Ultra-sensitive Cancer Liquid Biopsy Analysis with the Agilent SureSelect$^{XT\ HS}$ Target Enrichment Workflow

Application Note

Authors

Louise J Barber[1], **Sonia Mansukhani**[1], **Bram Herman**[2], **Bahram Arezi**[2] **and Marco Gerlinger**[1,3]

[1]Translational Oncogenomics Laboratory, Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

[2]Agilent Technologies

[3]Gastrointestinal Cancer Unit, Royal Marsden Hospital, London, UK

Summary

Low input cfDNA sequencing and molecular barcode error correction with Agilent SureSelect$^{XT\ HS}$ technology unlocks mutation and DNA copy number profiles of liquid biopsies.

Introduction

Circulating free DNA (cfDNA) sequencing allows minimally invasive genetic profiling of cancers and the detection of subclonal genetic variants which are frequently missed by single biopsies in heterogeneous cancers. Tumour derived DNA content within the cfDNA is low in many cancer types (<5%), with subclonal variants reaching even lower variant frequencies (often <1%). Mutation discovery in cfDNA hence requires highly sensitive and specific techniques to identify rare variants with confidence.

Solution-phase hybridization capture sequencing allows simultaneous investigation of nucleotide level variation, copy number aberrations and translocations across target gene panels, but detection of rare variants has previously been limited by PCR and sequencing error rates. Application of Agilent SureSelect$^{XT\ HS}$ technology, which integrates molecular barcodes (MBC) for sequencing error correction, achieved highly sensitive (detection limit ~ 0.1%) and easily customizable cfDNA sequencing. The analysis of off-target sequencing reads enabled reconstruction of genome-wide copy number profiles in addition to mutation and copy number changes of target genes.[1]

Agilent Technologies

Materials and Methods

Plasma Preparation and cfDNA Isolation

Blood was collected in K3EDTA Vacuette tubes (Greiner Bio-one #455036) and centrifuged within 2 hours (10 minutes, 1600xg). The plasma layer was carefully removed and stored at -80°C. Upon thawing, samples were centrifuged at high speed (10 minutes, 16000xg, 4°C) and cfDNA was extracted from 4 ml plasma and 8 ml for healthy donors using the Qiagen QIAamp Circulating Nucleic Acid Kit, following the manufacturer's protocol. cfDNA samples were evaluated on an Agilent Bioanalyzer High Sensitivity DNA chip, or if high concentration, on a Agilent DNA 7500 chip (Figure 1A) and concentrations were confirmed by Qubit HS assay (ThermoFisher).

SureSelect Target Enrichment Bait Library Design

A 164 kb custom SureSelect capture bait library was designed to target 32 Colorectal Cancer (CRC) driver genes or genes implicated in therapy resistance. The library was designed in Agilent SureDesign (https://earray.chem.agilent.com/suredesign/) with 5x tiling density, most stringent masking, and max performance boosting.

Library Preparation and Sequencing

The recommended SureSelect$^{XT\ HS}$ protocol was followed with amendments as detailed in the optimized protocol outlined below. PCR was performed on an Eppendorf Mastercycler nexus GSX1. Cycle number optimization may be required for different PCR machines and capture bait library size. Final prepared libraries were checked on a Bioanalyzer High Sensitivity DNA chip and quantified by qPCR using the Kapa Library Quantification kit before pooling. Pools were clustered using an Illumina cBot and sequenced with paired end 75 reads on an Illumina Hiseq2500 in Rapid output mode.

Data Analysis:

After demultiplexing, Agilent SureCall 4.0 software was used to align reads to the hg19 human genome and call variants. Default parameters for the SNPPET SNP caller were adjusted as follows: variant call quality threshold = 70; minimum allele frequency = 0.001; minimum number of reads supporting variant allele = 2. Called variants were reviewed using the Integrative Genomics Viewer (IGV, Broad Institute). Resulting bam files were used to generate genome-wide DNA copy number profiles using CNVkit[2].

Results

Optimizing Agilent SureSelect$^{XT\ HS}$ for Low Input cfDNA

25 ng of cfDNA was used as the standard input for SureSelect$^{XT\ HS}$. This amount can be obtained for most patients with metastatic colorectal cancer from 4 ml of plasma and should contain >7500 haploid genome equivalents. After accounting for expected conversion efficiency, this is theoretically sufficient to detect mutations down to 0.1% variant allele frequency (VAF) with a requirement for at least two independent MBC consensus families per variant.

