Title:

Gene Copy Number Estimation From Targeted Next Generation Sequencing Of Prostate Cancer Biopsies: Analytic Validation and Clinical Qualification.

Authors:

George Seed[1]*, Wei Yuan[1]*, Joaquin Mateo[1], Suzanne Carreira[1], Claudia Bertan[1], Maryou Lambros[1], Gunther Boysen[1], Roberta Ferraldeschi[1,2], Susana Miranda[1], Ines Figueiredo[1], Ruth Riisnaes[1], Mateus Crespo[1], Daniel Nava Rodrigues[1], Eric Talevich[4], Dan Robinson[3], Lakshmi P Kunju[3], Yi-Mi Wu[3], Robert Lonigro[3], Shahneen Sandhu[1,2], Arul Chinnayan[4], Johann de Bono[1,2].

AFFILIATIONS:

1. The Institute of Cancer Research, London, United Kingdom; 2. The Royal Marsden NHS Foundation Trust, London, United Kingdom; 3. Michigan Centre for Translational Pathology, Ann Arbor, Michigan, USA; 4. University of California San Francisco, San Francisco, USA.

* *Indicates these authors contributed equally.*

Running Title:

Targeted-seq and copy-number analysis of CRPC biopsies.

***CORRESPONDING AUTHOR:**

Professor Johann S de Bono, MB ChB, MSc, FRCP, PhD, FMedSci
Regius Professor of Experimental Cancer Medicine
Division of Clinical Studies, The Institute of Cancer Research
The Royal Marsden NHS Foundation Trust
London SM2 5NG
United Kingdom
Telephone: +44 (0)2087224028; Fax: +44 (0)2086427979
Email: *johann.de-bono@icr.ac.uk*

ACKNOWLEDGEMENTS: This study was supported by Cancer Research UK, Prostate Cancer UK, and the Stand Up To Cancer – Prostate Cancer Foundation Prostate Dream Team Translational Research Grant (Grant Number: SU2C-AACR-DT0712). Stand Up To Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. Support was also provided by an Experimental Cancer Medicine Centre grant and a Medical Research Council/Prostate Cancer UK fellowship to J. Mateo. G. Seed was supported by a Prostate Cancer UK PhD studentship (TLD-S15-006), and G. Boysen was supported by a Marie Curie International Incoming Fellowship (625792). This study was also supported by research funding from the EU FP7 project CTCTrap #305341. We thank the staff at the Prostate Targeted Therapy Group and all the patients who agreed to participate in this study.

Statement of Translational Relevance

Detection of oncogenic genomic aberrations has led to novel therapeutic strategies for treating cancer. We recently showed that 20-30% of metastatic prostate cancers have deleterious defects in DNA repair machinery affecting genes such as *BRCA2*, *ATM*, *PALB2*, *BRCA1* and *CDK12.* These aberrations are associated with sensitivity to PARP inhibitors. Robust methods are needed to identify these tumor subtypes. We have previously described a targeted next generation sequencing (NGS) method that identifies single nucleotide variants (SNVs) and copy number aberrations, with precision, from formalin-fixed, paraffin-embedded. tumor biopsy samples. We now describe the bioinformatics pipeline for this robust high throughput, cost efficient and rapid solution. We have implemented these methods into clinical trial protocols to enable patient stratification.

Abstract

Purpose

Precise detection of copy number aberrations (CNAs) from tumor biopsies is critically important to the treatment of metastatic prostate cancer. The use of targeted panel next generation sequencing (NGS) is inexpensive, high throughput and easily feasible, allowing single nucleotide variant calls, but CNA estimation from this remains challenging..

Experimental Design

We evaluated CNVkit for CNA identification from amplicon-based targeted NGS in a cohort of 110 fresh castration resistant prostate cancer biopsies, and used capture based whole exome sequencing (WES), array comparative genomic hybridization (aCGH), and fluorescent in situ hybridization (FISH) to explore the viability of this approach.

Results

We showed that this method produced highly reproducible CNA results (r=0.92), with the use of pooled germline DNA as a coverage reference supporting precise CNA estimation. CNA estimates from targeted next generation sequencing were comparable with WES (r=0.86) and aCGH (r=0.7); for key selected genes (*BRCA2*, *MYC, PIK3CA*, *PTEN* and *RB1)* CNA estimation correlated well with WES (r = 0.91) and aCGH (r = 0.84) results.

The frequency of CNAs in our population was comparable to that previously described (ie deep deletions: BRCA2 4.5%; RB1 8.2%; PTEN 15.5%; amplification: AR 45.5%; gain: MYC 31.8%). We also showed, utilizing FISH, that CNA estimation can be impacted by intra-tumor heterogeneity and demonstrated that tumor microdissection allows NGS to provide more precise CNA estimates.

Conclusion

Targeted NGS and CNVkit based analyses provide a robust, precise, high throughput and cost effective method for CNA estimation for the delivery of more precise patient care.

Introduction:

Prostate cancer is a highly heterogeneous disease with distinct genomic underpinnings[1] and generally presents a modest number of single point mutations and small insertions and deletions (indels) at approximately four per megabase[5,6], but frequently has large-scale copy number and structural alterations[7,8]. We recently showed that 20-30% of prostate tumours bear defects in DNA repair genes that render them sensitive to treatment with PARP inhibition, and that these are commonly mediated through deletions in these tumor suppressors[9].

