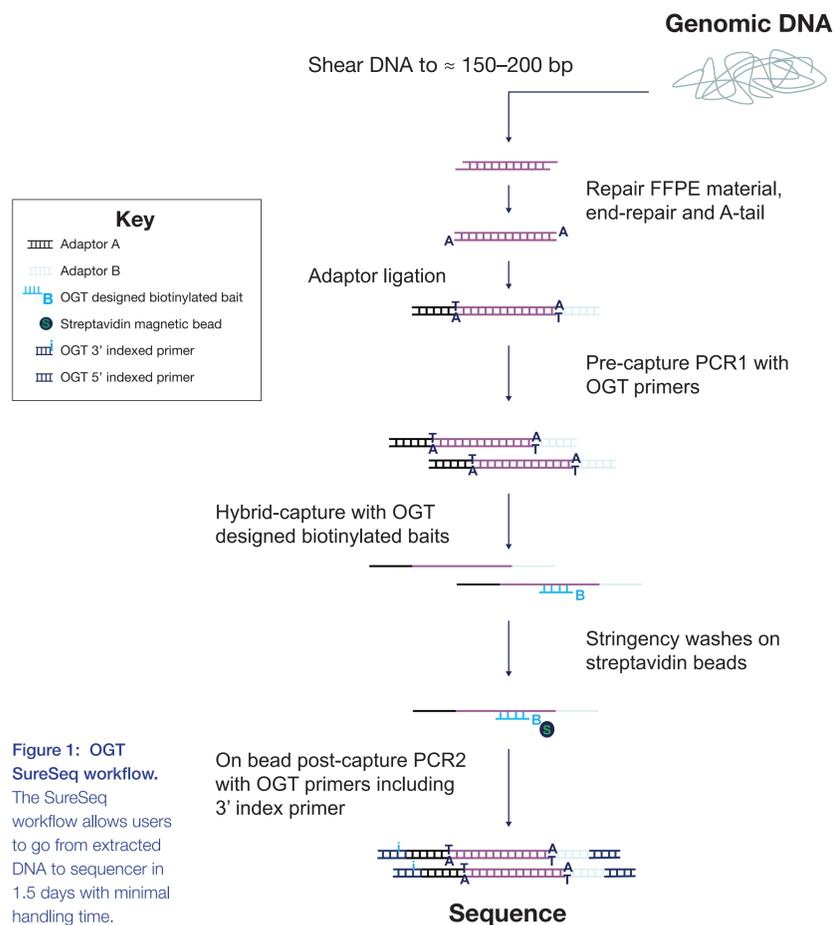
To evaluate the application of a hybridisation-based approach we:

- Compared the uniformity of coverage between a PCR amplification-based and the SureSeq™ hybridisation-based enrichment approach for *BRCA1* and *BRCA2* in solid tumour samples*.

- Assessed the performance of a custom panel (*ALK*, *KIT*, *EGFR*, *KRAS*, and *TP53*) from the SureSeq myPanel NGS Custom Cancer Panel range using the Quantitative Multiplex Reference Standard – gDNA and formalin-compromised DNA, from Horizon.

SureSeq hybridisation workflow

The SureSeq hybridisation-based approached was used throughout this study; the workflow of this is outlined in Figure 1.

Use of target-capture allows the removal of PCR duplicates which can obscure the minor alleles present within a sample.

Key
- Adaptor A
- Adaptor B
- OGT designed biotinylated bait
- Streptavidin magnetic bead
- OGT 3' indexed primer
- OGT 5' indexed primer

Genomic DNA

Shear DNA to ≈ 150–200 bp

Repair FFPE material, end-repair and A-tail

Adaptor ligation

Pre-capture PCR1 with OGT primers

Hybrid-capture with OGT designed biotinylated baits

Stringency washes on streptavidin beads

On bead post-capture PCR2 with OGT primers including 3' index primer

Sequence

Figure 1: OGT SureSeq workflow. The SureSeq workflow allows users to go from extracted DNA to sequencer in 1.5 days with minimal handling time.

Hybridisation-based enrichment generates highly uniform coverage of key targets

To confidently call low level variants, NGS reads need to be evenly distributed across all regions of interest. Uniformity of coverage is a useful value with which to compare this distribution and can be expressed as the percentage of target bases that have greater than 20% of the mean coverage.