End Repair and A-tailing, Ligation, and Sample Cleanup steps were performed according to the manufacturer's protocol. It was ensured that both library preparation buffers were completely thawed (especially after dry ice shipment) and mixed prior to use.

8 cycles of pre-hybridization PCR were optimal for 25 ng of input cfDNA to generate sufficient product (500-1000 nanogram prepared library recommended for SureSelect$^{XT\ HS}$) for hybridization. In a separate pilot, we achieved good reproducibility of the pre-hybridization PCR yield: for 3 independent cfDNA samples that were library prepared in quadruplicate each with 25 ng of input, the coefficient of variance of the mean PCR yield for each cfDNA was 0.10, 0.13 and 0.29 (data not shown).

After clean-up and quality control with a Bioanalyzer DNA 7500 chip, the entire library (13.5 µl) was used as input for hybridization. While using the entire reaction volume removes the ability to store some prepared library as backup, it avoids needing an additional PCR cycle and any respective potential sampling bias linked to this. Variation in input quantification and sample quality resulted in some samples having higher yield than others. Hybridization with up to 2 µg of DNA input was tested and no adverse effect on hybridization efficiency was identified. Pre-hybridization PCR yields correlated with median on-target sequencing depth (Figure 1B).

To allow for the extra library volume of prepared library used in the hybridization, only 1.5 µl of nuclease-free water was added to the subsequent capture library hybridization mix per reaction. Two µl of our CRC custom bait capture library were added to the capture library hybridization mix, as recommended for libraries <3 Mb.

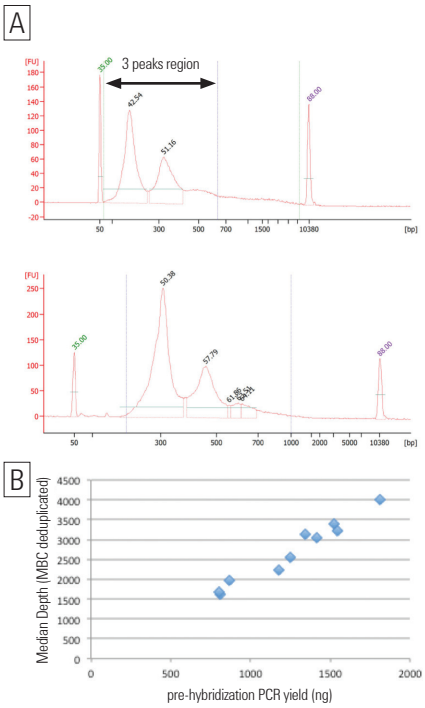

Figure 1. (A) Bioanalyzer DNA 7500 chip traces of a high yield cfDNA sample following extraction from plasma (top) and then on a Bioanalyzer High Sensitivity DNA chip after pre-hybridization PCR (bottom). The marked region in the top panel indicates the 3 peaks of cfDNA used in the quantification of the sample. (B) A direct correlation is seen between the yield of DNA obtained in the pre-hybridization PCR and the median on-target depth achieved after deduplication using MBCs. Data is shown for ten independent human plasma cfDNA samples prepared using our optimised protocol.

All other reagents were added according to the standard protocol, and 60 cycles of Fast Hybridization were performed.

Capture was started immediately after the final Hybridization cycle, and proceeded for 30 minutes at room temperature.

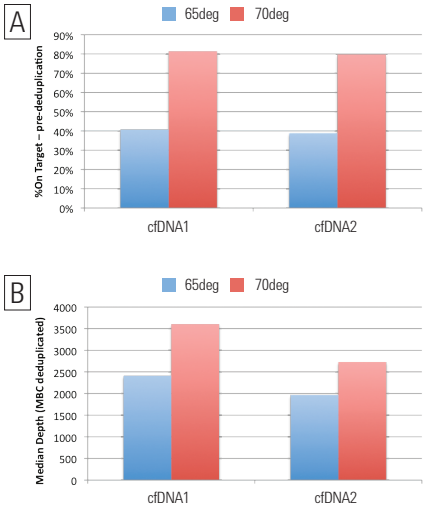

Figure 2. A comparison of two human plasma cfDNA samples prepared using either the 65°C or 70°C wash protocols and sequenced with an average 95M reads. Washing at 70°C increases %on-target (A) and median on-target depth (B).

