SureSeq™ FFPE DNA Repair Mix* and hybridisation-based enrichment provide superior variant detection in next-generation sequencing even from 10 ng of severely formalin compromised DNA

Introduction

With advances in next-generation sequencing (NGS) technology, genetic information can be extracted from an increasingly diverse range of samples. Formalin-fixed, paraffin-embedded (FFPE) storage is a standard method for archiving tissue biopsies and these biopsies can be used in NGS, for example to study cancer development and progression.

However, the quality of DNA in FFPE samples is often severely damaged and compromised compared to other methods of sample storage. Consequently, it may be difficult to distinguish between true low frequency mutations and damage-induced low frequency false positives. In these samples, hybridisation-based target enrichment (where input DNA has been sheared into short, fragments and selectively captured) provides superior performance compared to amplicon-based enrichment. This is due to the superior tolerance of the hybridisation approach for the fragmented DNA routinely found in FFPE samples. Hybridisation-based enrichment also provides greater uniformity of coverage, fewer false positives and superior variant detection due to use of fewer PCR cycles[1].

To improve the sequencing results of FFPE samples even further, a DNA pre-treatment step can be introduced to address different types of DNA damage. OGT's SureSeq FFPE DNA Repair Mix uses a mixture of enzymes to repair a range of DNA defects including deamination of cytosine, nicks and gaps, oxidised bases, and blocked 3' ends.

In this study, carried out in collaboration with Horizon Discovery, formalin-compromised DNA (fcDNA) samples of differing severity were repaired with the SureSeq FFPE repair mix and sequenced using a SureSeq custom hybridisation-based panel. The aim of the study was to investigate the effect of the repair mix and also the performance of the hybridisation-based enrichment method on DNA with varying levels of damage by measuring critical parameters at different points in the sequencing protocol:

- DNA length distribution after extraction
- DNA yield immediately before capture
- Mean target coverage in sequencing
- The ability to accurately detect known variants with excellent uniformity

When studying FFPE tumour biopsies, low DNA starting material is also a frequently encountered issue. In order to study the effect of this, tests were carried out with varying amounts of DNA starting material to determine sequencing sensitivity when DNA availability is limited.

Materials and Methods

To investigate the effect of the SureSeq FFPE DNA Repair Mix, Horizon Discovery provided three fcDNA samples with ‘mild’, ‘moderate’ and ‘severe’ damage[2]. The samples were investigated in duplicate before and after repair; the amounts of input DNA were 200, 100, 50 and 10 ng (Figure 1). All samples were sheared using a Covaris S220 focused-ultrasonicator and prepared using the SureSeq NGS Library Preparation Kit (cat. no. 500070).

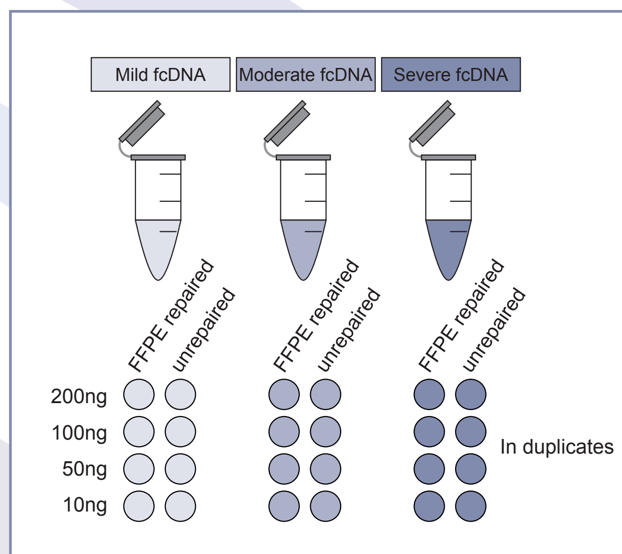


Figure 1: A total of 48 samples were investigated to study the effect of DNA quality, input amount and DNA repair.

Enrichment by hybridisation was completed with an 8.7 Kb custom hot-spot panel designed to target the variants listed in Table 1. The subsequent post-capture libraries were sequenced on an Illumina MiSeq® using a v2 300 cycles kit (cat. no. MS-102-2002). 16 samples were run on a MiSeq lane.

