

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Fettke, H;Steen, JA;Kwan, EM;Bukczynska, P;Keerthikumar, S;Goode, D;Docanto, M;Ng, N;Martelotto, L;Hauser, C;Southey, MC;Azad, AA;Nguyen-Dumont, T

Title:

Analytical validation of an error-corrected ultra-sensitive ctDNA next-generation sequencing assay

Date:

2020-01-01

Citation:

Fettke, H., Steen, J. A., Kwan, E. M., Bukczynska, P., Keerthikumar, S., Goode, D., Docanto, M., Ng, N., Martelotto, L., Hauser, C., Southey, M. C., Azad, A. A. & Nguyen-Dumont, T. (2020). Analytical validation of an error-corrected ultra-sensitive ctDNA next-generation sequencing assay. *Biotechniques*, 69 (2), pp.133-140. <https://doi.org/10.2144/BTN-2020-0045>.

Persistent Link:

<https://hdl.handle.net/11343/252478>

License:

CC BY-NC-ND

Analytical validation of an error-corrected ultra-sensitive ctDNA next-generation sequencing assay

Heidi Fettke¹ , Jason A Steen², Edmond M Kwan^{1,3} , Patricia Bukczynska⁴, Shivakumar Keerthikumar^{5,6}, David Goode^{5,6}, Maria Docanto¹, Nicole Ng⁷, Luciano Martelotto⁴, Christine Hauser⁴, Melissa C Southey^{2,8,9} , Arun A Azad^{‡,1,6,10}  & Tu Nguyen-Dumont^{*,‡,2,8} 

¹Department of Medicine, School of Clinical Sciences, Monash University, Melbourne, Australia; ²Precision Medicine, School of Clinical Sciences at Monash Health, Melbourne, Australia; ³Department of Medical Oncology, Monash Health, Melbourne, Australia; ⁴Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia; ⁵Computational Cancer Biology Program, Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia; ⁶Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Australia; ⁷Walter & Eliza Hall Institute of Medical Research, Melbourne, Australia; ⁸Department of Clinical Pathology, The University of Melbourne, Melbourne, Australia; ⁹Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia; ¹⁰Department of Medical Oncology, Peter MacCallum Cancer Centre, Melbourne, Australia; *Author for correspondence: tu.nguyen-dumont@monash.edu; ‡ Authors contributed equally

BioTechniques 69: 00–00 (August 2020) 10.2144/btn-2020-0045

First draft submitted: 10 April 2020; Accepted for publication: 27 May 2020; Published online: 13 July 2020

ABSTRACT

Plasma circulating tumor DNA (ctDNA) analysis has emerged as a minimally invasive means to perform molecular tumor typing. Here we developed a custom ultra-sensitive ctDNA next-generation sequencing assay using molecular barcoding technology and off-the-shelf reagents combined with bioinformatics tools for enhanced ctDNA analysis. Assay performance was assessed via a spike-in experiment and the technique was applied to analyze 41 plasma samples from men with advanced prostate cancer. Orthogonal validation was performed using a commercial assay. Sensitivity and specificity of 93 and 99.5% were recorded for ultra-rare somatic variants (<1%), with high concordance observed between the in-house and commercial assays. The optimized protocol dramatically improved the efficiency of the assay and enabled the detection of low-frequency somatic variants from plasma cell-free DNA (cfDNA).

METHOD SUMMARY

High-throughput multigene assays interrogating circulating tumor DNA (ctDNA) have the potential to significantly improve the field of precision oncology. However, existing methods are expensive and lack the sensitivity required to identify ultra-rare somatic variants. Here we present a next-generation sequencing assay that uses molecular barcoding to achieve a high sensitivity and specificity for ctDNA interrogation, while mitigating the high costs associated with a high proportion of unusable sequencing reads.

KEYWORDS:

cell-free DNA • cfDNA • circulating tumor DNA • ctDNA • liquid biopsy • molecular barcoding • next-generation sequencing

Advances in high-throughput technological platforms such as next-generation sequencing (NGS) have allowed comprehensive tumor profiling, fueling the rapidly expanding field of precision oncology.

Although solid tissue biopsy is the current gold standard for molecular tumor typing, it has several shortcomings, including invasiveness, risk of complications, and difficulties obtaining sufficient material [1]. Additionally, spatial and temporal tumor heterogeneity prevents single site biopsies from fully capturing the tumorigenic landscape and thus from identifying clinically relevant mutations [2,3]. These challenges are particularly apparent in advanced prostate cancer, which commonly metastasizes to bone and retroperitoneal lymph nodes. Both locations are challenging sites to biopsy, frequently yielding minimal if any tumor content for analysis [4,5]. Therefore minimally invasive approaches to profile the tumor genome are much more attractive in advanced prostate cancer.

