Assessment of the performance of a hybridisation-based NGS enrichment panel with as little as 10 ng of severely formalin-compromised DNA

Jacqueline Chan, Juliette Forster, Aysel Heckel, Venu Pullabhatla, Dave Cook, Graham Speight

Oxford Gene Technology (OGT), Begbroke Science Park, Begbroke Hill, Woodstock Road, Begbroke, Oxford, UK

Introduction

Formalin-fixed, paraffin-embedded (FFPE) storage is a standard method for archiving samples from solid tumours. It ensures the preservation of the ultrastructure of tissues and prevents degradation through formation of chemical links between macromolecules, for example between and within DNA molecules. FFPE samples contain a wealth of information which can be used to study cancer development and progression. Next generation sequencing (NGS) offers the capability of unlocking this information through the simultaneous study of multiple types of mutations in cancer-associated genes for a number of applications[1]. However, formalin treatment can significantly compromise the quality and amount of nucleic acids available for genomics research. As such it is technically challenging to examine the true genetic complexity present in a sample.

In this study DNA reference standards with different levels of formalin-induced damage were hybridised and sequenced with a SureSeq™ custom NGS panel in conjunction with the SureSeq FFPE DNA Repair Mix*. We assessed the impact of the repair mix on three levels of formalin compromised DNA (fcDNA) - 'mild', 'moderate' and 'severe', at 4 different DNA input amounts down to 10 ng. We then compared the uniformity and coverage of the enriched targets. We also assessed the concordance to the allele frequencies of the variants in the reference standards.

Study design

We tested three reference standards of fcDNA (provided by Horizon Discovery[†]) with 'mild', 'moderate' and 'severe' damage[2]. The samples were investigated in duplicate before and after repair with SureSeq FFPE DNA Repair mix; 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).
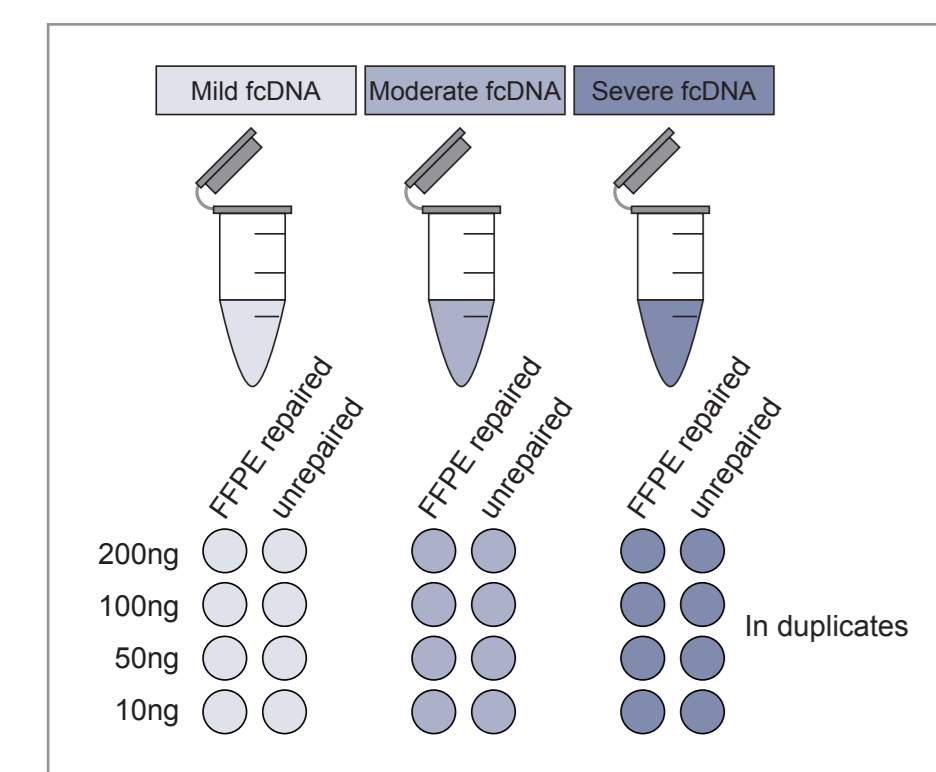
Enrichment by hybridisation was completed with a SureSeq custom NGS 8.7 Kb custom hot-spot panel designed to target the variants present in the reference standards (see Table 1 for variants targeted). The subsequent post-capture libraries were sequenced on an Illumina MiSeq® using a v2 300 cycles kit (cat. no. MS-102-2002). 16 samples were run on a MiSeq lane.