High throughput, inexpensive, multiplex biomarker assays for molecular stratification are needed to guide therapeutic choices and support clinical trial design. While high-coverage whole exome or genome sequencing can reliably assess tumour genomics, cost and bioinformatic demands currently restrict their routine clinical implementation to stratify patients for treatment. Targeted next-generation sequencing represents an opportunity for implementing genomics in clinical practice with the advantages of a rapid turnaround time, lower cost and the ability to concurrently analyse multiple genes with a limited bioinformatic analysis burden. Existing approaches to call somatic and germline single nucleotide variants and short indels from targeted sequencing data are well-established and have led to a number of FDA biomarker-dependent drug approvals including the PARP inhibitor olaparib in the last decade[10,11]. However, in the case of PARP inhibition for homologous recombination repair defective cancers, screening for tumours with these DNA repair defects relies on detecting gene copy number changes.

In this study, we identify gene copy number aberrations (CNA), based on targeted amplicon sequencing of a focused biomarker gene panel dedicated to identifying DNA repair defects. We used a patient cohort of metastatic castration-resistant prostate cancer to explore the utility of this approach. This targeted gene panel focused on the study of 113 genes important to DNA repair machinery including *BRCA2*, *BRCA1, ATM*, *PALB2* and *CDK12* as well as other important potentially actionable prostate cancer genes including *AR*, *SPOP*, *PIK3CA*, *PTEN*, *AKT1*, *AKT2* and *MYC*.[1,9]

Methods:

Patient Selection and Sample Preparation

Patient tumour (biopsies) and germline (buccal swab and saliva) samples were collected as part of the Royal Marsden ethics committee approved CCR2472 protocol. Samples were collected, annotated, stored and reviewed as described previously[9]. All samples were collected between 25/07/2015 and 17/10/2016.

Briefly, biopsies were paraffin embedded, DNA from biopsies was extracted using the QIAamp DNA FFPE Tissue kit (Qiagen), Quant-iT Picogreen High-Sensitivity double-stranded DNA (dsDNA) Assay Kits (Invitrogen, ThermoFisher Scientific Co) and QC check by FFPE QC kit (Illumina).

Targeted Panel Design

A customised Generead v2 DNAseq Panel (Qiagen) panel was used for library construction; the exonic regions of 113 genes were included in the targeted panel, these being selected for being potentially actionable and/or involved in DNA damage repair processes and/or in prostate carcinogenesis[9] **(Supplemental Table 1).** The panel covered ~564kb of the human genome, and each of the 113-genes was covered by an average of 18.69 probes (sd 13.64).

Targeted and Exome Sequencing

Targeted amplicon sequencing was performed using the Illumina MiSeq™ platform following the manufacturers' protocol. FASTQ files were generated using the Illumina MiSeq Reporter v2.5.1.3™ [12,13]. Reads were aligned to the human reference genome (hg19) using BWA. Whole exome sequencing was performed at the University of Michigan as previously described[1].

FastQC (v0.11.2) and Samtools (0.1.19) were used to assess sequencing quality. Targeted sequencing samples were excluded from copy number analysis on the basis of: insufficient (<0.5 million) total read counts, low (<95%) percentage of properly paired reads, and low (<99.99%) percentage of on-target reads. Exome sequencing samples were rejected if the FastQC per-base quality score for 75% of the reads was less than Q20 over the first 80 bases, and alignment quality was monitored with Picard, as described previously[1].

Copy Number Aberrations

We used the software CNVkit (v0.3.5)[14] to analyse sequencing coverage and copy number in the aligned sequencing reads from targeted amplicon sequencing of tumour and germline samples. Sequencing coverage of targeted regions in germline samples was assessed and used to create pooled reference data that included the technical variability at each covered region. Regions that were poorly captured or mapped were masked from further analysis. The analyses of germline samples were also performed with CNVkit to validate sample quality. The read depths of tumour samples were accessed, normalised (corrected for GC content, target footprint size and spacing, and repetitive sequences), and individually compared to the reference, and the circular binary segmentation (CBS) algorithm[15] was used

to infer copy number segments. Very small copy changes (≤3 bins) were treated as artifacts. Poor samples were also identified through high (>0.8) values of the segment inter-quartile range (IQR), a metric for the spread of bin-level copy ratios within each segment following application of the CBS algorithm. Copy number segments were annotated to genes, and regions bearing a Log_2 ratio of at least ±0.4 were identified as suggestive of shallow deletions or gains[14]. Segments with $Log_2 < -1.2$ were classified as deep deletions, and those with $Log_2 > 2$ were classified as amplifications. Experimental noise was identified as a Log_2 ratio standard deviation of ~0.2.

Whole-exome sequencing CNA data was extracted from a previously reported Su2C-PCF dataset[1]. Briefly, Log_2 ratios were derived on a per-gene basis from circular-binary segmented, Lowess-normalised Log_2 transformed coverage ratio between each tumour and matched normal sample.