The use of ctDNA – fragmented DNA shed into the circulation from tumor cells – is increasingly recognized as a minimally invasive approach for interrogating the tumor landscape. However, ctDNA can be highly diluted with normal cell-free DNA (cfDNA, released into the blood primarily by hematopoietic cells) [6], and thus robust liquid biopsy assays with high analytical sensitivity are required for ctDNA to reach its full clinical potential.

Low-throughput candidate gene analysis approaches such as Droplet Digital PCR have shown high clinical utility in tumors that share common genomic drivers, such as non-small-cell lung cancer [7,8]. However, single gene analyses are not optimal in tumors with high inpatient heterogeneity, or in the setting of biomarker discovery. In contrast, massively parallel NGS methodologies simplify multigene testing, despite historically having lower sensitivities due to sequencing ‘noise’ and high error rates [1]. Recently, the use of hybrid capture-based targeted molecular barcoding assays to generate error-corrected reads (Figure 1) has led to the ability to confidently detect variants

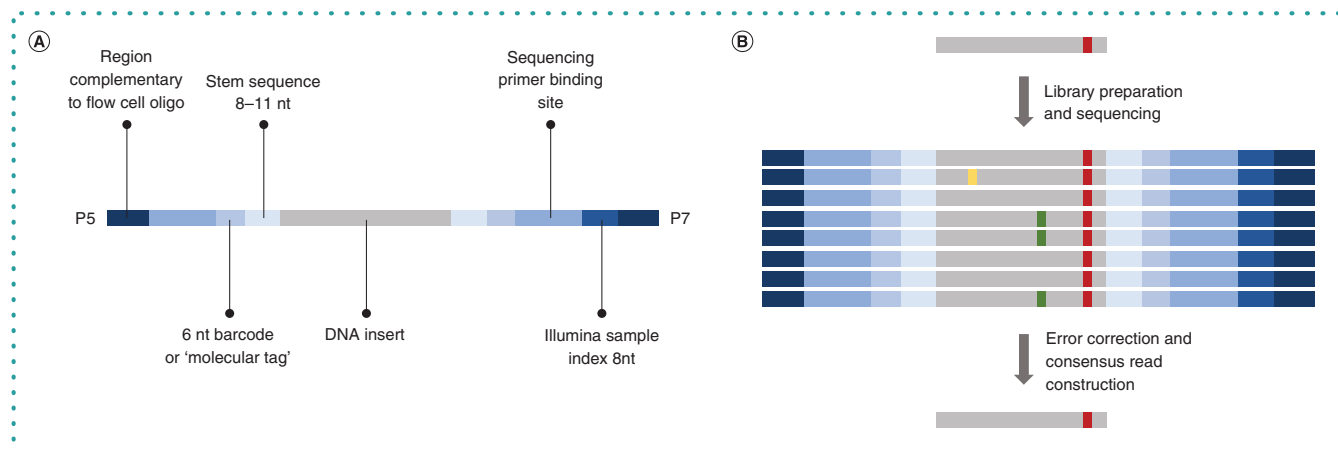


Figure 1. Structure of the Illumina-compatible molecularly tagged cfDNA libraries generated using the ThruPLEX Tag-Seq protocol incorporated into this assay. (A) The cfDNA insert is flanked by a stem sequence of 8–11 nucleotides in order to increase library complexity, followed by a unique barcode sequence. **(B)** The process of error correction and variant identification used in molecularly tagged libraries. Sequencing reads are grouped into 'families' – cfDNA fragment PCR duplicates that share the same unique barcode. Minority variants are attributable to PCR and/or sequencing errors and are removed, increasing the validity of identified 'majority' variants. cfDNA: Cell-free DNA; nt: Nucleotide.

at variant allelic fractions (VAFs) below 2% [1]. Unfortunately, high numbers of unusable off-target reads (up to 65%) that do not map to regions of interest significantly reduces the economic viability of such assays in a clinical setting [9].

To address these limitations, we developed a protocol for the generation of molecularly barcoded libraries with targeted hybrid capture, using off-the-shelf reagents and bioinformatics tools to generate an end-to-end streamlined assay. We adapted three key strategies to improve assay efficiency: first, adding a cocktail of four complementary, nonextendable custom oligonucleotides to mask degenerate library sequences and mitigate an off-target effect known as 'daisy-chaining' that occurs when independent complementary sequences bind to each other [10]; second, replacing primers used for post-capture enrichment in order to harmonize off-the-shelf kits; and third, performing two rounds of targeted gene capture (double-captured sequencing, dCAP-Seq) with extra washes to further improve on-target rate. Here we describe the characteristics and performance of our in-house assay, including a comparison with a commercial assay using plasma samples from a cohort of advanced prostate cancer patients.