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

Methods

The SureSeq hybridisation-based approached was used throughout this study; the workflow of this is outlined in Figure 2.

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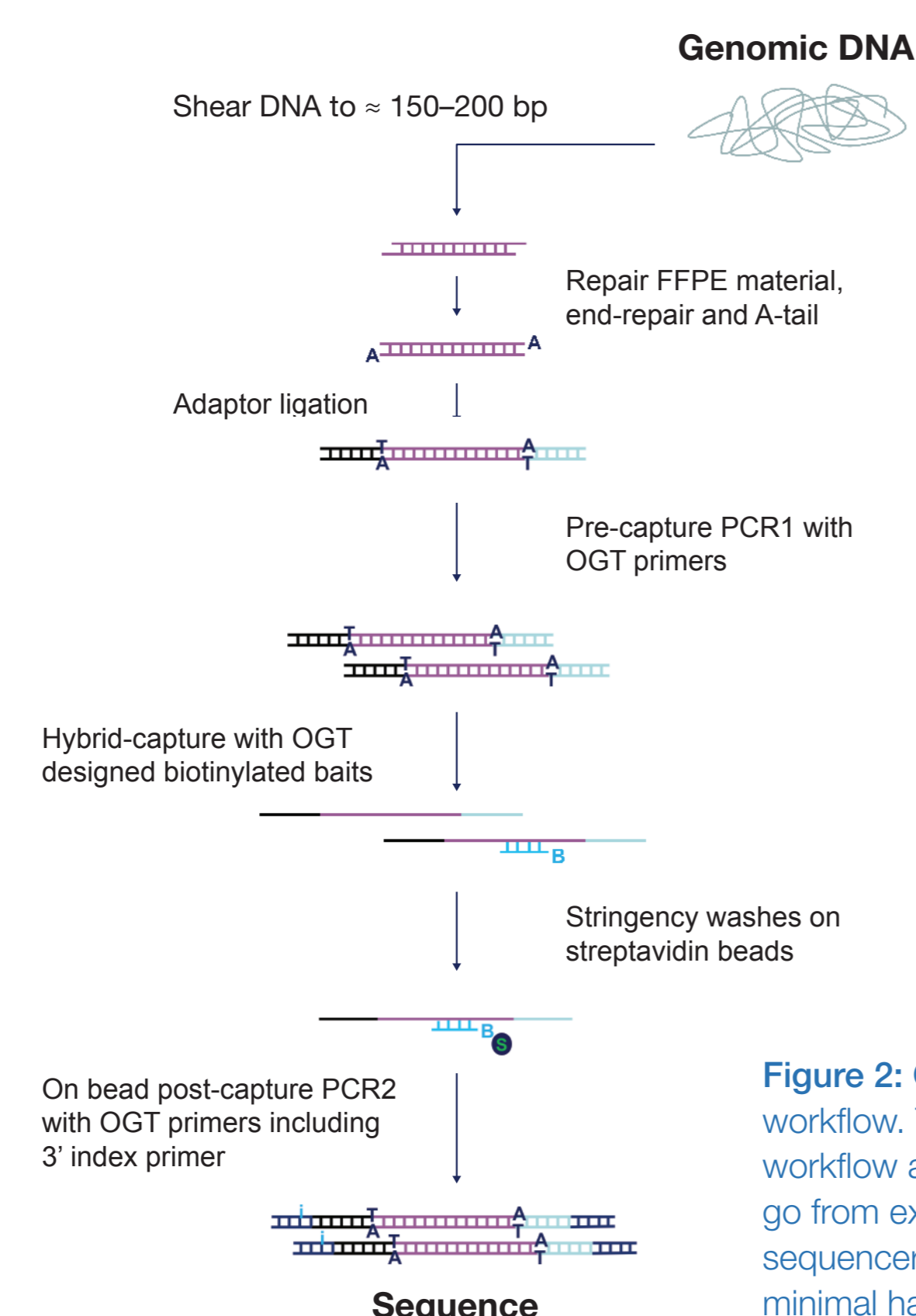

Figure 2: OGT SureSeq workflow. The SureSeq workflow allows users to go from extracted DNA to sequencer in 1.5 days with minimal handling time.