Gene	Variant	Exon	Allelic Frequency
ALK	P1543S	29	33.00%
APC	R2714C	16	33.00%
BRAF	V600E	15	10.50%
BRCA2	A1689fs	11	33.00%
cKIT	D816V	17	10.00%
EGFR	ΔE746 - A750	19	2.00%
EGFR	L858R	21	3.00%
EGFR	T790M	20	1.00%
EGFR	G719S	18	24.50%
FBXW7	G667fs	12	33.50%
*FLT3***	S985fs	24	10.50%
*FLT3***	V197A	5	11.50%
IDH1	S261L	7	10.00%
KRAS	G13D	2	15.00%
KRAS	G12D	2	6.00%
MET	V237fs	2	6.50%
NOTCH1	P668S	12	31.50%
NRAS	Q61K	3	12.50%
PI3KCA	H1047R	21	17.50%
PI3KCA	E545K	10	9.00%

Table 1: List of variants examined. All mutations verified by ddPCR except**.

Results

Improving DNA quality with FFPE repair mix

To confirm the quality of the three samples, DNA integrity numbers (DIN), an indication of DNA quality were determined using the Agilent TapeStation. Both DIN and DNA length distribution confirmed levels of DNA damage (Figure 2).

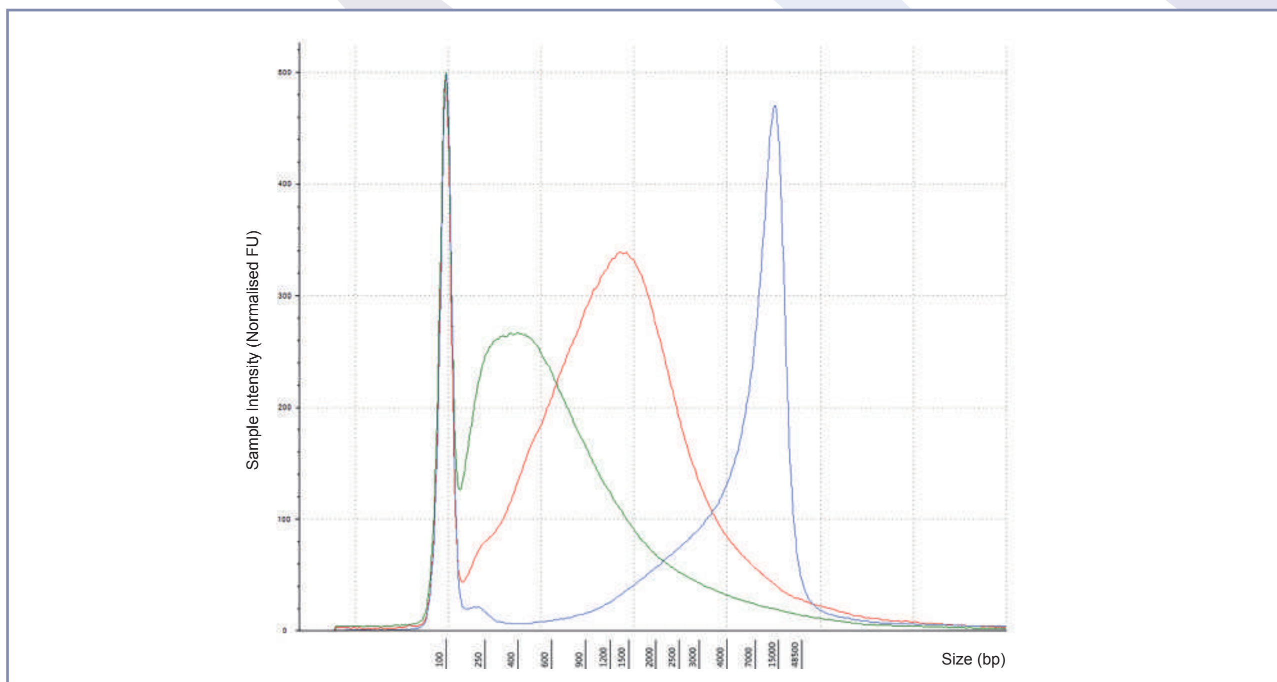


Figure 2: DNA length distribution and DIN can give an indication of the level of DNA damage in a sample. The sample with mild damage had the highest DIN (6.6) (blue), followed by the moderate (3.2) (red) and severe (1.9) (green) samples.