Array Comparative Genomic Hybridization

Tumour DNA from prostate cancer patients and male reference DNA from Agilent were amplified using Sigma WGA2 kit (Sigma-Aldrich, MO, USA) according to the manufacturer's recommendations and the amplified DNA was quantified by the Qubit fluorometric quantitation method (Thermo Fisher Scientific, MA, USA). 500ng of amplified tumour DNA was then fluorescently labeled with Cy5, and male reference DNA labeled with Cy3, using the SureTag Complete DNA Labeling Kit (Agilent Technologies CA, USA). Labeled DNA was subsequently hybridised utilizing the Agilent SurePrint G3 Human CGH Microarray Kit, 4x180K according to the manufacturer's instructions. The slides were then scanned and analysed using the CytoGenomics Software v 4.0.3.12 (Agilent Technologies CA, USA). In order to compare aCGH and NGS results we used the Log_2 ratio of the aCGH segments that overlap with genes present in the NGS panel. Genes not part of a copy-altered segment were included in the analysis by using the mean Log_2 ratio for probes, if the value was no greater than 0.25. Genes on the X-chromosome were excluded from comparison with results from the male-reference-based panel data since a female reference was used for aCGH.

FISH

RB1 FISH was performed using a standard FFPE hybridisation method[16] on 3-µM FFPE tissue slices adjacent to hematoxylin and eosin sections that were confirmed to contain a minimum of 50 intact cells. Briefly, RB1 status was determined using Vysis LSI 13 RB1 (13q14) probe (Catalogue # 08L65-020; Abbott Laboratories, Abbott Park, Illinois, U.S.A.) and a reference probe Vysis 13q34 (Catalogue # 05N34-020; Abbott Laboratories, Abbott Park, Illinois, U.S.A.). 19 z-stacks were used to assess cell status.

Results:

Metastatic Prostate Cancer Sample Characteristics

Freshly collected metastatic castration-resistant prostate cancer (mCRPC) biopsies (n=110) were collected from patients with progressing disease (**Supplemental table 2**) and assessed for copy-number changes by sequencing with our targeted panel. Germline DNA samples (buccal swab) were taken from 34 consenting patients to use as a baseline reference and for additional analyses. Tumour biopsies were from: bone (49.1%), lymph nodes (35.5%), liver (9.1%), soft tissue disease (3.6%) and transurethral resection of the prostate (TURP; 3.6%).

CNA Estimation Reproducibility

Accurately assessing the depth of coverage of genes can be biased by high GC content and repetitive regions[14,17], so we first evaluated CNA reproducibility on a per-gene basis between technical replicates. Thirteen samples were sequenced in duplicate to assess technical variation, twelve of these duplicate datasets included repeated library preparation from the same DNA extraction, while one involved resequencing of the same library twice.

We identified 15 genes that produced less concordant results between replicates, including five genes with insufficient number of targeted regions (four or less probes) per gene (*NRAS*, *NFKBIA*, *FANCF*, *FAM46C*, *CDKN1B*), and ten (*MAP2K2, CDKN2A, HRAS, RECQL4, NOTCH1, FGFR3, STK11, TSC2, SMARCA4, AXIN1*) that contained repetitive or polymorphic exonic regions that impeded accurate read alignment and possessed high coverage variability (**Supplementary Figure 1)**. These genes were excluded from downstream CNA analyses.

Following removal of the aforementioned genes, CNA estimates between duplicates showed high correlation (Pearson *r* correlation coefficient = 0.92) (**Figure 1a**). All genes with deep deletions (Log_2 ratio -1.2) or amplifications (Log_2 ratio > 2) had a similar (at least +- 0.4) CNA estimation replicated in the duplicate sample. 95% of samples with a log_2 ratio change of at least +- 0.4 had a similar result in the duplicate

The use of a matched germline sample is commonly used for identifying deviations in coverage ratios, so we sought to compare this with our pooled reference approach. For the pooled reference, we sequenced 34 unmatched germline samples, and assessed the depth of coverage as described previously[14]. Regions with low variability were identified in the pooled reference, and used to weight downstream CNA estimates[14]. This approach produced a highly similar result (overall Pearson *r* = 0.93) when compared with the CNA estimation approach using a matched normal as reference, when matched normal samples were available (**Figure 1b**); however the CNA estimation confidence score (IQR value) was improved when using pooled germline samples (paired *t* test p = 0.05). This was in line with other reports[17] that indicated that using pooled germline sequencing data as reference offered a robust method to identify gene copy number aberrations.

CNA Validation by exome sequencing and aCGH

We cross-compared copy number Log_2 ratios for all evaluable genes on the targeted panel with Log_2 ratios generated by exome-sequencing (n=13) (**Figure 1c**), resulting in a Pearson *r* correlation coefficient of 0.86. Moreover, of all the genes with a Log_2 ratio <-1.2 by targeted sequencing, in keeping with putative 'deep' deletions, exhibited a similar (at least +- 0.4) result in the exome-sequencing. For example, one sample bore AR amplification (Log_2 ratio 5.7 in targeted NGS and 4.91 in WES) alongside *WRN* deletion (-1.53 versus -1.17) and *ATM* deletion (-0.68 versus -0.56). We also compared Log_2 ratios from aCGH segments (n=9) with gene copy number estimations by panel NGS among the genes covered by both platforms, and found log_2 ratios were also concordant (**Figure 1d**), with a Pearson *r* value of 0.7.