Results

FFPE repair mix improves library complexity and increases mean target coverage

We found pre-treatment with the FFPE DNA Repair Mix improved library yields (Figure 3), which in turn led to an improvement in mean target coverage (Figure 4). Yield improvements varied from 1.2x in mildly compromised samples to 1.5x in severely damaged samples.
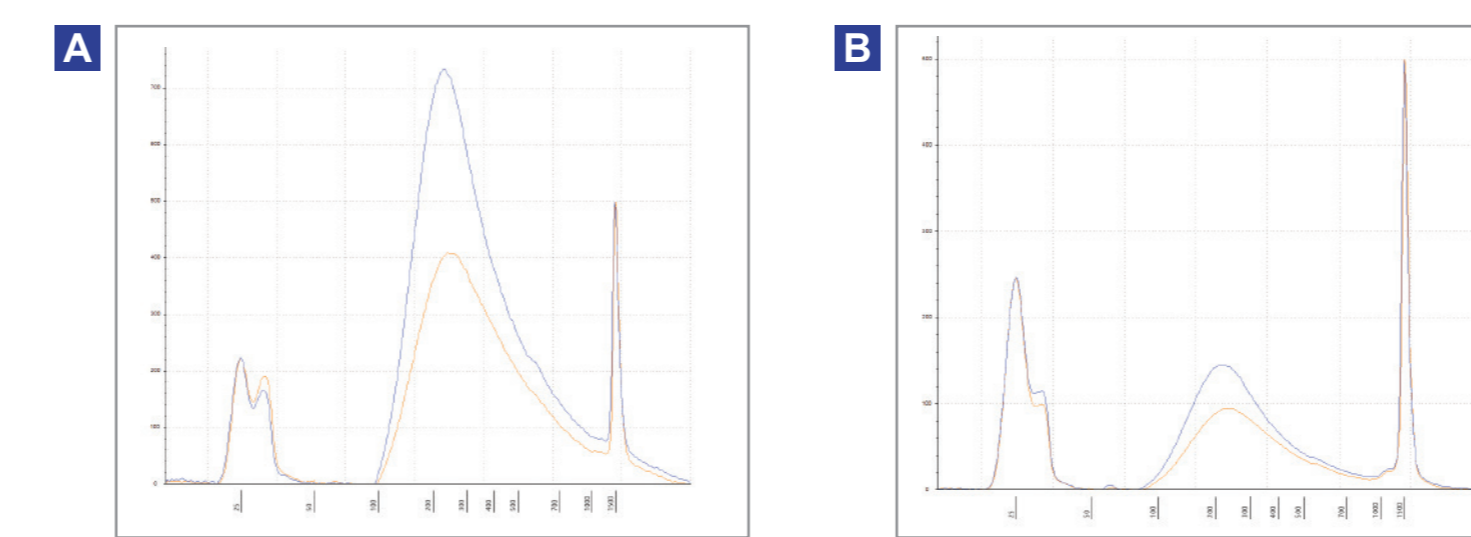