	65 °C	70 °C
Raw reads	92M	100M
% On-target before deduplication	34.0%	71.4%
Duplicates	48%	83%
Off target reads after deduplication	37M	8M
Median depth on-target before deduplication	14089	31604
Median depth on-target after deduplication	1921	2626
Median MBC amplification	4	6

Table 1. A comparison of sequencing metrics for a cfDNA sample prepared using either the 65°C or 70°C wash protocol. The data has not been normalized to an equal number of sequenced reads per sample to avoid bias. However, if the 70°C wash protocol data is down-sampled to match the 65°C wash protocol (92M reads) then the median depth observed (2588) is still higher than that achieved at 65°C (1858).

Importantly, it was found that changing the temperature of the post-capture wash steps alters the ratio of on-target reads, needed for accurate variant calling with high sensitivity, and off-target reads, which can be used to generate genome wide copy number profiles with CNVkit software. CNVkit infers genome wide copy number profiles by using off target reads generated through targeted DNA sequencing. Resulting data is corrected for GC content and normalized against a pooled dataset obtained from eight healthy donor cfDNA samples sequenced with the same panel.

Following the legacy protocol, washes carried out at 65°C (wash buffer 2) resulted in over 60% of reads aligned to off-target regions of the genome (Figure 2A). For example, samples sequenced with 92 million reads contained more than 35 million off target reads (Table 1). However, using the CNVkit algorithm, these off-target reads could be utilized to reconstruct genome-wide copy number profiles even when tumour content is low (Figure 3A). This approach enables maximum use of the generated sequencing data.

A high off–target fraction reduces the on-target sequencing depth and hence requires a compensatory increase in the total number of sequenced reads to achieve a given depth. Thus, it was investigated whether the on-target rate could be increased by performing more stringent washes with a temperature of 70°C. This significantly reduced the off-target fraction to 20% (Table 1 and Figure 2A) and increased depth (Figure 2B). However, in most

samples, the number of off-target reads was still sufficient for copy number reconstruction (Figure 3B).

Post-Capture Amplification was performed with 10 cycles of PCR. This was found to give sufficient final library for sequencing. Two rounds of 1x AMPure XP bead cleanup were performed to sufficiently remove unincorporated primers.

Sequencing Performance with Horizon Reference cfDNA

Horizon Multiplex I cfDNA Reference Standards (Horizon HD780) are commercially available cell line-derived DNA samples that include six common single-nucleotide mutations in cancer driver genes at pre-defined variant frequencies. This reference DNA is fragmented to an average fragment length of 160 bp, resembling the predominant DNA fragment size found in cfDNA.

25 ng of the 1% Horizon cfDNA Reference Standard was prepared using the 65°C wash 2 protocol and sequenced to a median depth of >7000x (>50M reads) before deduplication (Table 2A). All of the six mutations were detected with a sensitivity and specificity of 100% (Table 2B).

The SureSelect$^{XT\ HS}$ MBC technology reduced the false positive detection rate for such low-level variants. After standard deduplication (based on alignment coordinates of paired reads and insert length), 218 mutations with a VAF below 2% were called in the Horizon sample compared to the reference genome (Figure 4).
In contrast, when the data was deduplicated using the MBCs, only 18 variants were detected at VAF <2%. This is a 92% reduction of calls with low frequency through the use of MBC error correction while retaining the six known variants (Figure 4).

Sequencing Performance with cfDNA

cfDNA samples from two healthy donors were used in a mixing experiment to determine the limits of detection of this liquid biopsy assay. cfDNA from one donor was mixed into the cfDNA from the second donor at defined ratios (0.15%, 0.075%, and 0.0375%). 25 ng of DNA from each of these mixed samples were used for library preparation with 65°C wash 2.