We then sought to further validate copy-number calls for specific genes of clinical relevance such as *BRCA2*, *MYC, PIK3CA*, *PTEN* and *RB1* (**Figures 1e and 1f**). For these genes, we found that our panel produced results highly concordant with exome-sequencing and aCGH data; with Pearson *r* values of 0.91 and 0.84 respectively. All clinically relevant deletions of *BRCA2* in samples with exome or aCGH data were also identified in our panel results (**Table 1**).

CNA Profiles in CRPC

In the analyses of this cohort of advanced prostate cancer samples, the prevalence of the CNAs detected with this assay was consistent with that of previously reported studies (**Figure 2**). CNAs of *AR* (amplification; $Log_2 > 2$) were found in 45.5% (50/110) of tumour samples. Broad deletions of chromosome 13[1] involving the shared loss of *BRCA2* and *RB1* were detected as previously described with 23/26 (88%) of samples with any *BRCA2* loss also having some loss of *RB1*, while overall 23/66 (34.8%) of samples with *RB1* loss had loss of *BRCA2*) (**Figure 2a**).

Overall, the evaluation of the deep deletions of commonly aberrant CRPC genes revealed results in keeping with previous reports from two recent studies of CRPC genomics[1,2]; *PTEN* genomic deep deletions were detected in 17/110 (15.5%), *BRCA2* deep deletion in 5/110 (4.5%), and *RB1* deep deletion in 9/110 (8.2%) of CRPC samples. In addition, gain of the *MYC* locus was found in 36/110 (32.7%). We found that our values were in line with expected values (**Figure 2b**).

Exploring tumour purity and heterogeneity

In order to further explore how tumour purity altered the capacity of the assay to detect CNAs, we pursued the serial dilution of tumour DNA acquired from a sample estimated by pathology review as having 80% tumour content, with same-patient germline DNA. This sample harboured a somatic, *BRCA2* homozygous deletion (independently confirmed WES). We then sequenced the resulting dilutions (**Figure 3a**) and found that homozygous deletions were easily distinguishable with tumour content purities >60%, and that clonal CNAs were detectable with purity as low as 30%.

We also performed micro-dissection of a tumour biopsy (pathologist tumour purity estimate ~40%) bearing borderline *BRCA2* homozygous deletion (Log_2 ratio of -1.15), and re-sequenced the resulting tumour-enriched sample (purity estimate ~80%). This step further shifted the CNA estimates away from normal, with the Log_2 ratio for *BRCA2* dropping to -2.82, an unambiguous deletion (**Figure 3b**).

The Log_2 copy ratios estimated from sequencing a bulk of cells, as in our protocol, can be biased by tumour sample heterogeneity. To assess this relationship, we performed FISH assays of the *RB1* gene, counting the number of cells where each copy number (0 through 3+) was observed (**Figure 3c**). Our results showed an association between the proportion of cells with 1 or 0 copies of *RB1* (hemizygous or homozygous deletion) and putative copy loss inferred from our targeted sequencing panel (unpaired t-test p-value 0.02, n=18, df=15.884). However, the full complexity of aneuploidy in individual tumour cells is masked by bulk tissue sequencing, which effectively averages the estimated DNA content across all cells in a sample. FISH analyses revealed that many tumour samples contained a mixture of cells with different copy number changes at the RB1 locus, in highly heterogeneous cell populations (**Supplemental Figure 2**).

Discussion

The ability to rapidly, inexpensively, and accurately identify cancers with actionable genomic defects in tumour biopsies is now becoming critically important for advanced prostate cancer care. We have previously shown that a proportion of metastatic castration-resistant prostate cancers have DNA repair defects and are sensitive to treatment with PARP inhibitors[9], with emerging studies of platinum-based chemotherapies also indicating important antitumour activity in this population[18]. These and other data support further studies of genomic stratification for treatment selection for patients suffering from mCRPC. Here we demonstrate the analytic validation and clinical use of a targeted next generation sequencing assay that allows for copy-number screening for predictive biomarkers of response, enabling the integration of more precise patient stratification into trial design.

Our method reliably estimates gene copy number changes from amplicon based targeted next generation sequencing with the acquired results being comparable to widely used assays including exome sequencing, array comparative genomic hybridization, digital droplet PCR and FISH.[19]. Critically, the data processing time for this panel-based CNA estimation is substantially shorter of that for exome sequencing and can be easily completed even with a personal computer. In addition, we have confirmed that using pooled germline samples as a reference can produce equivalent and robust results to utilizing matched normal samples. This feature provides the possibility of diagnosing

patients with only tumour samples for whom germline DNA is not available, and can dramatically reduce the overall sequencing costs of large studies.