Figure 3: Agilent TapeStation traces of pre-capture libraries. Illumina-compatible libraries were prepared with severely damaged fcDNA that was treated with SureSeq FFPE DNA Repair Mix (blue), or was untreated (orange). The amount of input DNA were A 200 ng and B 50 ng.
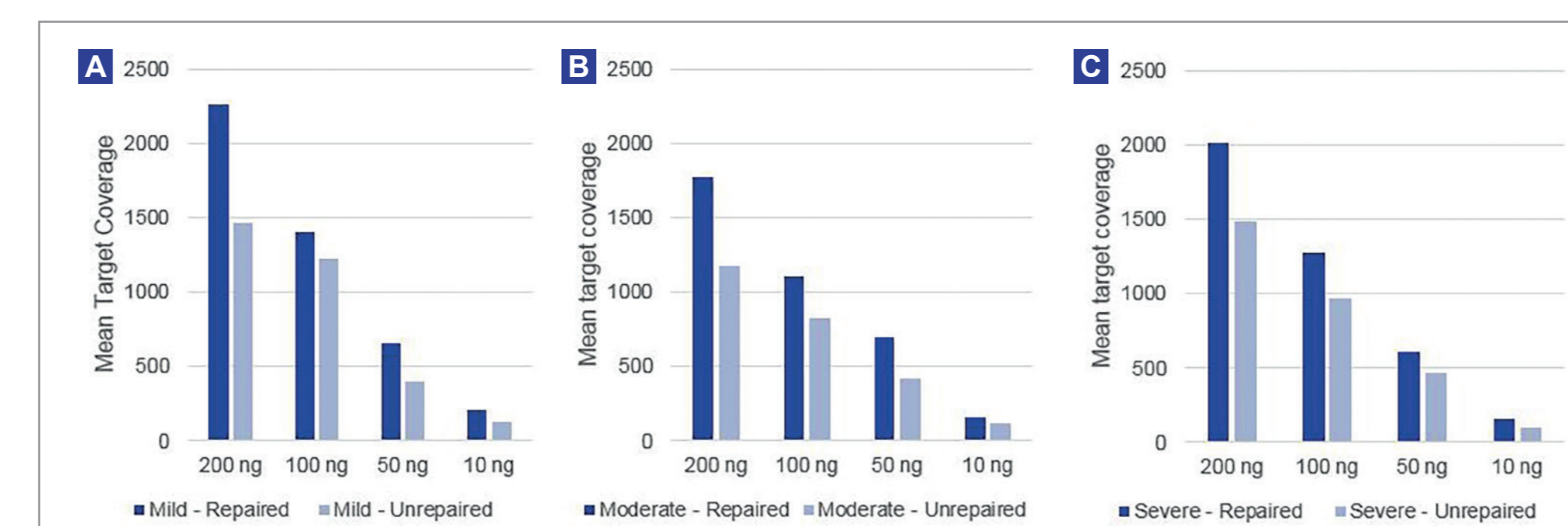

Figure 4: Improvement in mean target coverage can be observed across all amounts of starting material and all levels of DNA damage. A – mild fcDNA; B – moderate fcDNA; C – severe fcDNA.

The amount of archived sample tissue available can be limited and may only contain a low percentage of tumour cells of interest, therefore assays need to demonstrate strong sequencing performance at low input amounts. High depth and uniformity of coverage enables the accurate detection of low frequency mutations.