Higher library yields improve the complexity of the library, which in turn leads to better quality sequencing data. Figure 3 shows the effect of DNA repair on pre-capture library yields in the most severely damaged sample with different quantities of input DNA; the mild and moderate samples showed similar results. The difference in peak height at around 200 base pairs demonstrates that more DNA was available for capture and sequencing when the FFPE repair mix was used.

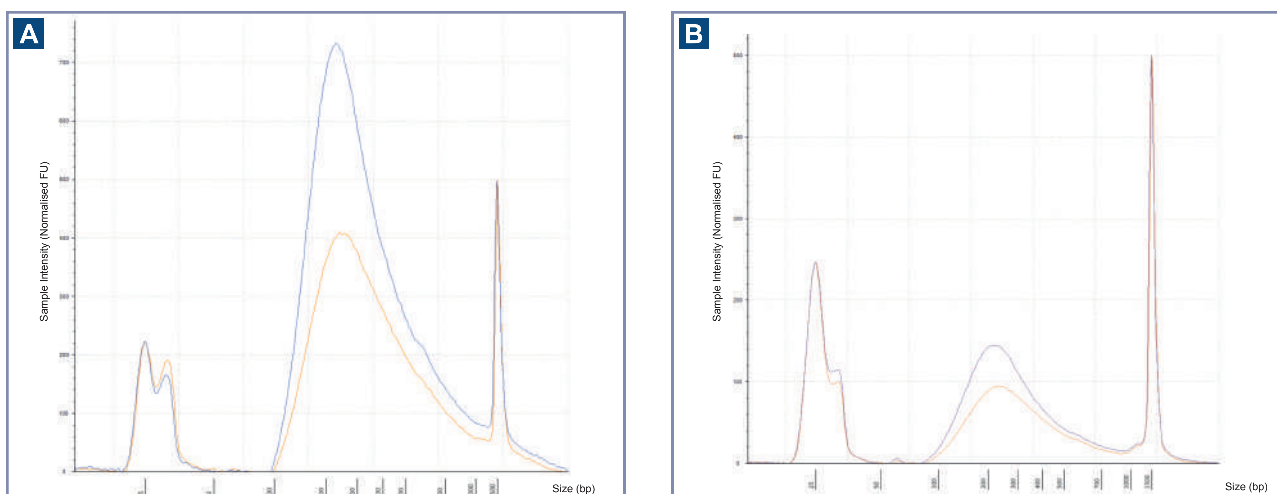


Figure 3: Pre-capture library yields of fcDNA with ‘severe’ damage treated with SureSeq FFPE DNA Repair Mix (blue) and without treatment (orange). The amounts of input DNA were 200 ng **A** and 50 ng **B** .

DNA repair increases mean target coverage and remains high at low amounts of input DNA

The increase in library yield had a positive effect on target coverage in all tested samples (Figures 4 and 5). Improvements in mean target coverage (MTC) varied from 20% up to 50%, showing the positive effect of higher library yields.

DNA treated with the FFPE repair mix also demonstrates a strong sequencing performance at low input amounts, maintaining a MTC (after removal of PCR duplication) of over 1000x for all samples at 100 ng input and over 500x at 50 ng input. This result shows that even when there are only small amounts of DNA available for analysis, it may still be possible to achieve a high level of coverage when using the FFPE repair mix. The result in Figure 5 also shows the direct link between both the quality and the quantity of DNA on variant calling accuracy.