Patients, materials & methods

Study participants

Peripheral blood samples from a total of 44 men with metastatic castrate-resistant prostate cancer (mCRPC) commencing systemic therapy were collected between September 2016 and August 2018 at Monash Health (Melbourne, Australia). Healthy donor samples were collected from volunteers who self-reported as healthy, without a history of cancer, at Monash Health. Participants provided written informed consent, with ethics approval obtained from the Monash Health Human Research Ethics Committee.

Sample collection, processing & DNA extraction

Peripheral blood (up to 30 ml for mCRPC cases, up to 300 ml for healthy controls) was collected in EDTA-containing or dedicated cfDNA-stabilizing tubes (Streck, NE, USA) depending on estimated time to processing [1]. Tubes were centrifuged at $1900 \times g$ at 4°C for 10 min. The resulting separated plasma and buffy coat (containing peripheral blood mononuclear cells; PBMCs) layers were isolated, with plasma being further purified via a two-step centrifugation method (at $16,000 \times g$ for 10 min at 4°C , with any resulting pellets being discarded between each spin). Plasma and PBMCs were stored at -80°C until batch processing.

cfDNA was extracted using the QIAamp circulating nucleic acid kit (Qiagen, Hilden, Germany) and eluted in 50 μl buffer AVE. Because cfDNA fragments are between 150 and 500 bp in length [1], contaminating large molecular weight fragments (> 500 bp) were removed using a 0.5:1 ratio of AMPure XP paramagnetic beads (Beckman Coulter, CA, USA) [11]. The size distribution, quality and quantity of the resulting cfDNA were assessed using automated gel electrophoresis (Agilent TapeStation 2200 with High Sensitivity D1000 ScreenTapes, and Expert software; Agilent, CA, USA) and the Qubit double-stranded DNA high-sensitivity assay (Thermo Fisher Scientific, MA, USA).

Germline DNA was extracted from PBMCs using the QIAamp DNA blood mini kit (Qiagen) and eluted in 200 μl buffer AE. DNA was then sheared to 150–200 bp in length using ultrasonication (Covaris, NC, USA) in order to match cfDNA fragment length, and quantification and quality control procedures were performed as for the extracted cfDNA.

Panel design

Genes were selected based on publicly available mutational data across 1047 advanced prostate cancer samples available at cBioPortal [12]. Forty-two commonly mutated genes of clinical relevance for prostate cancer (Supplementary Table 1) were selected to increase the likelihood of detecting an aberration while also limiting the genomic size of the panel.

The 42-gene panel was designed using the SureDesign software (Agilent). A pool of 5760 120-mer biotinylated RNA oligonucleotide 'baits' was designed to target the coding regions of the 42 genes, along with 1042 single nucleotide polymorphism (SNP) loci. The SNP targets were evenly distributed across the genome to improve whole-chromosomal coverage for large duplication/deletion event identification, as well as serving as a patient-specific DNA 'fingerprint' to allow confirmation of individuals across longitudinal sampling. Exonic regions had 2× bait tiling, while SNP regions had 1×, resulting in a total library size of 447.031 kbp.

Library construction

A total of 25–50 ng of DNA was used to generate indexed and barcoded (Figure 1) cfDNA and paired genomic DNA libraries using the ThruPLEX Tag-Seq kit (Takara Bio, CA, USA). Library fragment sizes were confirmed and double-stranded DNA quantified using the TapeStation 2200 and Qubit, as previously described. Samples yielding ≥400 ng proceeded to hybridization capture.

Hybridization capture, sequencing & data processing

The SureSelect^{XT} target enrichment system (Agilent) was used on DNA libraries between 400 and 750 ng. Protocol modifications were introduced to improve assay efficiency and on-target rate (see 'Assay optimization' in Results section). Final capture libraries were assessed via quantitative PCR (KAPA SYBR FAST qPCR kit; Roche, Basel, Switzerland), and pooled in equimolar amounts. Up to four cfDNA libraries were loaded onto an Illumina HiSeq 2500 2×150-bp flow cell and sequenced using the Rapid Run mode by a service provider.

Raw read pairs were demultiplexed using per-sample indices and mapped to the human reference genome hg19 using BWA-MEM (v0.7.12) and converted to bam files for further analysis (Samtools v1.6). Mapped reads were then demultiplexed based on the sequence of molecular barcodes on both ends of the read using Connor (v0.5; <https://github.com/umich-brcf-bioinf/Connor>). This software identifies PCR duplicates by grouping reads that share the same molecular barcode into 'families' (Figure 1). Only reads with perfectly matching barcodes and at least three replicates were clustered together as a family to increase the validity of each base call. Error correction was performed on a per-family basis to generate a single collapsed read, with each base called if a consensus frequency of >0.6 was reached.