Samples treated with FFPE Repair Mix also demonstrated good sequencing metrics at low input amounts, maintaining a high and uniform level of coverage over a range of input amounts (Figure 5).
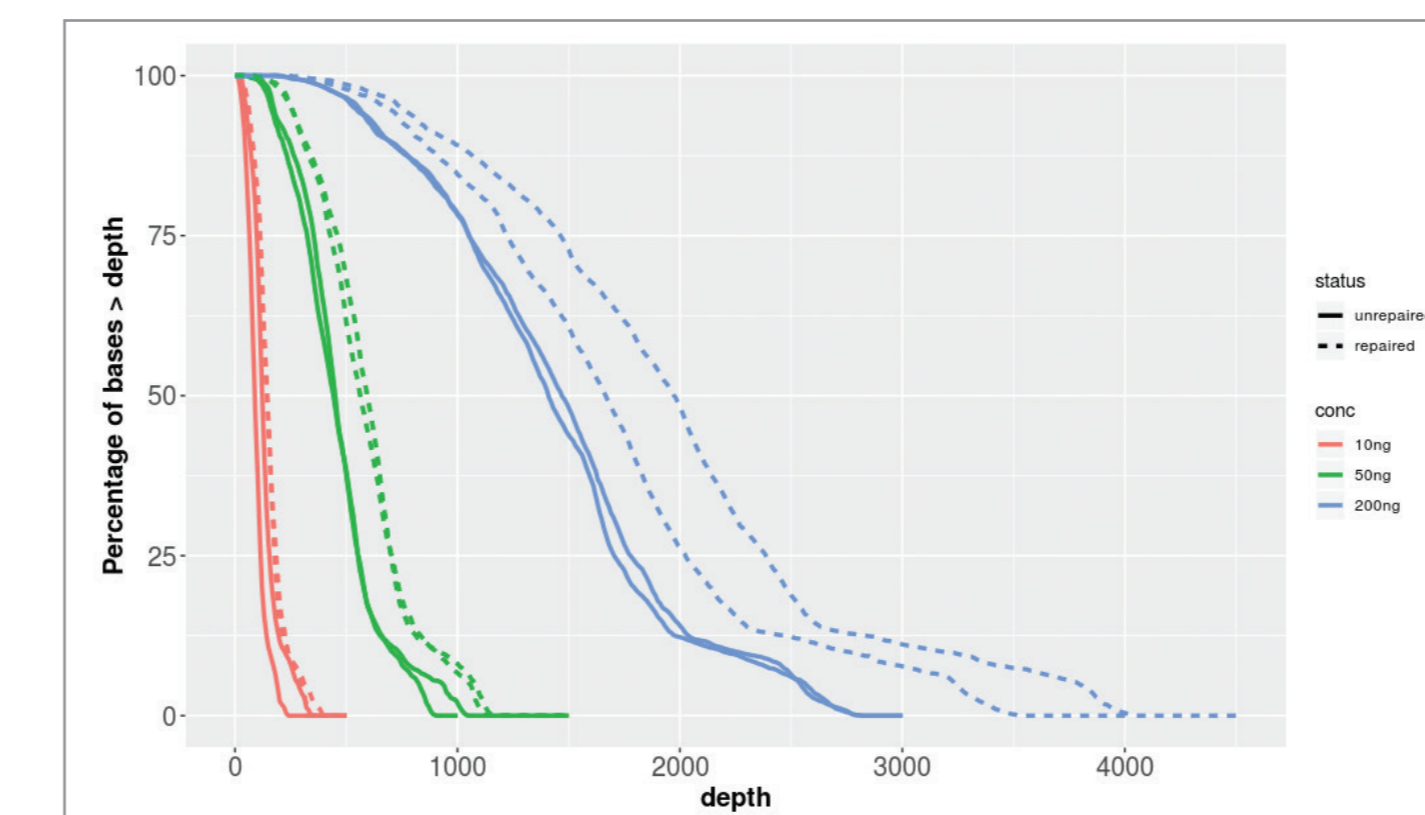

Figure 5: Plot of % bases at greater than x depth for treated (dotted lines) and untreated severely fcDNA (solid lines) samples using 10 (red), 50 (green), and 200 (blue) ng input.

The quantity and quality of sequencing data has a direct impact on the confident identification of variants, in particular, low frequency somatic variants. We found the number of supporting reads increased in repaired samples as shown in Figure 6 and reduced the number of false positive calls.
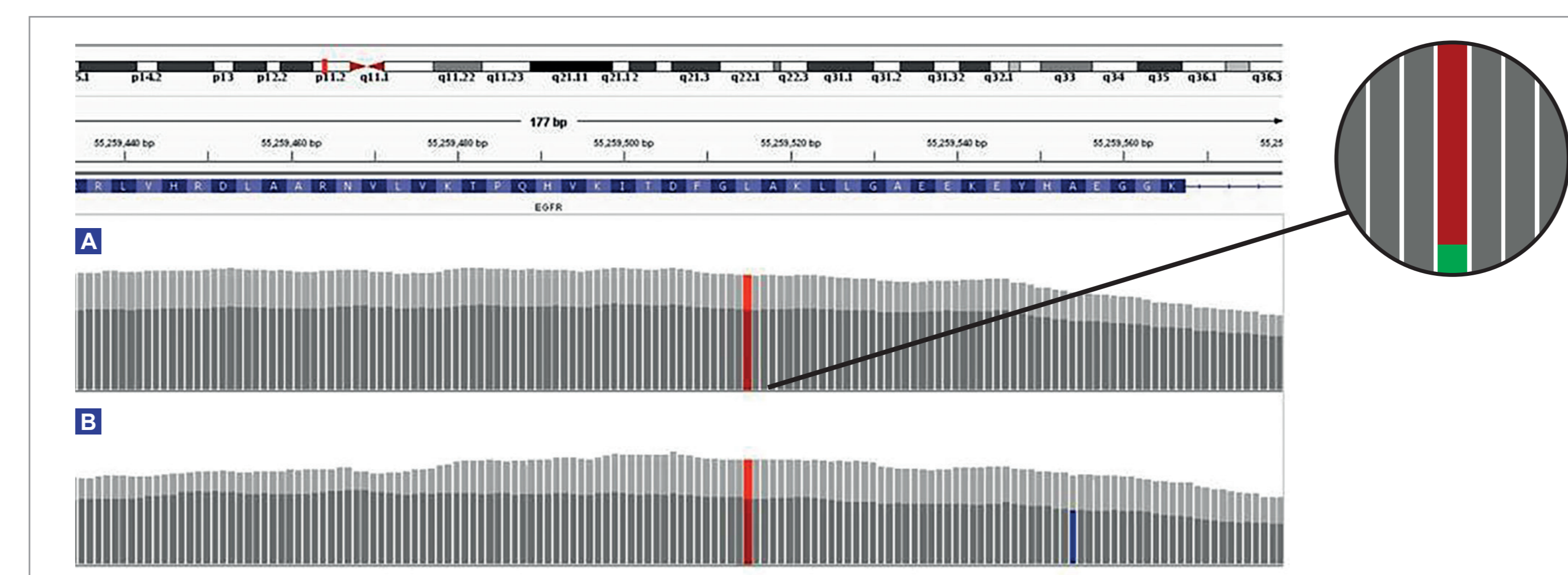