This approach is currently being applied in the TOPARP-B (NCT01682772) study as part of an ongoing attempt to clinically qualify DNA repair pathway aberrations as a predictive biomarker for treatment with PARP inhibition with olaparib[18]. However, this assay has utility beyond this study. The gene panel we utilised covers several key genomic aberrations that are of emerging interest as possible clinical predictive biomarkers that may be actionable in prostate cancer, including loss of *PTEN,* gain of *PIK3CA* (both may sensitise tumours to PI3K/AKT pathway inhibition[20,21]), and amplification of *MYC* (MYC driven tumours have been reported to sensitize to AURA and BET inhibition and can result in replication stress[22]). Prospectively identifying tumours bearing these CNAs and other deleterious mutations, utilizing such validated methods, is likely to become critically important in defining future treatment strategies for this commonest of male cancers.

We explored the functional limits of assessing CNAs and found that our pipeline can confidently estimate deletion when tumour content is reduced to 60% (homozygous and heterozygous). As tumour content decreases, the pipeline struggles to differentiate between homozygous and heterozygous deletions, and we estimate that interpreting results from samples with <30% purity may be extremely challenging for targeted NGS panels. We have shown, however, that the use of micro-dissection to enrich tumour content offers a possible solution to resolve these issues.

This approach also suffers from the nature of amplicon based next generation sequencing, including difficulty identifying precise breakpoints, very focal gains or deletions, unequal amplification among genes and coverage variation across polymorphic repeat regions. Furthermore, because the number of single nucleotide variants identified in targeted panels will be much lower when compared to exome or whole genome sequencing, SNV allele frequencies may not be sufficient to assist with copy number estimation or to detect allelic imbalance.

It is important to note that we did not systematically compare software tools in this study, and that different clinical requirements will require individual optimisations. As the field develops, other software tools will emerge to improve the precision of CNA identification in targeted sequencing, and independent validations will be critical in guiding further work.

In conclusion, we have described the validation and clinical application of an assay evaluating 100 genes utilised to assess both deleterious tumour gene copy number, and mutations, from targeted next generation sequencing of prostate cancer samples from the most common metastatic sites, including bone biopsies. We have validated these results by orthogonal methods including not only exome sequencing, but also aCGH and FISH. Such targeted panel NGS assays are likely to become central to future patient care in cancer therapy.

References:

1. Robinson D, Van Allen EM, Wu YM, et al: Integrative clinical genomics of advanced prostate cancer. Cell 161:1215-28, 2015
2. Kumar A, Coleman I, Morrissey C, et al: Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med 22:369-78, 2016
3. Gao J, Aksoy BA, Dogrusoz U, et al: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6:pl1, 2013
4. Cerami E, Gao J, Dogrusoz U, et al: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2:401-4, 2012
5. Baca SC, Prandi D, Lawrence MS, et al: Punctuated evolution of prostate cancer genomes. Cell 153:666-77, 2013
6. Barbieri CE, Baca SC, Lawrence MS, et al: Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 44:685-9, 2012
7. Mitchell T, Neal DE: The genomic evolution of human prostate cancer. Br J Cancer 113:193-8, 2015
8. Grasso CS, Wu YM, Robinson DR, et al: The mutational landscape of lethal castration-resistant prostate cancer. Nature 487:239-43, 2012
9. Mateo J, Carreira S, Sandhu S, et al: DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. N Engl J Med 373:1697-708, 2015
10. Ledermann J, Harter P, Gourley C, et al: Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a

preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. Lancet Oncol 15:852-61, 2014

11. Fong PC, Boss DS, Yap TA, et al: Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med 361:123-34, 2009

12. Ong M, Carreira S, Goodall J, et al: Validation and utilisation of high-coverage next-generation sequencing to deliver the pharmacological audit trail. Br J Cancer 111:828-36, 2014

13. Carreira S, Romanel A, Goodall J, et al: Tumor clone dynamics in lethal prostate cancer. Sci Transl Med 6:254ra125, 2014

14. Talevich E, Shain AH, Botton T, et al: CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol 12:e1004873, 2016

15. Olshen AB, Venkatraman ES, Lucito R, et al: Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5:557-72, 2004

16. Ferraldeschi R, Nava Rodrigues D, Riisnaes R, et al: PTEN protein loss and clinical outcome from castration-resistant prostate cancer treated with abiraterone acetate. Eur Urol 67:795-802, 2015

17. Grasso C, Butler T, Rhodes K, et al: Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. J Mol Diagn 17:53-63, 2015

18. Cheng HH, Pritchard CC, Boyd T, et al: Biallelic Inactivation of BRCA2 in Platinum-sensitive Metastatic Castration-resistant Prostate Cancer. Eur Urol 69:992-5, 2016

19. Strom CM, Rivera S, Elzinga C, et al: Development and Validation of a Next-Generation Sequencing Assay for BRCA1 and BRCA2 Variants for the Clinical Laboratory. PLoS One 10:e0136419, 2015

20. Yamamoto Y, De Velasco MA, Kura Y, et al: Evaluation of in vivo responses of sorafenib therapy in a preclinical mouse model of PTEN-deficient of prostate cancer. J Transl Med 13:150, 2015

21. Park H, Kim Y, Sul JW, et al: Synergistic anticancer efficacy of MEK inhibition and dual PI3K/mTOR inhibition in castration-resistant prostate cancer. Prostate 75:1747-59, 2015