Variant identification

Somatic variant identification was performed using VarDict (v1.5.8) in paired sample mode using standard filter settings, with the exception of minimum variant frequency (reduced to 0.001%) [13]. Somatic and germline variants (small indels and single nucleotide variants) were filtered and identified from paired VCF (variant call format) files annotated using VarSeq (v2, Golden Helix, Inc., MT, USA; www.goldenhelix.com) [14]. Only rare intragenic loss-of-function or missense variants with a CADD PHRED [15] score ≥20 variants termed 'strong somatic' by VarDict (i.e., only found in the cfDNA sample) with a minimum read depth of 500× and at least three supporting consensus reads were considered for downstream manual curation. cfDNA variants that were also present in the matched PBMC sample at VAF >5% were assumed to be derived from clonal hematopoiesis and removed. Germline variants were required to have at least 8× consensus read coverage. Manual curation was performed by visualizing VCF files using Integrative Genomics Viewer (Broad Institute, MA, USA) in order to remove any remaining sequencing artifacts (suspected polymorphic variants present across multiple independent patient samples, variants in homopolymeric regions or in 'messy' regions with multiple low-quality calls).

Copy number analysis

Copy number variants (CNVs) were identified using VarSeq CNV caller (Golden Helix). Coverage log ratios were calculated against a median reference generated using germline DNA samples that underwent the same library preparation and sequencing process. Only references with an overall divergence from the cfDNA sample of <30% were used for CNV calling. Copy number gains were required to have a p-value of <0.0001, a z-score of >0.25, and a minimum target depth of 50×. Deletions were required to have either a coverage ratio of <0.87 and a p-value of <0.00001, or have a coverage ratio of <0.7, a p-value of <0.05, and a z-score of <-0.85. Further details are presented at <http://protocols.io/view/an-error-corrected-panel-based-next-generation-seq-bb3siqne>.

Spike-in experiment

Plasma samples were obtained from two healthy donors (A and B). Individual B's cfDNA was spiked into individual A's cfDNA in known ratios (1, 0.5, 0.25 and 0.1%). Targeted sequencing data were generated and processed as described above. VarSeq was then utilized to identify loci where the two samples were of opposite genotypes, that is, sample A was homozygous to the reference and sample B was homozygous alternate. The result was 133 variants across 21 chromosomes that were expected to be heterozygous in the spiked-in samples (true positive variant calls). In addition, true negative calls were defined as loci in which both donors were homozygous alternate and where spiked-in samples were thus expected to be homozygous alternate (142 loci across 22 chromosomes).

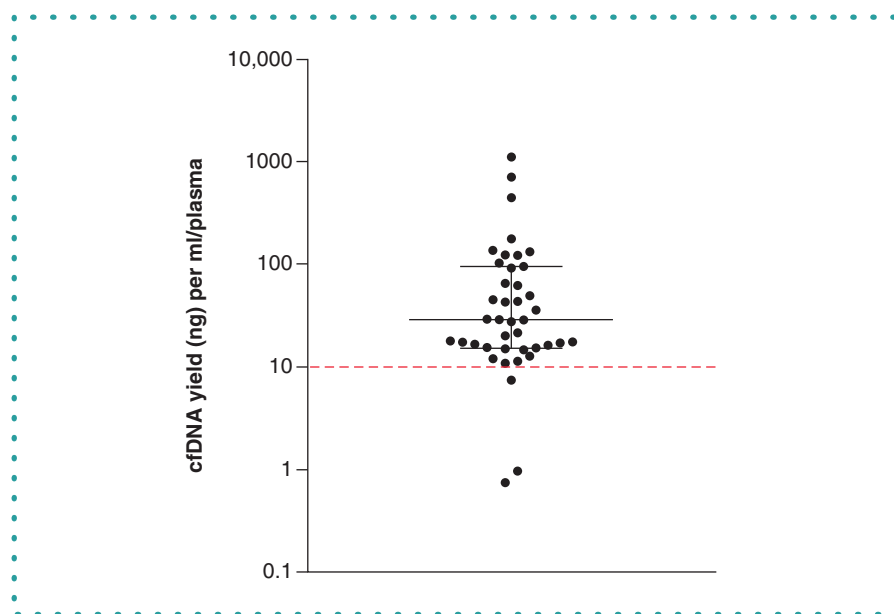


Figure 2. Total cfDNA yields in nanograms per milliliter of plasma obtained from 41 men with metastatic castrate-resistant prostate cancer. DNA was extracted from 5 ml plasma. Median and interquartile range indicated (nonparametric data). Samples above the dashed line yielded a minimum of 50 ng cfDNA for library preparation. cfDNA: Cell-free DNA.