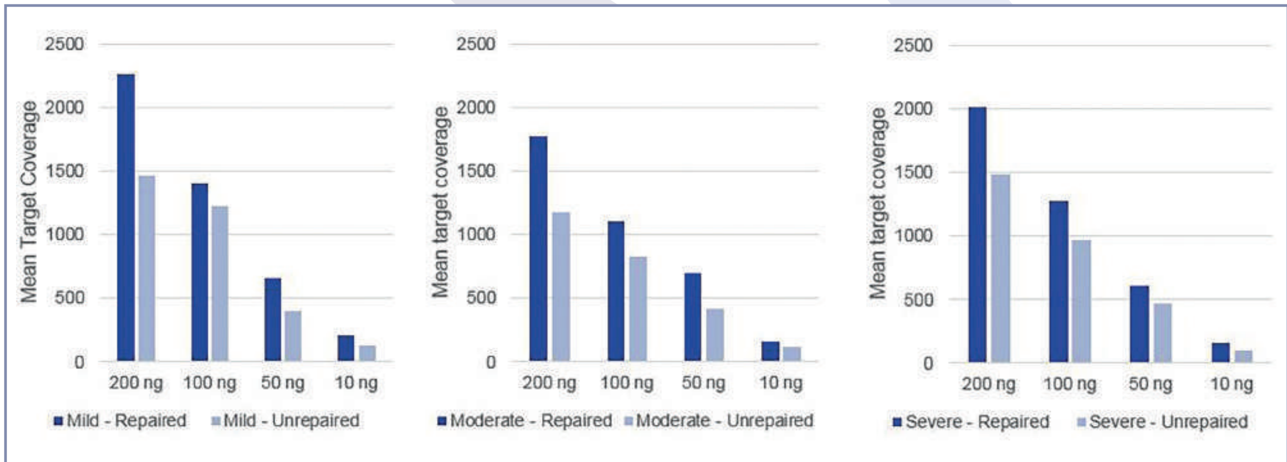


Figure 4: DNA repair significantly improves MTC and helps to improve coverage at all amounts of starting material and all levels of DNA damage; a: mild damage, b: moderate damage, c: severe damage.

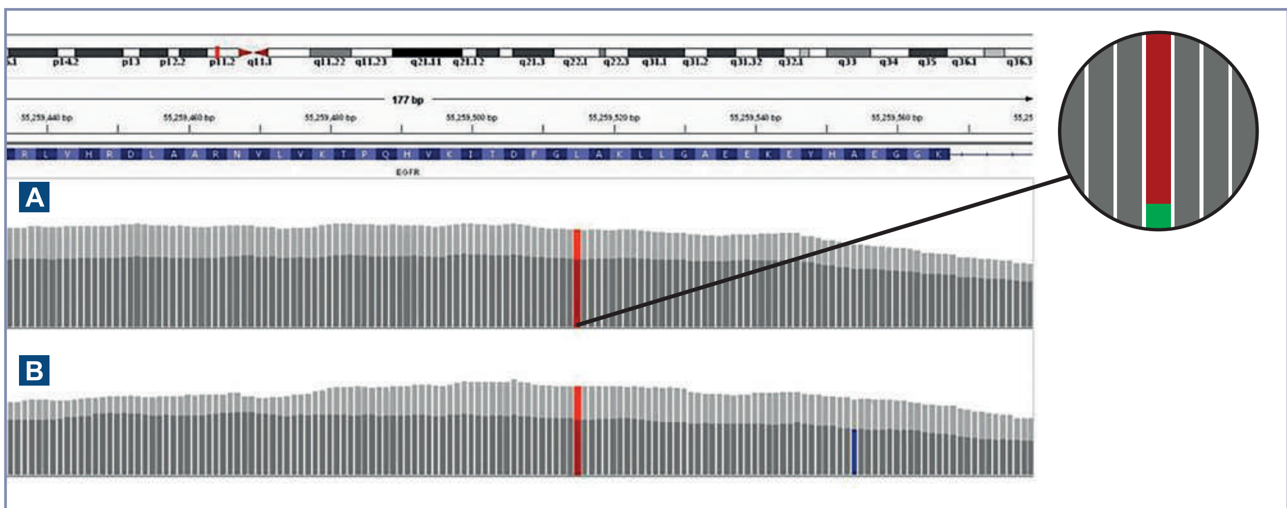


Figure 5: Comparison of the coverage in the region of an *EGFR* L858R mutation, 3% variant allele frequency (VAF), between a severely FFPE compromised sample treated with the SureSeq FFPE Repair Mix (light grey) and one without treatment (dark grey). At 200 ng input, **A**, following treatment, the total number of reads increases from 2707 to 3860, with the number of supporting reads for the mutation (see expanded illustration) increasing from 85 to 112 (after removal of PCR duplication). At 10 ng input, **B**, following treatment, the total number of reads increases from 216 to 341, with the number of supporting reads for the mutation increasing from 5 to 13. Note that the combination of low (10 ng) input and poor DNA quality increases the likelihood of the generation of false positive mutations (blue) and other spurious data. Total read counts are after the removal of PCR duplication.

