Supplementary Methods

Study Design

Our study was designed to determine whether assessment of genomic alterations in the *AR* enhancer and gene body (collectively referred to as *AR*/enhancer) in cell-free DNA (cfDNA) could predict resistance to AR-directed systemic therapy. The sample size of 40 was justified to achieve 90% power by a 2-sided normal test at a 5% alpha to detect a difference of 75% versus 25% rate of resistance for patients with positive versus negative cfDNA results, assuming a 50% rate of *AR*/enhancer alteration in cfDNA[1] and a 5% attrition rate. We obtained peripheral blood at the time of enrollment, which was processed within six hours of phlebotomy for cfDNA analysis. A separate blood sample was submitted for CTC AR-V7 analysis (Genomic Health) in a subset of patients at the discretion of the treating oncologist. Laboratory research investigators were unaware of the AR-V7 status of study participants at the time of cfDNA analysis. For four patients, blood was drawn serially for cfDNA analysis with timepoints being at least two weeks apart and at least one timepoint occurring during AR-directed treatment.

Specimen Collection and Processing

Between 10 and 20 mL of peripheral blood was collected in K_2EDTA Vacutainer tubes (Becton Dickinson) at the time of study enrollment. Tubes were centrifuged at 1,200g for 10 minutes, then plasma separated and centrifuged for another 5 minutes at 1,800g. Plasma was then frozen at -80°C prior to cfDNA processing and analysis. Leukocyte-enriched plasma-depleted whole blood (PDWB) was also collected and frozen at -80°C for isolation of germline genomic DNA as previously described[2,3]. Peripheral blood was separately collected in a subset of patients using collection tubes provided by Genomic Health for the Oncotype DX AR-V7 Nucleus Detect CTC assay. After collection, tubes were immediately sent to Genomic Health for analysis following their protocol.

DNA Isolation and Quantification

cfDNA was extracted from plasma using the QiaAmp Circulating Nucleic Acid Kit (Qiagen) according to the manufacturer's instructions. cfDNA concentration was measured with a Qubit 4.0 Fluorometer (Thermo Fisher Scientific) using the dsDNA High Sensitivity Assay Kit (Thermo Fischer Scientific). cfDNA fragment size was determined using an Agilent 2100 Bioanalyzer with the High Sensitivity DNA Kit (Agilent Technologies). A median of 32ng was inputted into sequencing library preparation based on the percentage of cfDNA in the 70-450bp region of the bioanalyzer electropherogram. The QIAamp DNA Micro Kit (Qiagen) was employed to extract genomic DNA from 100ul of PDWB. Genomic DNA from PDWB was fragmented prior to library preparation using a LE220 focused ultrasonicator (Covaris).

Development of EnhanceAR-Seq Gene Panel

To develop a next-generation sequencing (NGS) assay for metastatic prostate cancer cfDNA analysis, we designed a hybrid-capture gene panel to target the complete *AR* gene body (including introns), 30kb of the *AR* enhancer, and exons of 84 other genes that have been shown to harbor genomic alterations in mCRPC[4,5]. To gain finer detail for copy number analysis in the full *AR*/enhancer locus, we evenly placed 500bp targeted regions

(1kb apart) between 500kb upstream of the *AR* enhancer and 500kb downstream of the *AR* gene body. Our panel also included the *TMPRSS2-ERG* gene fusion hotspot intronic region (13kb) in the *TMPRSS2* gene to detect a subset of *TMPRSS2-ERG* gene fusions. Additionally, 12 genes least frequently affected by copy number alteration in mCRPC (surveyed in prior WGS data[5]) were included in the panel as controls for copy number analysis, and three genes included to assess clonal hematopoiesis[6-8]. NimbleDesign (Roche) was used to convert our desired gene panel into a SeqCap EZ Prime Choice probe set (Roche).

DNA Processing and Analysis

We performed cfDNA and PDWB DNA library preparation using the Cancer Personalized Profiling by deep Sequencing (CAPP-Seq) workflow[2] with duplex barcoded adapters[3], then performed NGS on an Illumina HiSeq4000 with 2x150bp paired-end reads, with 12 samples sequenced per lane, dedicating ~60 million reads per sample. We then applied a custom bioinformatics pipeline detailed in the sections below.