As reported extensively in the literature[1-3], the uniformity of coverage from captured-based approaches such as SureSeq consistently outperforms those enriched using an amplicon method (Figure 2). Furthermore, in our sample set we found the high levels of uniformity are maintained when starting with ~250 ng DNA (light blue bars).

The uniformity of coverage for most samples is greater than 99% of bases covered at 20% of the mean, ensuring that all bases within the panel can be confidently assessed.

Uniformity of coverage using amplicon and hybridisation capture-based approaches
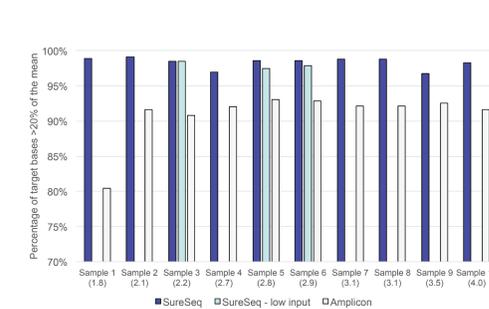

Figure 2: Assessment of the uniformity of coverage from FFPE-derived DNA using an amplicon and the SureSeq hybridisation capture-based approaches. Enrichment by SureSeq sequence capture (dark blue bars) demonstrates better uniformity than that of an amplicon-based approach (white bars). The level of uniformity is maintained when starting with ~250 ng DNA (light blue bars). Samples are ordered by increasing DNA Integrity Number (DIN) determined by Agilent 2200 TapeStation – value in brackets.

The superior uniformity of coverage of key exons in *BRCA1*, and *BRCA2* from Sample 6 with SureSeq compared to an amplicon-based panel is shown in Figure 3. Regions of low complexity such as repetitive sequences and high/low GC content often appear skewed in NGS data. However, as a hybridisation capture-based approach will also provide data from the flanking regions adjacent to the target the coverage profile is "smoother" than seen in panel iii. This enables reliable identification of somatic single nucleotide variants (SNVs) and indels in solid tumour samples.
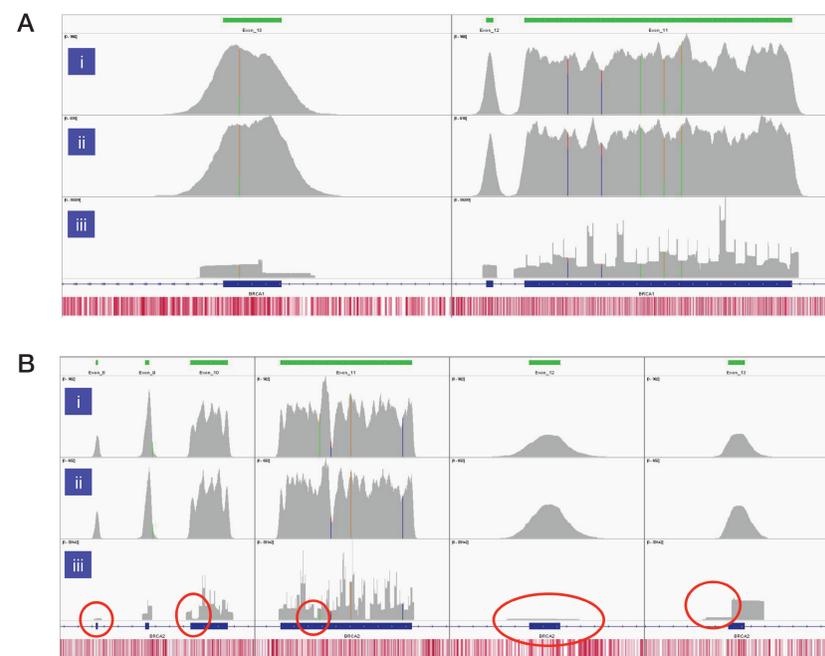

Figure 3: Assessment of the uniformity of sequencing coverage from FFPE-derived DNA using an amplicon-based and the SureSeq hybridisation-based capture approaches. Integrated Genomics Viewer[4] images comparing of depth of coverage of A: *BRCA1* exons 11-13, and B: *BRCA2* exons 8-13. Panels i and ii - SureSeq Ovarian panel starting with 600 and 300 ng of FFPE-derived DNA respectively. Panel iii – data from an amplicon-based approach. Areas circled in red highlight target regions of very low coverage in the amplicon-based panel. Depth of coverage per base – grey; targeted region – green; gene coding region as defined by RefSeq – blue; GC percentage – red.