Reliable detection of low-frequency variants

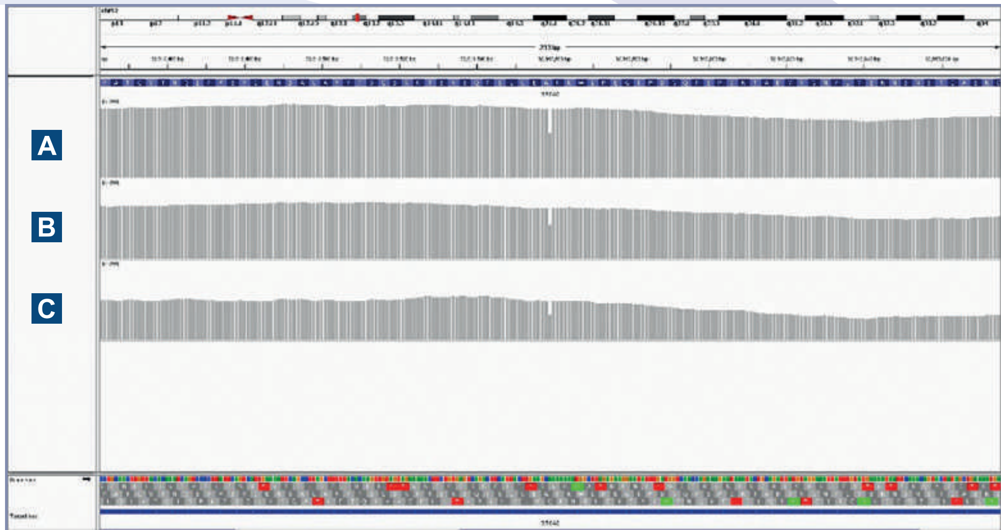
In order to assess the effect of increased MTC on the ability to call variants with greater accuracy the concordance between the NGS results of the fcDNA samples and Horizon Discovery's Reference Standards was investigated using OGT's SureSeq Interpret software. Table 2 details the 20 variants analysed: 15 single nucleotide variants (SNVs) and 5 deletions, with VAF varying between 1% and 33%.

Gene	Variant	Expected frequency	Mild				Moderate				Severe			
			200 ng	100 ng	50 ng	10 ng	200 ng	100 ng	50 ng	10 ng	200 ng	100 ng	50 ng	10 ng
EGFR	**T790M**	**1.0%**	**1.5%**	**1.0%**	**0.5%**	**0.8%**	**0.8%**	**1.2%**	**0.5%**	**0.8%**	**0.9%**	**1.1%**	**1.4%**	**0.9%**
EGFR	**ΔE746 - A750**	**2.0%**	**1.3%**	**1.3%**	**1.8%**	**0.7%**	**1.1%**	**1.8%**	**0.6%**	**ND**	**1.4%**	**1.7%**	**1.4%**	**1.1%**
EGFR	**L858R**	**3.0%**	**3.5%**	**3.9%**	**4.0%**	**2.4%**	**3.2%**	**3.5%**	**3.1%**	**4.3%**	**3.0%**	**3.2%**	**3.7%**	**2.9%**
KRAS	**G12D**	**6.0%**	**5.8%**	**6.4%**	**6.3%**	**6.4%**	**5.3%**	**4.1%**	**6.4%**	**4.2%**	**6.9%**	**5.3%**	**6.5%**	**4.2%**
MET	V237fs	6.5%	5.4%	6.2%	6.7%	5.0%	6.9%	6.7%	4.4%	7.2%	5.4%	6.1%	5.4%	5.4%
PI3KCA	**E545K**	**9.0%**	**9.9%**	**9.8%**	**9.5%**	**8.7%**	**9.2%**	**8.2%**	**9.4%**	**11.0%**	**8.0%**	**6.9%**	**9.6%**	**7.7%**
cKIT	**D816V**	**10.0%**	**8.8%**	**11.4%**	**10.6%**	**11.2%**	**9.1%**	**9.4%**	**9.7%**	**11.9%**	**9.1%**	**7.7%**	**9.5%**	**9.8%**
IDH1	S261L	10.0%	7.3%	8.4%	7.6%	8.7%	7.5%	9.6%	8.5%	6.0%	8.1%	7.6%	8.3%	9.2%
BRAF	**V600E**	**10.5%**	**12.0%**	**12.9%**	**11.3%**	**12.6%**	**11.0%**	**11.1%**	**8.2%**	**4.4%**	**11.7%**	**12.3%**	**9.9%**	**11.1%**
FLT3	S985fs	10.5%	7.1%	9.0%	8.1%	6.5%	7.8%	8.0%	8.4%	38.4%	7.7%	7.7%	7.9%	4.8%
FLT3	V197A	11.5%	8.5%	7.8%	7.1%	6.4%	8.9%	8.4%	9.4%	4.6%	7.8%	7.7%	8.5%	7.5%
NRAS	**Q61K**	**12.5%**	**13.6%**	**15.8%**	**14.7%**	**11.8%**	**11.2%**	**12.6%**	**14.5%**	**17.4%**	**13.2%**	**13.2%**	**14.1%**	**9.7%**
KRAS	**G13D**	**15.0%**	**14.5%**	**14.1%**	**16.4%**	**11.5%**	**14.7%**	**14.4%**	**12.9%**	**13.8%**	**12.8%**	**13.2%**	**13.7%**	**12.9%**
PI3KCA	H1047R	17.5%	16.6%	16.1%	17.1%	19.5%	18.7%	17.8%	21.0%	20.9%	18.9%	17.5%	16.4%	14.6%
EGFR	**G719S**	**24.5%**	**26.3%**	**25.9%**	**26.0%**	**22.0%**	**24.2%**	**26.1%**	**25.8%**	**27.1%**	**24.9%**	**24.9%**	**26.4%**	**24.7%**
NOTCH1	P668S	31.5%	28.5%	28.8%	25.8%	13.4%	25.6%	28.9%	26.8%	22.2%	28.3%	30.0%	27.0%	17.9%
ALK	P1543S	33.0%	31.7%	30.9%	30.3%	25.8%	29.1%	29.1%	31.6%	29.5%	30.3%	32.8%	32.7%	32.4%
APC	R2714C	33.0%	32.6%	30.7%	31.3%	31.1%	31.2%	30.2%	30.6%	26.0%	30.3%	30.0%	26.8%	36.2%
BRCA2	A1689fs	33.0%	33.0%	31.4%	34.3%	27.0%	31.0%	29.6%	26.7%	29.8%	33.9%	31.3%	34.9%	31.8%
FBXW7	G667fs	33.5%	26.5%	28.4%	28.5%	26.2%	29.4%	29.1%	28.9%	19.3%	29.8%	29.9%	30.6%	19.2%