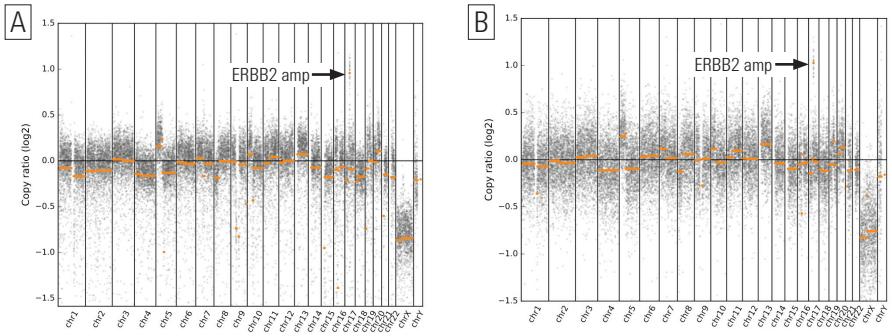

Figure 3. A human plasma cfDNA sample with estimated 10% tumour content was prepared using either the 65°C (A) or 70°C (B) wash protocols and sequenced with 113M (65°C) and 95M (70°C) reads. MBC deduplicated bam files were analysed using CNVkit to generate a genome-wide copy number profile using the off-target reads (49M and 5M for 65°C and 70°C, respectively). Equivalent copy number profiles were generated at both wash temperatures, despite the low tumour content of the cfDNA sample (truncal APC and TP53 variants identified in archival FFPE biopsies were detected at 8 and 6% respectively). However, the higher number of off-target reads obtained at 65°C provides greater resolution, and a tighter profile. An ERBB2 amplification identified in the archival FFPE biopsies (chromosome 17) is detected at both temperatures. Whilst ERBB2 is within the targeted region and therefore amplification can be inferred from the raised read depth (4600 vs median depth 1700), CNVkit analysis shows that the amplification is focal, rather than part of a larger chromosomal region copy number change.

A

Raw reads	58M
% On-target before deduplication	28.1%
Duplicates	40%
Median depth on-target before deduplication	7404
Median depth on-target after deduplication	947
Median MBC amplification	3

B

EGFR L858R	0.83%
EGFR T790M	1.09%
KRAS G12D	0.94%
NRAS Q61K	1.69%
NRAS A59T	1.68%
PIK3CA E545K	0.84%
Sensitivity	100%
Specificity	100%

Table 2. Sequencing metrics (A) and detected VAF (B) for the six known single nucleotide variants for 25 ng of the 1% Horizon cfDNA Reference Standard. Specificity was calculated based on the detection of other bases at the indicated position.

Libraries were sequenced to a median depth of >20000x (>70M reads) before deduplication (Table 3A) with the aim to re-identify 16 homozygous SNPs, which were known to be unique to the sample spiked in with low frequency. At 0.15% spike-in, all 16 SNPs were detected, indicating a sensitivity of 100% (Table 3B). Fifteen of 16 SNPs (93.75%) were detected at 0.075%, and eleven (68.75%) could still be detected at 0.0375% (Table 3B). The mean VAF was close to the spike-in ratio (0.15%: mean 0.188, CV 0.6; 0.075%: mean 0.099, CV 0.7). The read depth at the missed SNP (0.075% spike-in) was much lower than those successfully detected (1253 vs mean depth across SNP regions of 3975). It is likely that where more than 25 ng of cfDNA is available, increasing the input and therefore the number of genome equivalents incorporated into the resulting library would increase the sensitivity of this sequencing technology further.

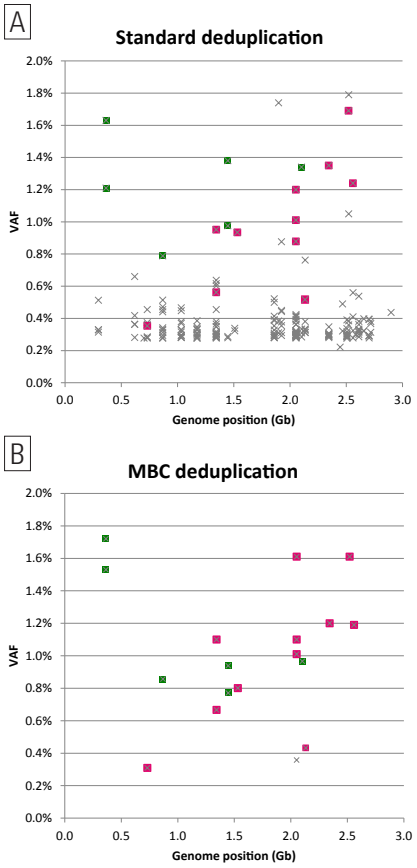

Figure 4. Horizon 1% standard was prepared using the 65°C wash protocol and sequenced with 58M reads. Data was analysed in SureCall using standard (A) or MBC (B) deduplication, and low level variants (<2% VAF) are shown. The six known variants are highlighted in green, with unknown variants common to both MBC and standard deduplication analysis highlighted in pink. Other base changes detected in the targeted region relative to the hg19 reference genome are marked as grey crosses. Only the known variants (green) were observed by Horizon in their independent Exome dataset. Orthogonal validation of the other variants was not performed.