22. Delmore JE, Issa GC, Lemieux ME, et al: BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell 146:904-17, 2011

Figure Legends

Figure1

Copy number estimation results by targeted NGS and the described bioinformatic method were highly reproducible. Correlations of Log_2 results (overplotted) shown from: **Figure 1a)** Replicate samples (n = 13); **Figure 1b)** Matched germline reference (25 pairs available) against pooled germline reference. Pearson *r* values presented. Comparing tools for generating informative Log_2 ratios showed that targeted panel NGS of CRPCs produced valid results for identification of CNAs when compared with orthogonal methods. Shown are correlations of Log_2 results (overplotted) from: **Figure 1c)** Panel data with exome data (n = 13); and **Figure 1d)** Panel data with aCGH results (n = 9). Pearson *r* values depicted. CNA orthogonal validation focusing on selected genes of therapeutic relevance to CRPC presented as Log_2 ratios in comparison with exome (**Figure 1e)** and aCGH (**Figure 1f)** data from the same samples compared to panel data. Colors indicate gene: BRCA2 = red, MYC = orange, PIK3CA = dark blue, PTEN = light blue, RB1 = yellow. Pearson *r* values are shown.

Figure 2a.

Binned heatmap of copy number aberrations for all evaluated (n = 110) samples with colors defined by log2 ratio thresholds (dark blue, < -1.2; light blue, < -0.4; pink, > 0.4; red, > 2). Rows and columns clustered by complete Euclidean distance for visualization purposes only. Commonly deleted genes

such as PTEN, RB1 and BRCA2 cluster towards the left, while commonly amplified genes such as AR, MYC, BRAF and KRAS trend to the right. Genes which are closely located in chromosomal coordinates (eg CHEK2 and NF2, BLM and MAP2K1,RB1 and BRCA2) often share similar copy-number aberrations.

Figure 2b
Scatter plot showing that the percentage of samples with specific copy-number changes in our targeted sequencing cohort was similar to two recently published studies of CNAs in CRPC[1,2]. Colors indicate type of aberration (red = amplification, pink = gain, blue = deletion). Shape of points indicate study data source. Data was derived from copy number frequency data publicly available from cbioportal[3,4], retrieved at 17/11/2016.

Figure 3a
Serial dilution of a tumour sample with matched germline DNA and resequencing highlights functional limits of copy number detection. Homozygously deleted (BRCA2, PTEN – dark blue), hemizygously deleted (RB1 – light blue), copy-neutral (SPOP – grey) and copy-gained (PIK3CA – pink) genes are shown.

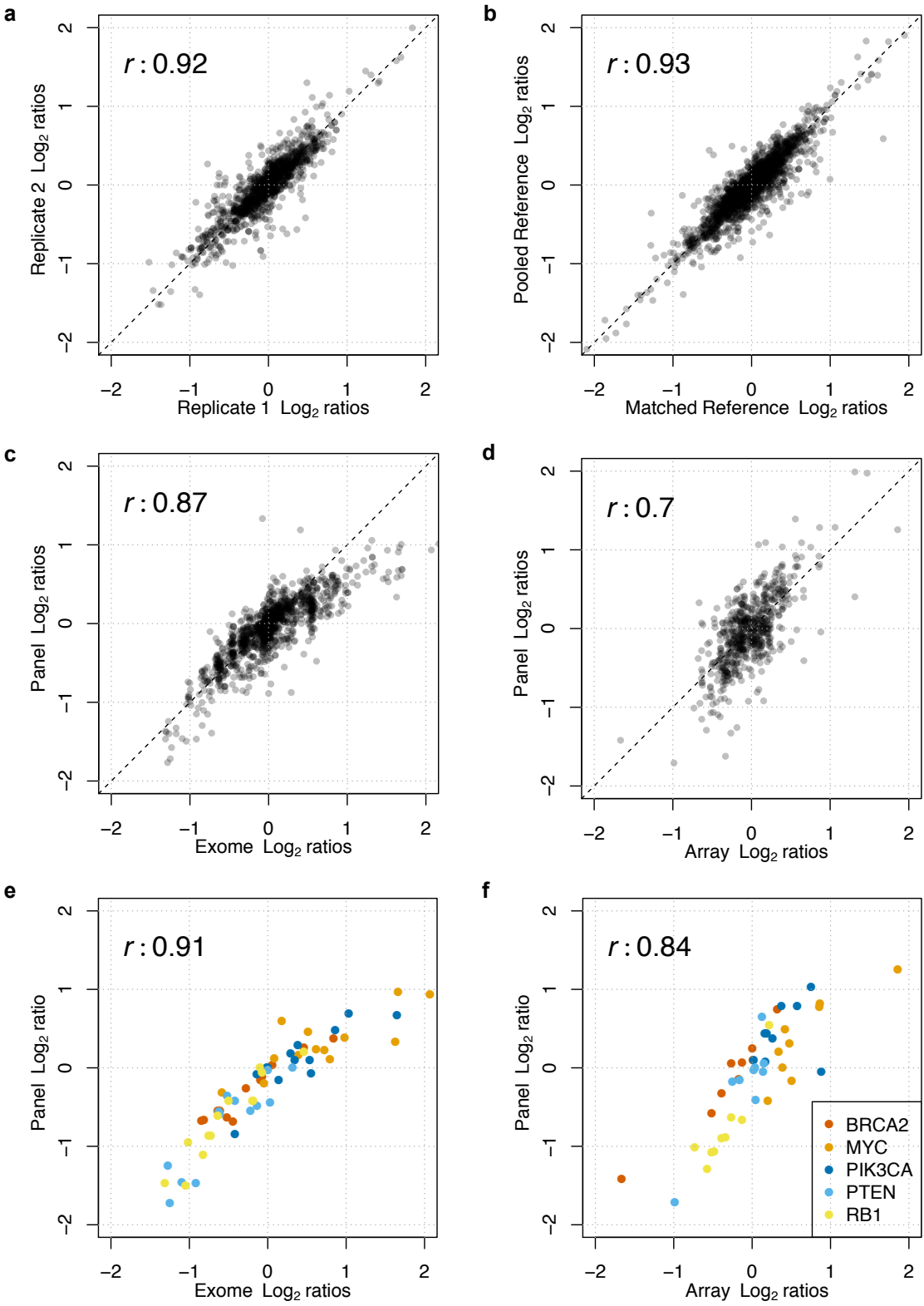
Figure 3b
Micro-dissection of a biopsy to enrich tumour content emphasizes a *BRCA2* deletion, as the segment Log_2 ratio (orange line) shifts from -1.15 to -2.82. Also shown are individual exon-level copy ratios (blue points), from which the segment is assigned.