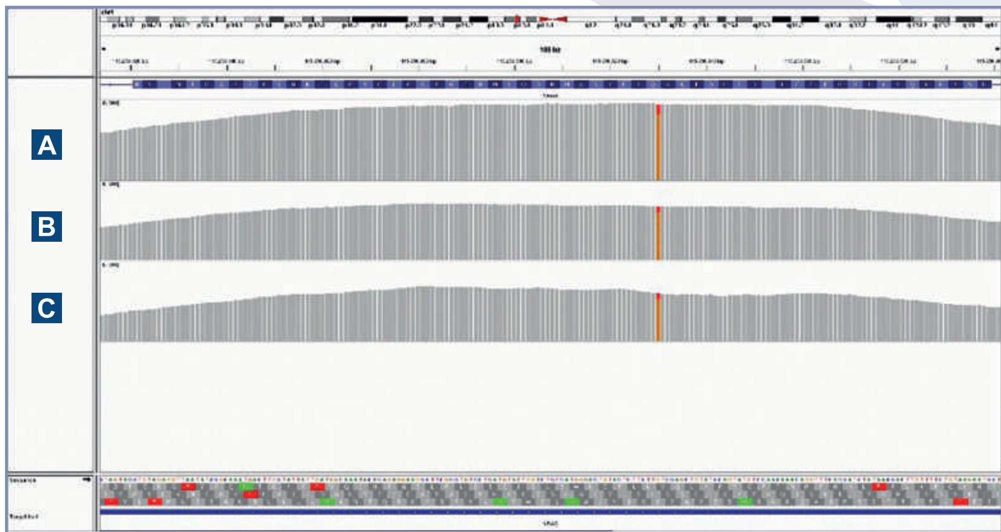
Table 2: VAFs of variants identified. Variants in bold have been confirmed by Horizon Discovery using droplet digital PCR, the presence of the remaining variants have been confirmed in the parental cell line. ND: not detected.