	0.15%	0.075%	0.0375%
Raw reads	74M	84M	71M
% On-target before deduplication	70.0%	71.3%	71.9%
Duplicates	78%	81%	83%
Median depth on-target before deduplication	22130	25361	22030
Median depth on-target after deduplication	3077	3034	2340
Median MBC amplification	4	5	6

Table 3A. Sequencing metrics for the healthy donor cfDNA mixing experiment.

	0.15%		0.075%		0.0375%	
	VAF	depth	VAF	depth	VAF	depth
5:112162854	0.09 %	4333	0.05 %	4374	0.030 %	3290
5:112164561	0.11 %	1826	0.19 %	1574	ND	1425
5:112175770	0.17 %	4791	0.08 %	4784	0.231 %	3465
5:112176325	0.07 %	4572	0.07 %	4490	ND	3489
5:112176559	0.30 %	3996	0.13 %	3926	0.033 %	3056
5:112177171	0.44 %	3666	0.06 %	3412	0.073 %	2730
6:111695268	0.26 %	3905	0.29 %	3756	0.031 %	2929
6:111696852	0.37 %	2961	0.20 %	2940	0.128 %	2351
7:116339282	0.13 %	4514	0.02 %	4561	0.029 %	3467
7:116435768	0.14 %	5054	0.15 %	5175	0.075 %	3988
7:116436022	0.29 %	4829	0.11 %	4734	0.027 %	3680
7:116436097	0.17 %	4771	0.09 %	4456	ND	3529
7:55268916	0.04 %	4825	0.04 %	4845	ND	3512
18:44346224	0.13 %	2999	0.03 %	2948	0.07 %	2516
18:51170937	0.23 %	1300	ND	1253	ND	1120
18:61164602	0.08 %	5258	0.07 %	5130	0.10 %	3673
Sensitivity	100 %		94 %		68%	
Specificity	99.98 %		99.99 %		99.99 %	

Table 3B. Detected VAF and depth (number of consensus read families) for the sixteen homozygous SNPs specific to the spiked-in sample. All expected SNP positions were reviewed manually using IGV and were called if the variant was present in more than one MBC consensus read family, or in one MBC read family consisting of at least 5 supporting reads. Specificity was calculated based on the detection of other bases at the indicated position.

detected	
not detected	

Conclusions

We have optimized Agilent SureSelect$^{XT\ HS}$ Target Enrichment Workflow for the ultra-deep and error corrected sequencing of small cfDNA quantities. The MBC technology reduced calls with low VAF by 92% compared to standard deduplication, which has a major impact on false positive calls, while retaining the known variants. At a median sequencing depth of 20000x before deduplication, we could reliably detect all of 16 variants in plasma-derived cfDNA at 0.15% VAF (average depth after MBC deduplication across these regions: 3975, CV 0.29) and 93.75% at 0.075% VAF (average depth after MBC deduplication across these regions: 3897, CV 0.31), suggesting that the sensitivity of this technology for 25 ng input cfDNA is close to 0.1%. We also demonstrate that reconstruction of genome-wide copy number profiles from off-target reads using CNVkit can be reliably combined with the highly sensitive detection of low frequency variants with small (164 kb) SureSelect$^{XT\ HS}$ target panels. We were able to detect focal amplifications in cfDNA samples with tumour content as low as 10%.

References

1. Mansukhani S, Barber LJ, *et al.* (2017) BioRxiv 213306. Doi: https://doi.org/10.1101/213306

2. Talevich, E., Shain, A. H., Botton, T., & Bastian, B. C. (2014) PLOS Computational Biology 12(4): e1004873.

Trusted Answers. Together.

PR7000-0733
© Agilent Technologies, Inc., 2017
Published in the USA, September 20, 2017
5991-8464EN

LEARN MORE AT:
www.agilent.com/genomics

Find an Agilent customer center in your country:
www.agilent.com/genomics/contactus

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Europe
info_agilent@agilent.com

Asia Pacific
info_agilent@agilent.com

For Research Use Only. Not for use in diagnostic procedures.

Agilent Technologies