Cell-free DNA SNV and Indel Analysis

Cell-free DNA (cfDNA) sequencing results were analyzed for single nucleotide variants (SNVs) and insertions/deletions (indels) using the CAPP-Seq bioinformatic pipeline[2,3,9]. Briefly, cfDNA sequencing reads were de-multiplexed using sample-level index barcodes, mapped to the human reference genome, filtered for properly paired reads, filtered for bases with Phred quality score ≥30, then de-duplicated using unique molecular identifiers. Background-polishing using 12 healthy donor plasma samples was performed to reduce stereotypical base substitution errors as previously described using the integrated digital error suppression (iDES) method[3]. Variant-calling using the CAPP-Seq pipeline was then performed to call SNVs and indels from patient plasma using matched plasma-depleted whole blood (PDWB) as the background reference, filtered further to remove potential SNPs with variant allele fraction (vAF) >45%, loci with de-duplicated depth <100, and mutations in the canonical clonal hematopoiesis genes *ASXL1*, *DNMT3A* and *TET2*[6-8]. Nonsynonymous SNVs and indels ≥2 base pairs in plasma, not present in matched PDWB, not present in the Genome Aggregation Database (gnomAD)[10] at a >0.0001 frequency, and indexed in the Catalogue of Somatic Mutations in Cancer (COSMIC)[11] were reported in the final dataset shown in Figure 2 and Appendix Table A10. Mutations in *AR* that met these criteria were considered positive by EnhanceAR-Seq. An additional SNV analysis using the filters described above but not requiring COSMIC indexing was performed to measure overall ctDNA SNV burden (number of SNVs detected per patient) and levels (based on mean vAF and cfDNA concentration), shown in Appendix Figure A5 and Appendix Table A16.

Cell-free DNA Copy Number Analysis

Cell-free DNA sequencing results were de-multiplexed using sample-level index barcodes, mapped to the human reference genome, then de-duplicated using Picard[12] based on identical start/end coordinates. Copy number analysis was performed based on a read depth approach. First, the genome was binned (larger bins for non-targeted regions and smaller bins for targeted regions) and read depth ratios for bins between plasma cfDNA and matched PDWB control samples were calculated and corrected for biases in GC content, sequence repeats, and target density using CNVkit[13].

Subsequently, read depth ratios were centralized by subtracting the mean log2 ratios of all bins across chromosomes and normalized using read depth ratios from bins overlapping with copy number control genes. Copy number segmentation was performed using DNACopy[14]. To obtain the background read depth ratio for individual genes, we performed the same analysis on 24 pairs of plasma and matched PDWB control cfDNA samples from male healthy donors. Finally, a gain (or loss) event in patient plasma was called when the calculated log2 ratio was four standard deviations above (or below) the median log2 ratio of that locus in healthy plasma. Genes whose log2 ratios showed high variability or deviation from zero in healthy plasma samples (median>0.2 or standard deviation>0.2) were excluded from copy number analysis.

Cell-free DNA Structural Variation Analysis

Our targeted panel was designed to capture structural variation (SV) breakpoints targeting full-length *AR* (including intronic regions) and the *TMPRSS2-ERG* fusion hotspot in an intron of *TMPRSS2*. SVs including tandem duplications were called using Lumpy[15] and Manta[16] using plasma samples with matched PDWB control samples. Subsequently, SVs with breakpoints overlapping the blacklist and low complexity regions[17] or those with both breakpoints falling in non-targeted regions were removed. Additional filtering was applied to retain only SVs with at least 2 supporting discordant read pairs or split reads and with high confidence regarding breakpoint positions (based on the width of the confidence interval provided by Manta or Lumpy being <5 bases), and filtering out SVs with abnormally high read support (>150 discordant read pairs or split reads) in patient plasma cfDNA.

Tissue Molecular Analysis

For some cases, at the discretion of the treating oncologist, matched formalin-fixed paraffin embedded (FFPE) tumor tissue of a metastatic site was available for molecular analysis. Tissue was submitted to Tempus Laboratories where DNA was isolated and targeted NGS performed with ~500x coverage using one of two panels — Tempus xO[18] (1,714 genes) or Tempus xT[19] (596 genes). Both panels included the *AR* coding region.