Sequencing by an independent commercial assay

Whole blood from 41 patients was processed and sequenced in our facility as above, with a minimum of 5 ml of frozen plasma and matched buffy coat additionally being shipped for testing using an independent panel-based NGS assay that also uses molecular bar-coding and hybridization capture to assess 90 cancer-related genes. Eighteen of the 42 genes included on our panel were also present in the commercial assay gene panel. A variant report was returned, with raw sequencing data not available for analysis.

Results & discussion

Determination of DNA input

Optimal DNA input for library preparation was determined based on the feasibility of obtaining such an amount from the majority of patients. Total cfDNA was extracted and quantified from the plasma of 41 men with mCRPC. A median of 29 ng were extracted per milliliter of plasma, with >92% (38/41) samples yielding at least 50 ng (the library input maximum) from 5 ml of plasma (Figure 2). Because cfDNA input into NGS assays is inversely correlated with the limit of detection for variant alleles [16], the ThruPLEX Tag-Seq and SureSelect^{XT} protocols (including amplification cycles) were optimized for 50 ng input.

Assay optimization

To determine initial assay performance and on-target rate, three patient cfDNA libraries were prepared using the ThruPLEX Tag-Seq kit. Our custom probe set was then used in tandem with the Agilent SureSelect^{XT} target enrichment system, which performs targeted hybridization capture, with two key modifications. First, SureSelect Block 3 was replaced with a cocktail of four complementary, nonextendable custom oligonucleotides (Integrated DNA Technologies, IA, USA) to mitigate the ‘daisy-chaining’ of degenerate library sequences that enables off-target capture [10]. Secondly, during post-capture PCR enrichment (16 cycles), the SureSelect ILM forward primer was replaced with 0.5 μ M Illumina P5 and P7 primers (Integrated DNA Technologies) in order to allow tandem use of the two kits. Primer and oligonucleotide sequences are available as Supplementary materials.

Following sequencing, an average of 38 million reads were obtained from the three patient libraries, with 35–42% of reads mapping to target regions (Supplementary Table 2). This low on-target rate was not unexpected and has been observed with small hybrid capture sequencing panels with low input DNA, impacting both assay cost-efficacy and sensitivity [17,18]. We hypothesized that the on-target rate could be improved by two further modifications to the protocol: first, increasing the post-capture washes (four washes with 400 μ l WB2 for 20 min each) to remove mismatched and thus weakly binding library fragments; and second, performing dCAP-Seq (a second capture for 5 h followed by a further five cycles of amplification) to further enrich for targeted libraries [9]. Adopting these strategies, a further 41 patient samples were sequenced, with a significant improvement in on-target rate to a median of 88% ($p < 0.001$, two-tailed Mann–Whitney U test; Supplementary Table 3). Crucially, only 0.16–0.23% of the target region lacked sequencing read coverage, which is advantageous compared with the large gaps of coverage that can be observed with amplicon sequencing panels [19].

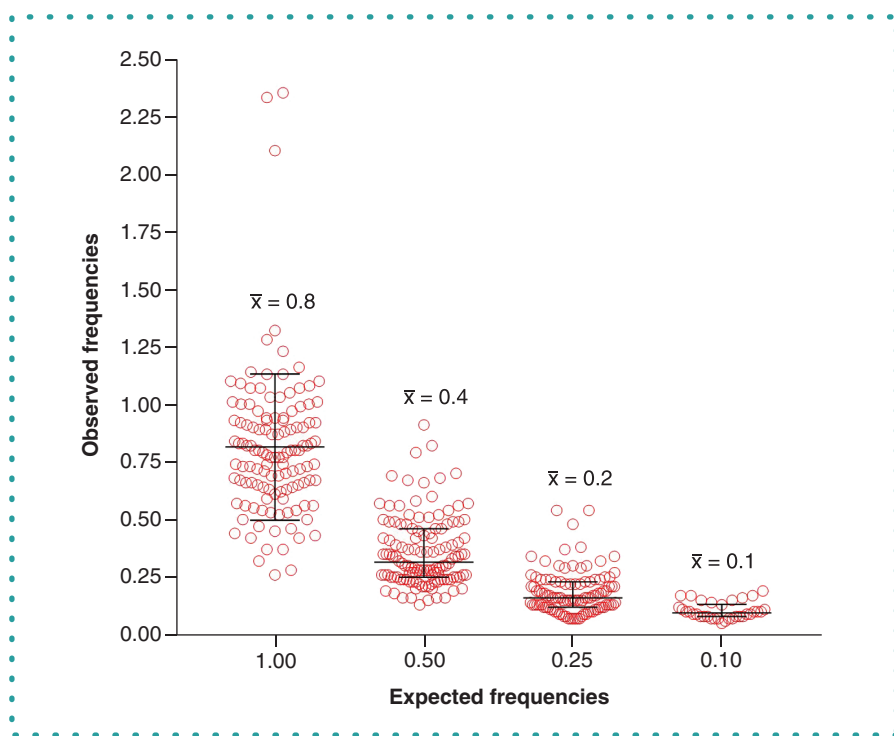


Figure 3. Expected variant allelic fraction versus observed variant allelic fraction. Results are shown for the 133 variants identified in the cfDNA mixing experiment. Mean and 95% CI indicated.
cfDNA: Cell-free DNA.