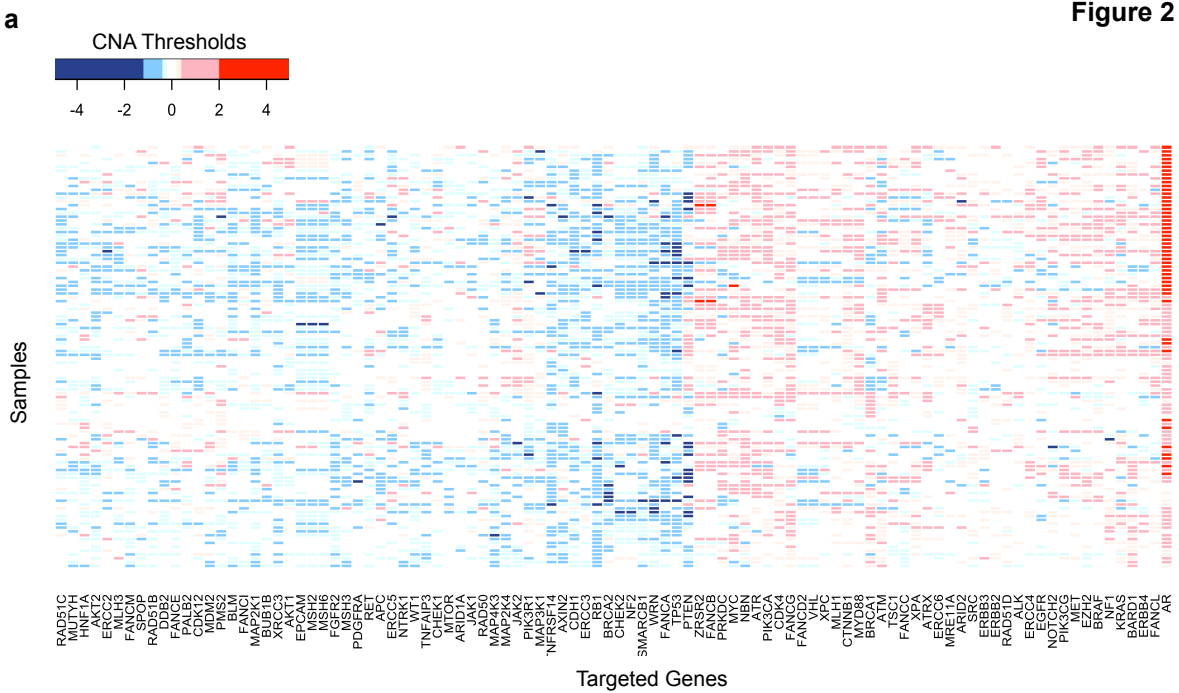
Figure 3c
Stacked bar plot showing RB1-FISH results for 18 samples (50 cells each), with colors indicating observable copies of RB1 for individual cells. Navy, 0 copies; light blue, 1 copy; white, 2 copies; pink, >2 copies. Black bar indicates samples which had a deletion of any kind reported by our targeted NGS protocol.