Accurate detection of variants from reference standards

The high mean target coverage and uniformity of coverage was maintained across the two sample types and different starting amounts of gDNA (Table 2) which ranged from 1000 to 45 ng of DNA.

Sample Type	Starting amount (ng)	DIN	Mean target coverage	Percentage of regions covered at 20% of mean
gDNA	285	8.3	1906[a]	99.16
gDNA	100	6.8	1294[a]	99.16
gDNA	100	6.8	1276[a]	99.11
gDNA	45	5.8	647[a]	99.46
Formalin-compromised DNA - mild	1000	6.1	1364[b]	99.82
Formalin-compromised DNA - moderate	1000	2.9	1208[b]	99.70

Table 2: Performance of custom SureSeq myPanel with characterised Horizon samples. Sequencing was conducted on a MiSeq® using a v2 300 bp cartridge (Illumina). [a]Data generated from a 16 sample run; [b]data generated from a 32 sample run.

All samples had 100% concordance for reported SNVs and all allele frequencies were within 5% of the expected value (Figure 4).
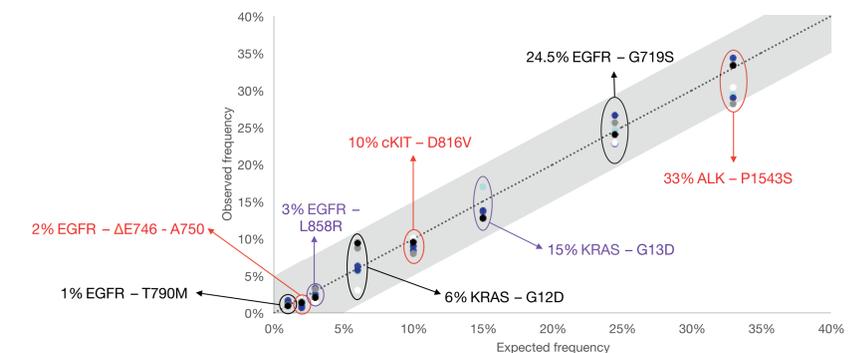

24.5% EGFR – G719S
10% cKIT – D816V
33% ALK – P1543S
2% EGFR – ΔE746 - A750
3% EGFR – L858R
15% KRAS – G13D
1% EGFR – T790M
6% KRAS – G12D

Figure 4: Comparison of the expected and observed allele frequencies from characterised samples. The variants – 7 SNVs and one deletion, ranging from 1 to 33%, were determined using OGT's SureSeq Interpret™ software. Shaded area denotes +/-5% of expected allele frequency, dotted line - guideline.

Conclusions

- Superior uniformity of coverage from a hybridisation-based enrichment using a SureSeq Cancer Panel.

- High levels of uniformity are maintained across a range of starting DNA amounts in both formalin-compromised and genomic DNA.

- 100% concordance in variant detection in both genomic and formalin-compromised DNA.

- Accurate detection of low frequency variants, (sub 5%), from as little as 45 ng of DNA.

- The SureSeq hybridisation-based approach is a robust and reproducible method for the identification of somatic variants from genomic DNA and FFPE tumour samples.

Visit www.ogt.com for more information.

References

1. García-García, G., Baux, D., Faugère, V., Moclyn, M., Koenig, M., Claustres, M., & Roux, A. F. (2016). Assessment of the latest NGS enrichment capture methods in clinical context. *Scientific reports*, 6.

2. Samorodnitsky, E., Jewell, B. M., Hagopian, R., Miya, J., Wing, M. R., Lyon, E., ... & Roychowdhury, S. (2015). Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Human mutation*, 36(9), 903-914.

3. Xu, Y., Jiang, H., Tyler-Smith, C., Xue, Y., Jiang, T., Wang, J., ... & Wang, J. (2011). Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome biology*, 12(9), R95.

4. Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 14(2), pp.178-192.