Supplemental Methods

### Study cohort inclusion criteria (extended)

Patients were recruited from one of 6 comprehensive BC Cancer centres and the Vancouver Prostate Centre in British Columbia. We required consenting patients to have histological diagnosis of prostate cancer (any histologic subtype) and evidence of metastatic disease on conventional imaging at time of first cfDNA blood draw. Castration-sensitive or -resistant disease were both allowed (castration-resistance defined according to Prostate Cancer Working Group 3 criteria). Blood samples were scheduled to be collected at baseline (generally prior to next line of therapy), on treatment at ~12 weeks following therapy initiation, and at clinical progression. No other clinicogenomic factors were considered in evaluating metacohort eligibility (including treatment history, Easter Cooperative Oncology Group performance status, or known germline/somatic alteration status).

Mutation pathogenicity was assessed on the basis of predicted protein-level consequence. Truncating variants—constituting frameshift insertions/deletions, splice site mutations (mutations within ±2bp from an exon/intron junction), and nonsense (stopgain) mutations—were presumed deleterious. Germline missense mutations classified as ‘pathogenic’ or ‘likely pathogenic’ in the ClinVar database were also considered deleterious[1](https://paperpile.com/c/SisDfj/AJeLv). Because of the challenge in assessing potential pathogenicity of previously unreported somatic missense variants, we excluded this mutation class from our selection criteria. An exception was made for *CDK12*, where missense variants disrupting the kinase domain (amino acids 728-1020; RefSeq ID: [NM\_016507](https://www.ncbi.nlm.nih.gov/nuccore/NM_016507)) are linked to protein inactivation[2,3](https://paperpile.com/c/SisDfj/juQgf%2BmDp6g). We additionally required somatic *CDK12* missense variants affecting the kinase domain to be predicted as at least ‘probably damaging’ by Polyphen2[4](https://paperpile.com/c/SisDfj/ieufI). One truncating variant in *CDK12* (p.W1459fs) was excluded since it was located 3’ of the kinase domain and presumed to be non-pathogenic.

### Blood collection, processing, and DNA isolation

Whole blood was collected in 2×9mL Streck Cell-Free DNA BCT tubes and maintained at room temperature prior to and during processing. Samples were centrifuged at 1600 rcf for 15 minutes; buffy coat was subsequently aliquoted and plasma was transferred to a new tube and centrifuged for an additional 10 minutes at 3200 rcf. Prior to DNA extraction, plasma aliquots were stored at -80°C. Cell-free DNA (cfDNA) was extracted from up to 6 mL of plasma with the QIAGEN Circulating Nucleic Acids kit. We introduced a one hour lysis incubation at 60°C and cfDNA was eluted in 60 μL water. Following extraction, cfDNA was quantified with the Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit, or the Quantus Fluorometer and QuantiFluor ONE dsDNA system. Buffy coat fraction aliquots from centrifuged whole blood samples (buffy coat and plasma were obtained simultaneously) were stored at -80°C prior to DNA extraction. Leukocyte (i.e. germline) DNA (gDNA) was extracted from the buffy coat fraction using the QIAGEN DNeasy Blood and Tissue kit or Promega Maxwell RSC Blood DNA kit and Maxwell RSC system, as per the manufacturer's instructions. Following extraction, gDNA was quantified with a NanoDrop spectrophotometer.

### CfDNA library preparation and sequencing

Libraries for next-generation sequencing were prepared from 10-25ng of plasma cfDNA. gDNA was sheared (modal fragment length ~180 bp) with a Covaris focused-ultrasonicator or via enzymatic digestion prior to A-tailing, end repair, Illumina-compatible adapter ligation and PCR amplification with the KAPA Biosystems KAPA Hyper Prep or Hyper Plus Kits. Library quantification was performed with a NanoDrop spectrophotometer; each library was run on an ethidium bromide gel to confirm success.

Purified sample libraries were multiplexed to obtain single pools with a combined mass of 1-2 μg. Library pools for initial sequencing were hybridized to a custom targeted panel capturing the exonic regions of 73 genes with known relevance to prostate cancer (Roche NimbleGen SeqCap EZ Choice). A subset of sample libraries (determined to have high tumor content from initial targeted sequencing) were sequenced on a second custom panel covering the full gene lengths (including flanking sequences) of key tumor suppressors (**Supplementary Table 2**). SeqCap EZ HyperCap protocol was used for hybridization and subsequent wash, recovery, and amplification of the captured regions. Agencourt AMPure beads enabled purification of library pools prior to quantitation with the Qubit 2.0 or Quantus Fluorometer. Pools were then diluted and sequenced on an Illumina MiSeq (V3 600 cycle kit) or HiSeq 1500/2500 (V4 250 cycle kit) machine.

### Tissue analysis

Patient records were reviewed for the availability of primary archival tumor tissue derived from diagnostic biopsy, prostatectomy, or transurethral resection of the prostate. Formalin-fixed paraffin-embedded tissue was retrieved and reviewed by a pathologist; regions with the highest Gleason Grade Group were selected for sampling and DNA extraction. Tumor tissue DNA extraction and targeted sequencing were performed as previously described[5](https://paperpile.com/c/SisDfj/XentL). One archival tissue sample (belonging to patient P19) failed post-sequencing quality-control and was excluded from analysis.

### Sequence alignment

Paired-end reads were mapped to the hg38 human reference genome using Bowtie-2.3.0[6](https://paperpile.com/c/SisDfj/51ShN). For cfDNA or samples that contained unique molecular identifiers (UMI), duplicate reads were identified if they aligned identically and harbored at most one degenerate base within their UMIs. Samblaster-0.1.24 was used to mark duplicate reads in leukocyte control samples and in cfDNA samples sequenced without UMIs[7](https://paperpile.com/c/SisDfj/FepJv). Cutadapt-1.11 in paired mode was used to trim 3'-end read adaptors. Low-quality read tails (smoothed baseq <30) were trimmed using an in-house algorithm. Bedtools-2.25.0 was used to quantify per-base read statistics in target regions following duplicate removal. Germline single-nucleotide polymorphisms were used to verify cfDNA-gDNA sample pairs.