The data in Table 2 show that 99.6% of the expected variants were detected, even in the most severely damaged sample. Furthermore, the measured VAFs are close to the expected coverage with 91.25% of the 240 variants lying within 5 percentage points of their expected values

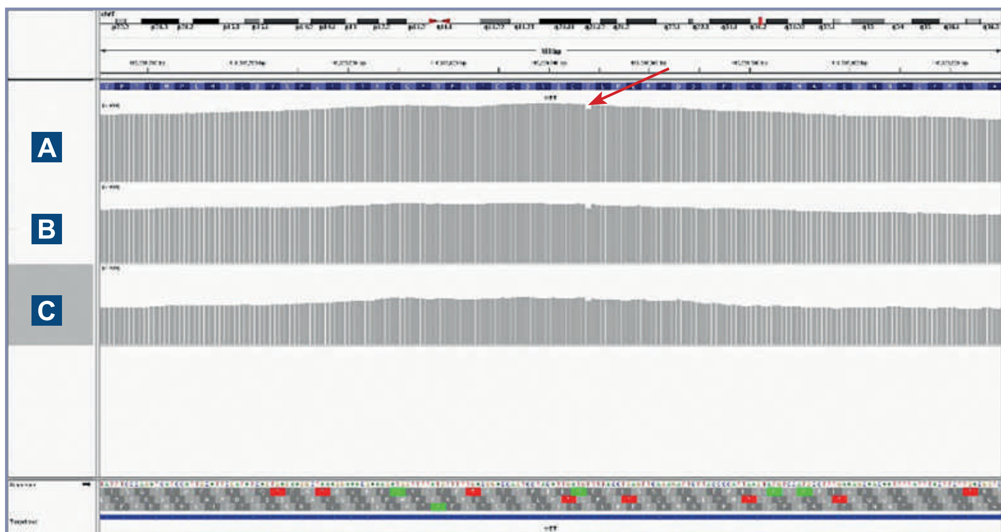
For a more detailed view of the coverage, targeted regions were visualised with the Integrative Genomics Viewer (IGV) software. Using 200 ng of starting material, coverage of one SNV and two deletions are given for each of the three samples (Figure 6); Figure 7 shows the effect of decreasing the amount of input DNA.