Table 1. Sensitivity and specificity calculations at various variant allele fractions generated during the cfDNA spike-in experiment.

Variant allelic fraction	Sensitivity	Specificity
1%	93%	99.5%
0.5%	93%	99.5%
0.25%	78.5%	99.5%
0.1%	25%	99.5%

To further improve assay performance and cost efficiency, molecular barcoding and associated family sizes were also assessed to ensure we were not over-sequencing or over-amplifying the patient libraries. A minimum of three reads are required to form a family, with a consensus call only made if at least two of the three (>66%) share the same base at the same position; smaller family sizes (<10 reads) are therefore optimal for error correction, with larger families providing no further information and only serving to increase sequencing costs. We assessed family sizes across the 41 patient samples and found that with a median of 59.2 million reads obtained per sample, over 78% of families were between three and ten reads (Supplementary Figure 1). This confirms the optimal use of molecular barcoding.

Assay limit of detection, sensitivity & specificity

A median of 70 million reads were generated per spike-in sample. After deduplication and error correction, a median depth of 2632× was obtained (Supplementary Table 4). The expected versus observed allelic frequencies for each titration are shown in Figure 3. Assay sensitivity was determined for each simulated allelic frequency by dividing the number of identified true positive SNPs by the total number of true positives, while specificity was determined by dividing the number of identified true negatives by the identified true negatives plus the number of false positives (loci at which alleles other than those already expected were called) (Table 1).

From our spike-in experiment, we determined that our assay had a sensitivity of 93% and specificity greater than 99% for VAFs of 0.5% or above. The assay retained high specificity for lower VAFs while still having the capability to identify variants down to 0.25% allelic frequency.

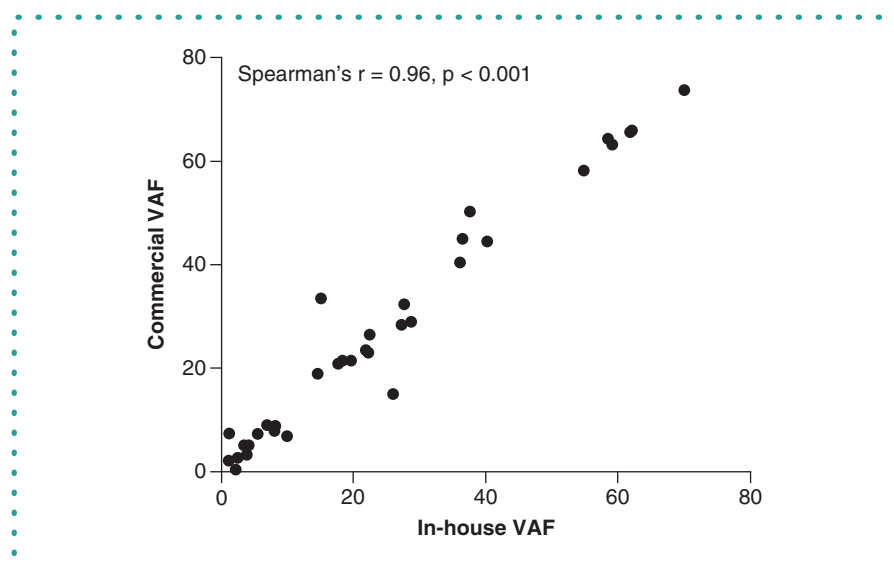


Figure 4. Allelic frequencies (%) of somatic variants identified in both the in-house and commercial targeted panel assays (n = 35). Spearman rank correlation used (two-tailed, nonparametric data). VAF: Variant allelic fraction.

Assay concordance with commercial assay

cfDNA and DNA extracted from matched buffy coat samples from 41 men with mCRPC were concurrently sequenced by an independent commercial laboratory using an error-corrected hybrid capture panel paired with Illumina deep sequencing. The full coding regions of 18 genes were shared between the two panels. The median error-corrected depth for the commercial assay was 7812×.

Using our in-house assay, a total of 63 CNVs were identified across 12 genes, representing 41 losses and 22 gains (Supplementary Table 5). The most common alterations were *AR* gain (8/41; 20%), and *PTEN* and *RB1* loss (both 14/41; 34%). 51 of these CNVs were concurrently identified using the commercial assay, leading to an assay concordance of 81%. The 12 nonconcordant CNVs comprised 8 copy number losses and 4 gains.