Figure 6: Comparison of the depth of coverage over a 3% EGFR L858R mutation in a severely formalin-compromised sample treated with FFPE repair mix (light grey) and not treated (dark grey). At 200 ng input A following treatment, the total depth increased by 43%, from 2707 to 3860, and the number of reads supporting the variant increased from 85 to 112, an increase of 32%. At 10 ng input B following treatment, the total depth increased by 58%, from 216 to 341, with the number of supporting reads increasing from 5 to 13 (160%). The combination of low input and poor quality DNA increases the likelihood of false positive mutations (blue).

All values are based on de-duplicated data. Visualised using Integrated Genomics Viewer[2].
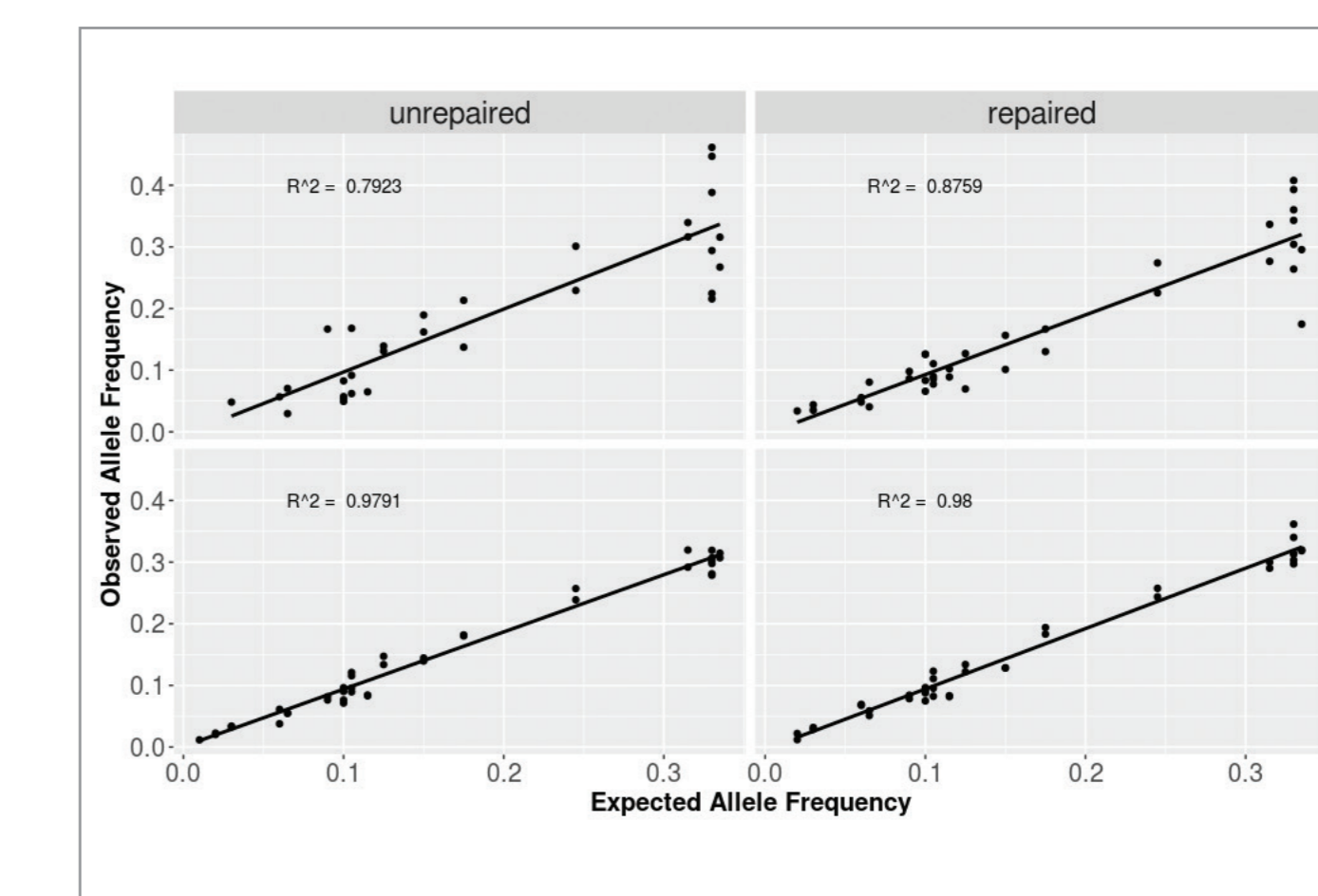
Accurate detection of variants from reference standards