Clinical Outcomes and Statistical Analysis

The primary clinical endpoint, primary or secondary resistance to AR-directed therapy, was scored by a board-certified academic medical oncologist specializing in genitourinary cancers. Primary resistance was defined as PSA progression, change of therapy or death within 4 months of treatment initiation, or radiographic progression within 6 months. Secondary resistance was defined as PSA progression, change of therapy, radiographic progression or death outside of this timeframe. PSA progression was defined as an increase of ≥25% above nadir and ≥2 ng per milliliter, with confirmation ≥3 weeks later (PCWG3[20]). Secondary endpoints for our study were progression-free survival (PFS) defined as the time to PSA progression by PCWG3[20] criteria or death, or last known date of PSA measurement in non-progressors, and overall survival (OS) defined as time to death or to last follow up for alive patients. PFS and OS were calculated from time of study enrollment.

We performed survival and statistical analyses using R version 3 (http://www.r-project.org) and Prism 8 (Graphpad Software). Fisher's exact test was used to assess the

significance level of associations between assay results and resistance to AR-directed therapy. For PFS and OS Kaplan-Meier analyses, the log-rank test was used to estimate P values and the Mantel-Haenszel method used to estimate hazard ratios. Multivariate Cox proportional hazards models were fitted with incorporation of important baseline covariates including PSA, ctDNA levels, number of lines of prior therapy, prior abiraterone vs. enzalutamide treatment, metastatic disease burden, and time since diagnosis in order to further assess the independent impact of *AR*/enhancer alterations detected in cfDNA. Proportional hazards assumptions were confirmed for these analyses by evaluating the Schoenfeld residuals.

Supplementary References

1. Vandekerkhove G, Struss WJ, Annala M, et al: Circulating Tumor DNA Abundance and Potential Utility in De Novo Metastatic Prostate Cancer. Eur Urol 75:667-675, 2019
2. Newman AM, Bratman SV, To J, et al: An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 20:548-54, 2014
3. Newman AM, Lovejoy AF, Klass DM, et al: Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol 34:547-555, 2016
4. Annala M, Vandekerkhove G, Khalaf D, et al: Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. Cancer Discov 8:444-457, 2018
5. Quigley DA, Dang HX, Zhao SG, et al: Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell 174:758-769 e9, 2018
6. Genovese G, Kahler AK, Handsaker RE, et al: Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 371:2477-87, 2014
7. Hu Y, Ulrich BC, Supplee J, et al: False-Positive Plasma Genotyping Due to Clonal Hematopoiesis. Clin Cancer Res 24:4437-4443, 2018
8. Jaiswal S, Fontanillas P, Flannick J, et al: Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 371:2488-98, 2014
9. Chaudhuri A, Chabon JJ, Lovejoy AF, et al: Analysis of circulating tumor DNA in localized lung cancer for detection of molecular residual disease and personalization of adjuvant strategies. Journal of Clinical Oncology 35, 2017
10. Karczewski KJ, Francioli LC, Tiao G: Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv, 2019
11. Tate JG, Bamford S, Jubb HC, et al: COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 47:D941-D947, 2019
12. https://github.com/broadinstitute/picard
13. 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
14. Seshan VE OA: DNAcopy: DNA copy number data analysis. R package version 1.60.0, 2019
15. Layer RM, Chiang C, Quinlan AR, et al: LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15:R84, 2014

16. Chen X, Schulz-Trieglaff O, Shaw R, et al: Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32:1220-2, 2016
17. Li H: Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30:2843-51, 2014
18. Beaubier N, Tell R, Huether R, et al: Clinical validation of the Tempus xO assay. Oncotarget 9:25826-25832, 2018
19. Beaubier N, Tell R, Lau D, et al: Clinical validation of the tempus xT next-generation targeted oncology sequencing assay. Oncotarget 10:2384-2396, 2019
20. Scher HI, Morris MJ, Stadler WM, et al: Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. J Clin Oncol 34:1402-18, 2016