We also compared detection of somatic variants between the platforms. The identification of ultra-rare somatic variants (VAFs <1%) can vary significantly between platforms [20–22]; for example, a recent study comparing four different NGS liquid biopsy assays found an average paired concordance rate of only 55% [21]. The variable performance between individual assays is likely attributable to a lack of standardization across the liquid biopsy workflow, including plasma processing, cfDNA input, on-target rate, sequencing read depth, uniformity of coverage, and bioinformatics processing pipelines [23–25]. Stochastic sampling errors further reduce the likelihood of identifying these ultra-rare variants. Such discrepancies highlight the requirement for standardized methodologies in future [26].

For these reasons, we only assessed concordance between our in-house and commercial assay for variants at a VAF of >1% (Supplementary Table 6). Of the 38 somatic variants (including single nucleotide variants and small insertions/deletions) identified across 7 genes using our assay, 35 (92%) were validated using the commercial assay. The three unconfirmed alterations were *AR* T878A (1.2% VAF), *AR* V716M (44% VAF), and *CTNNB1* I588V (6% VAF). All three variants were subsequently confirmed in a serial plasma sample sequenced and analyzed as above (Supplementary Table 6). For concordant variants, VAF was highly correlated between the two assays (Figure 4), with a correlation coefficient of 0.96 ($p < 0.001$, Spearman rank correlation).

Limitations of this report include the lack of orthogonal validation data for ultra-low-frequency variants in the cohort of mCRPC patients and the limited information available on the commercial assay methodology, including access to raw sequencing data.

We demonstrate the high sensitivity and specificity of a customized, end-to-end targeted sequencing assay using optimized off-the-shelf reagents with molecular barcoding technology, along with readily available bioinformatics tools. Our assay can reliably detect ultra-rare variants (under 1% VAF) with a low false-positive rate, and exhibited high concordance with an existing commercial assay.

Future perspective

High-throughput ctDNA analysis has the potential to offer minimally invasive, cost-effective and comprehensive profiling of tumor landscapes in both the research and clinical settings, and is likely to continue to rise in popularity. Further technological advancements and increases in assay sensitivity will allow increased use of liquid biopsies in early stage diagnosis, minimal residual disease monitoring and the identification of small subpopulations of treatment-resistant clones. As such, there exists a growing need to identify standardized protocols for ctDNA analysis for the future clinical implementation of liquid biopsies such as the one described here.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2020-0045

Author contributions

H Fettke: acquisition of data, analysis and interpretation of data, statistical/bioinformatic analysis, drafting and approval of manuscript. J Steen: statistical/bioinformatic analysis, drafting and approval of manuscript. E Kwan: acquisition of data, manuscript revision and approval. P Bukczynska: acquisition of data, manuscript revision and approval. S Keerthikumar: statistical/bioinformatic analysis, manuscript revision and approval. D Goode: statistical/bioinformatic analysis, manuscript revision and approval. M Docanto: acquisition of data, manuscript revision and approval. N Ng: acquisition of data, administrative support, manuscript revision and approval. L Martelotto: acquisition of data, supervision, manuscript revision and approval. C Hauser: analysis and interpretation of data, manuscript revision and approval. M Southey: supervision, manuscript revision and approval. A Azad: concept and design, supervision, manuscript revision and approval. T Nguyen-Dumont: study design, supervision, data analysis, manuscript revision and approval.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Acknowledgments

The authors acknowledge S Wilson from Monash Bioinformatics Platform for this work and the Australian Genome Research Facility for their services.

Financial & competing interests disclosure

H Fettke: Australian Government Research Training Program (RTP) Scholarship. EM Kwan: NHMRC Postgraduate Scholarship, Monash University Postgraduate Publications Award. Melissa Southey: National Health and Medical Research Council (NMHRC, Australia) Senior Research Fellowship (APP1155163). A Azad: NHMRC Project Grant, Victorian Cancer Agency Clinical Research Fellowship, Astellas Investigator-Initiated Grant. T Nguyen-Dumont: National Breast Cancer Foundation (Australia) Career Development Fellowship (ECF-17-001). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: • of interest