The custom hot-spot panel was designed to capture 20 variants present in the reference standards: 15 single nucleotide variants (SNVs) and 5 deletions with variant allele frequencies varying between 1 and 33%. We found 100% concordance in repaired samples with >500x coverage, including three sub-5% EGFR variants. Overall 99.6% of the expected variants were detected with 91.25% of the 240 variants lying within 5 percentage points of the expected values (Table 1).

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%	0.0%	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%
P13KCA	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%
P13KCA	H1047R	17.5%	16.6%	16.1%	17.1%	7.9%	16.7%	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%
NOTCH 1	P668S	31.5%	28.5%	28.8%	25.8%	13.4%	25.6%	26.8%	26.8%	22.2%	28.3%	30.0%	27.0%	17.9%
ALK	F1543S	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 1: Difference between the expected and observed allele frequency in Horizon Discovery's fcDNA standards treated with FFPE Repair Mix (unrepaired data not shown). The variants were identified using OGT's SureSeq Interpret™ software. Mean of duplicates shown.
*variants confirmed by Horizon Discovery droplet digital PCR, the presence of the remaining variants were confirmed in the parental cell line.

Figure 7: Correlation between expected and observed allele frequency in severely compromised fcDNA unrepaired/repaired samples at 10 and 200 ng input. The R[2] values are improved in repaired samples, most noticeably at 10 ng, the lowest input tested.

We found the correlation between expected and observed values was improved in samples treated with FFPE Repair Mix (Figure 7). The greatest improvement in accuracy of MAF value was found in severely formalin-compromised samples, in particular when starting with 10 ng of DNA.

Conclusions

- We have demonstrated a superior depth of coverage across a range of starting DNA input amounts in formalin compromised DNA of varying quality treated with FFPE Repair Mix.
- 100% concordance in variant detection from reference standards with >500x coverage.
- Difference between expected and observed allele frequency is reduced in repaired low input samples, <50 ng, compared to un-repaired samples.
- Accurate detection of variants of less than 5% frequency, from as little as 10 ng of DNA.
- The SureSeq hybridisation-based approach is a robust and reproducible method for the identification of low frequency somatic variants from technically challenging samples.

References

1. C. Meldrum, M. A Doyle, and R. W. Tothill. Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. Clin Biochem Rev. 2011 Nov; 32(4): 177–195.
2. H. Thorvaldsdóttir, J. T. Robinson, and J. P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief bioinform 2013 Mar; 14(2):178-92