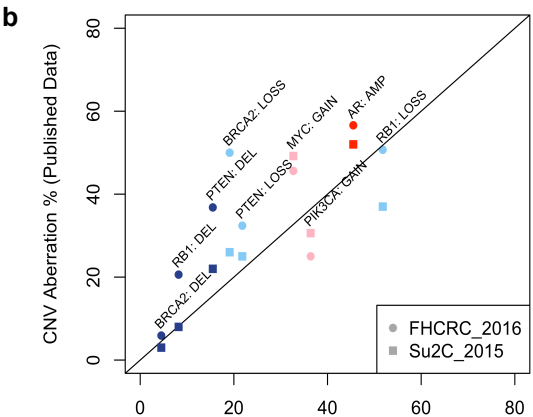
Table 1
Validation of *BRCA2* results from targeted panel CNV analysis with either whole-exome (E) sequencing or array (A) comparative genome hybridisation, log2 ratios are shown. Colours indicate direction of shift away from 0, with losses represented by blue and gains represented by red.

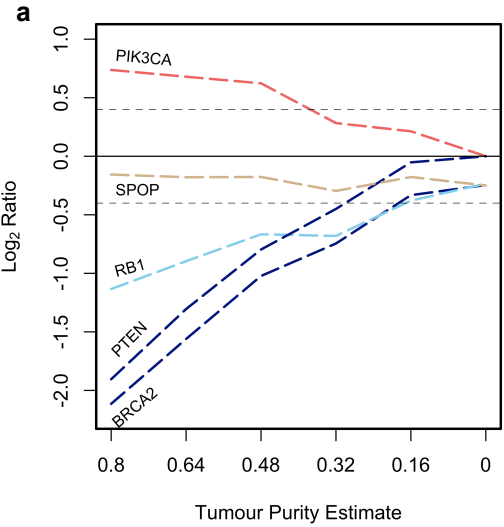
Sample	Panel *BRCA2*	Exome *BRCA2*	aCGH *BRCA2*
p.4.1	-0.670	-0.845	
p.5.0	0.262	0.46	
p.7.0	-0.542	-0.611	
p.8.0	-0.152	-0.095	
p.9.0	-0.682	-0.442	
p.13.0	-0.109	-0.074	
p.14.1	-0.663	-0.815	
p.15.0	-0.546	-0.639	
p.18.0	0.036	0.057	
p.20.0	0.377	0.838	
p.22.0	-0.101	-0.101	
p.23.0	-0.259	-0.277	
p.25.0	-0.633	-0.52	
p.39.0	-0.324		-0.389
p.41.0	0.056		-0.264
p.45.0	-0.139		-0.17
p.110.0	-0.576		-0.516
p.69.0	-1.419		-1.665
p.98.0	0.102		0.008
p.102.0	0.065		-0.128
p.104.0	0.75		0.321
p.38.2	0.468		0.003

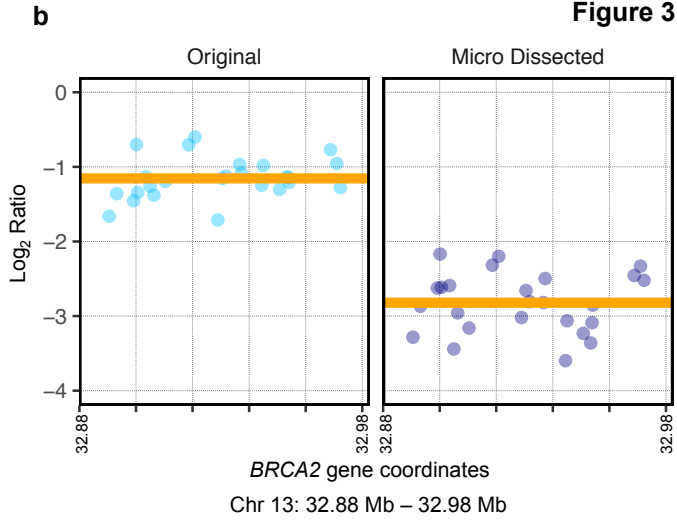
Table 1

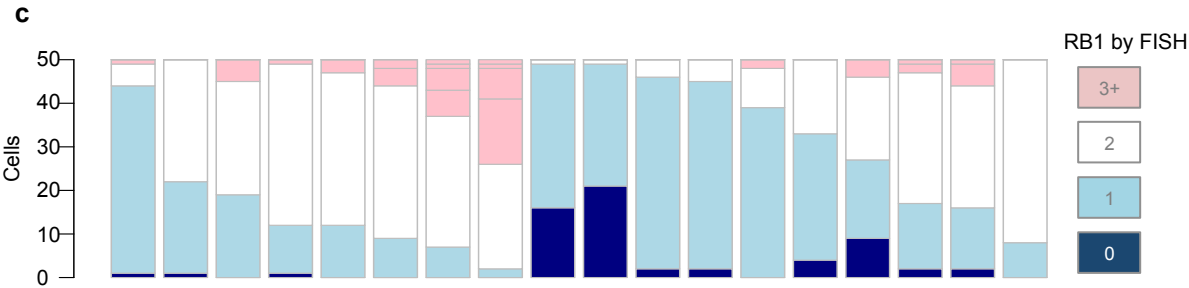












AACR American Association for Cancer Research

Clinical Cancer Research

Gene Copy Number Estimation From Targeted Next Generation Sequencing Of Prostate Cancer Biopsies: Analytic Validation and Clinical Qualification.

George Seed, Wei Yuan, Joaquin Mateo, et al.

Clin Cancer Res Published OnlineFirst July 27, 2017.