- 1 Fettke H, Kwan EM, Azad AA. Cell-free DNA in cancer: current insights. *Cell. Oncol.* 42(1), 13–28 (2019).
 - Detailed contemporary review on cell-free DNA in the context of cancer.
- 2 Wyatt AW, Annala M, Aggarwal R *et al.* Concordance of circulating tumor DNA and matched metastatic tissue biopsy in prostate cancer. *J. Natl Cancer Inst.* 109(12), doi: 10.1093/jnci/djx118 (2017).
 - Demonstrated that cfDNA profiling can capture clinically relevant mutational heterogeneity in advanced prostate cancer that may be missed during conventional biopsy analysis.
- 3 Seoane J, De Mattos-Arruda L. The challenge of intratumour heterogeneity in precision medicine. *J. Intern. Med.* 276(1), 41–51 (2014).
- 4 Kwan EM, Fettke H, Docanto MM *et al.* Prognostic utility of a whole-blood androgen receptor-based gene signature in metastatic castration-resistant prostate cancer. *Eur. Urol. Focus* S2405–4569(19), 30139–7 doi:<https://doi.org/10.1016/j.euf.2019.04.020> (2019).
- 5 Azad AA, Volik SV, Wyatt AW *et al.* Androgen receptor gene aberrations in circulating cell-free DNA: biomarkers of therapeutic resistance in castration-resistant prostate cancer. *Clin. Cancer Res.* 21(10), 2315–2324 (2015).
- 6 Lehmann-Werman R, Neiman D, Zemmour H *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl Acad. Sci. USA* 113(13), E1826–E1834 (2016).
- 7 Luo J, Shen L, Zheng D. Diagnostic value of circulating free DNA for the detection of EGFR mutation status in NSCLC: a systematic review and meta-analysis. *Sci. Rep.* 4, 6269 (2014).
- 8 Jiang XW, Liu W, Zhu XY, Xu XX. Evaluation of EGFR mutations in NSCLC with highly sensitive droplet digital PCR assays. *Mol. Med. Rep.* 20(1), 593–603 (2019).
- 9 Zhang Y, Song J, Day K, Absher D. dCATCH-Seq: improved sequencing of large continuous genomic targets with double-hybridization. *BMC Genomics* 18(1), 811 (2017).
 - First description of performing a double-capture protocol.
- 10 Cronn R, Knaus BJ, Liston A *et al.* Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99(2), 291–311 (2012).
- 11 Underhill HR, Kitzman JO, Hellwig S *et al.* Fragment length of circulating tumor DNA. *PLoS Genet.* 12(7), e1006162 (2016).
- 12 Cerami E, Gao J, Dogrusoz U *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2(5), 401–404 (2012).
- 13 Lai Z, Markovets A, Ahdesmaki M *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 44(11), e108–e108 (2016).
 - First description of VarDict.
- 14 Bozeman MGH, Inc. VarSeq™ [Software]. <http://www.goldenhelix.com>
- 15 Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47(D1), D886–D894 (2018).

- 16 Albrelli I, Generali D, Jermann P *et al.* Cell-free DNA analysis in healthy individuals by next-generation sequencing: a proof of concept and technical validation study. *Cell Death Dis.* 10(7), 534 (2019).
- 17 Nguyen HT, Tran DH, Ngo QD *et al.* Evaluation of a liquid biopsy protocol using ultra-deep massive parallel sequencing for detecting and quantifying circulation tumor DNA in colorectal cancer patients. *Cancer Invest.* 38(2), 85–93 (2020).
- 18 Yoon JG, Hahn HM, Choi S *et al.* Molecular diagnosis of craniosynostosis using targeted next-generation sequencing. *Neurosurgery* doi:10.1093/neuros/nyz470 (2019).
- 19 Samorodnitsky E, Jewell BM, Hagopian R *et al.* Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. *Hum. Mutat.* 36(9), 903–914 (2015).
- 20 Kuderer NM, Burton KA, Blau S *et al.* Comparison of 2 commercially available next-generation sequencing platforms in oncology. *JAMA Oncol.* 3(7), 996–998 (2017).
- Demonstrated the low concordance seen between commercial next-generation sequencing assays at low allelic frequencies.
- 21 Stetson D, Ahmed A, Xu X *et al.* Orthogonal comparison of four plasma NGS tests with tumor suggests technical factors are a major source of assay discordance. *JCO Precis. Oncol.* 3, 1–9 (2019).
- Demonstrated the low concordance seen between commercial next-generation sequencing assays at low allelic frequencies.
- 22 Taavitsainen S, Annala M, Ledet E *et al.* Evaluation of commercial circulating tumor DNA test in metastatic prostate cancer. *JCO Precis. Oncol.* 3, 1–9 (2019).
- 23 Page K, Shaw JA, Guttery DS. The liquid biopsy: towards standardisation in preparation for prime time. *Lancet Oncol.* 20(6), 758–760 (2019).
- 24 Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. *Clin. Chem.* 61(1), 112–123 (2015).
- 25 Devonshire AS, Whale AS, Gutteridge A *et al.* Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal. Bioanal. Chem.* 406(26), 6499–6512 (2014).
- 26 Greytak SR, Engel KB, Parpart-Li S *et al.* Harmonizing cell-free DNA collection and processing practices through evidence-based guidance. *Clin. Cancer Res.* doi:10.1158/1078-0432.ccr-19-3015 (2020).
- Recent recommendations for specimen handling and cfDNA sample quality assessment.