BRAC2

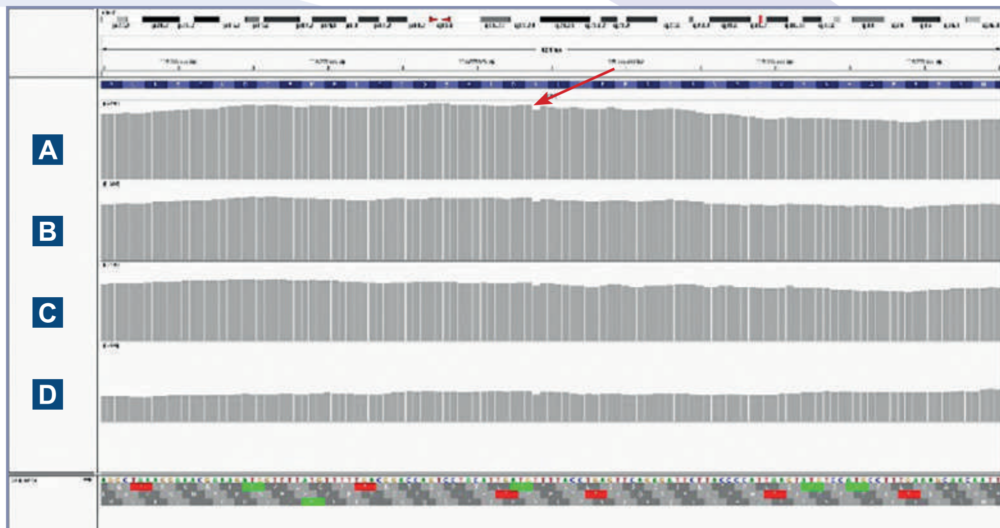


NRAS

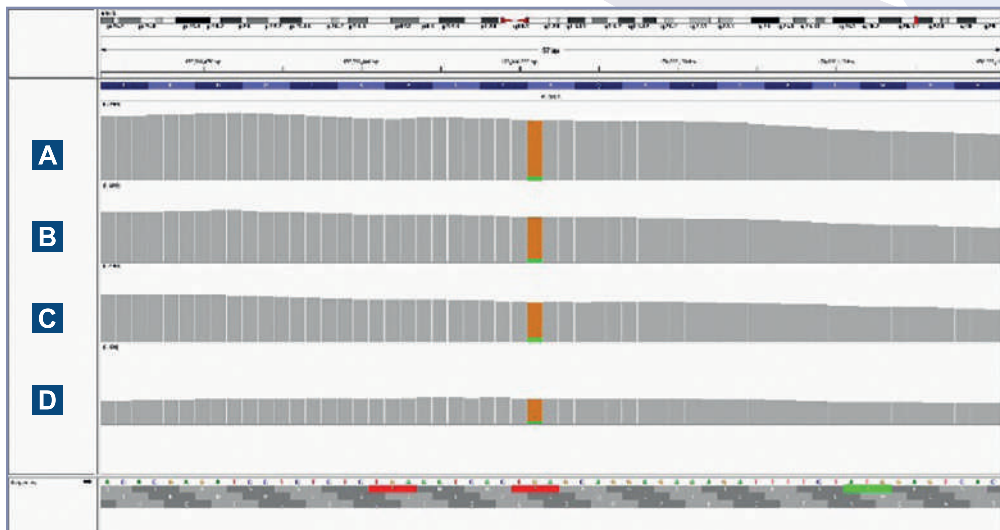


MET

Figure 6: IGV plots of regions of the *BRCA2, NRAS* and *MET* genes with mild **A**, moderate **B** and severe **C** damage, 200 ng input DNA. All variants were detected at all levels of DNA damage. Deletions *(BRCA2, MET)* appear as single-nucleotide drops in coverage; the SNV (NRAS) appears as one line consisting of two colours with each colour representing a different nucleotide.



MET



P13KCA

Figure 7: IGV plots of regions of the *MET* and *PI3KCA* genes of the sample with severe damage. The amounts of input DNA are 200 ng **A**, 100 ng **B**, 50 ng **C** and 10 ng **D**. Variants with a VAF of 6.5% and 9.0% respectively can be detected at all tested concentrations. *MET* - Y-axis set to 2500x, 2000x, 1000x and 500x for 200, 100, 50 and 10 ng respectively. *PICK3CA* - 2000x for 200 and 100ng and 750x for 50 and 10ng respectively.

The plots in Figures 6 and 7 show that using the SureSeq FFPE DNA Repair Mix, reliably detect low-level variants in fcDNA even in as low as 10ng of severely damaged material. It also shows the high uniformity of coverage across the exon that can be achieved by using hybridisation-based enrichment.

Conclusion

In areas such as cancer research, the use of FFPE samples for sequencing has the potential to unlock a vast amount of information from tissue biopsies. The use of SureSeq FFPE DNA Repair Mix in addition to hybridisation-based enrichment has been shown to improve DNA yield and MTC, allowing greater confidence in calling low-frequency variants.

When studying 20 known variants, a 99.6% concordance was reported at all three levels of DNA damage (mild, moderate and severe). Of the measured allele frequencies, 91.25% of values were within 5 percentage points of the expected values. These results highlight that carrying out DNA repair in addition to hybridisation-based enrichment leads to better results in every step of the process – resulting in superior coverage, even when the quality or quantity of the starting material is low.

References

1. *Selecting the best NGS enrichment assay for your needs,* available at: https://www.ogt.com/resources/literature/1357_selecting_the_best_ngs_enrichment_assay_for_your_needs
2. Horizon Discovery catalogue numbers: HD798 (mild), HD799 (moderate) and HD803 (severe). *https://www.horizondiscovery.com/reference-standards/our-formats/formalin-compromised-dna*

For more information about SureSeq myPanel™ custom NGS panels or the SureSeq FFPE DNA Repair Mix, visit www.ogt.com or contact us at products@ogt.com.

*The SureSeq™ FFPE DNA Repair Mix can only be purchased in conjunction with SureSeq NGS panels, not as a standalone product.

Oxford Gene Technology
T: +44(0)1865 856826 (US: 914-467-5285) E: products@ogt.com W: www.ogt.com
Technical support: support@ogt.com

SureSeq: For Research Use Only; Not for Diagnostic Procedures. This document and its contents are © Oxford Gene Technology IP Limited – 2018. All rights reserved. OGT™ and SureSeq™ are trademarks of Oxford Gene Technology IP Limited. The SureSeq NGS Library Preparation Kit was jointly developed between Oxford Gene Technology and Bioline Reagents Limited. MiSeq is a registered trademark of Illumina Inc.