### Somatic variant calling

Somatic cfDNA variants were called when supported by ≥10 unique sequencing reads and a variant allele fraction (VAF) of ≥1%. To eliminate stereotypical artifacts and germline single-nucleotide polymorphisms, we required the VAF of putative cfDNA mutations to be ≥20× higher than the average position-matched background error-rate calculated from a collection of leukocyte control samples, and ≥3× higher than the allele frequency in the paired gDNA. We required the paired leukocyte sample to have at least 20× raw sequencing depth at that position. For base substitutions, we required the average mapping quality of reads supporting the variant to be ≥10, and required the average distance of the mutant allele from the nearest read end to be ≥15 bases. Variants were manually inspected in the Integrative Genomics Viewer (IGV)[8](https://paperpile.com/c/SisDfj/Z8r5a), and protein-level consequences were predicted using ANNOVAR[9](https://paperpile.com/c/SisDfj/6KiIf). Each putative SNV and InDel was examined for imbalance in forward/reverse read representation, clustering toward the 5’ or 3’ read-ends, genomic sequence context (low complexity / highly repetitive tracts), as well as a high density of putative variants or artifacts in the local neighborhood surrounding the candidate mutation (potentially suggestive of high sequence homology and/or low mapping fidelity). We approximated DNA-damage repair variant clonal status by examining a mutation’s variant allele frequency (VAF) relative to the sample’s ctDNA fraction: clonal and subclonal variants were respectively defined by VAF/ctDNA ≥ 25% and VAF/ctDNA < 25%, as applied previously[10](https://paperpile.com/c/SisDfj/XoI1n) The identical variant-calling pipeline was used for tissue sequencing data, but with a minimum VAF of ≥5% required for calling somatic variants. Somatic structural rearrangements were identified using a split-read approach implemented via the software Breakfast ([github.com/annalam/breakfast](https://github.com/annalam/breakfast), revision e94e922), with the --merge-duplicates option. Rearrangements required a minimum of four unique supporting reads. The above thresholds used for variant calling were determined in previously published dilution series experiments and study of patient-matched tissue and cfDNA collections[11,12](https://paperpile.com/c/SisDfj/GwuQS%2B6MbcA). To calculate longitudinal mutation re-detection rates, we searched across all available patient-specific cfDNA collections for mutant reads corresponding to the set of independently identified variants (from all their samples). Variants with at least one mutant read (and independently called in ≥1 other serial sample) were categorized as being non-independently re-detected.

### Germline variant calling

Germline variants were identified in gDNA samples by searching for non-reference variants with a VAF of ≥15% and at least 8 supporting reads. We discarded germline variants with population allele frequency 0.5% or higher in the KAVIAR or ExAC databases[13,14](https://paperpile.com/c/SisDfj/yToaV%2BNZNHi). Protein-level consequences were predicted using ANNOVAR[9](https://paperpile.com/c/SisDfj/6KiIf). This minimum VAF of 15% was selected to accommodate the lower median sequencing depth of leukocyte samples relative to cfDNA, since lower depth can widen the observable range of VAFs corresponding to true heterozygous variants (**Supplementary Figure 1**). Since a lower VAF threshold is potentially liable to include variants arising from clonal hematopoiesis of indeterminate potential (CHIP), we compared the VAFs of putative germline variants across patient-matched gDNA and cfDNA samples. Evidence of concordantly low VAFs across both compartments (utilizing serial cfDNA collections where available) was used to eliminate plausible CHIP variants from the set of detected germline mutations. Note that this procedure was only performed for germline variants, as CHIP variants are eliminated from the set of somatic cfDNA mutations via removal of putative variants with gDNA read-support.

### Circulating tumor DNA purity estimation

ctDNA fraction was estimated using the highest observed VAFs of autosomal somatic mutations on non-amplified genes (coverage log-ratio < 0.2)[11,15](https://paperpile.com/c/SisDfj/GwuQS%2BrdL6g). Mutant allele frequencies are elevated in cases of concomitant loss-of-heterozygosity (LOH) (i.e. deletion of the non-mutated allele). In this circumstance, the highest observed VAF is related to ctDNA fraction as ctDNA fraction = 2 / (VAF-1 + 1). Deletion detection sensitivity is diminished in circumstances of low ctDNA purity, and we therefore conservatively assume in all cases that the observed mutation is accompanied by somatic LOH. Tissue tumor content was estimated using the identical technique.

### Copy number evaluation

Deduplicated sequencing reads were counted in all capture regions using bedtools-2.25.0[16](https://paperpile.com/c/SisDfj/Kf6lQ), and coverage logratios were calculated against a median reference derived from all leukocyte samples. A cfDNA sample from a healthy volunteer served as a control and reference for guanine-cytosine (GC) correction. GC fraction was calculated for all target regions, and Loess regression was applied to coverage logratios to correct for GC content bias. After correction, a median reference profile representing non-tumor derived cfDNA was constructed based on a pool of cfDNA samples with no detectable somatic mutations or amplifications. The final coverage logratio for each gene was calculated as the median coverage logratio of all intragenic target regions. Fixed log-ratio thresholds were used to assign categorical copy-number status: homozygous loss: (-inf, -1], heterozygous loss: (-1, -0.3], copy gain: [0.3, 0.7), amplification: [0.7, inf), and no evidence of copy number alteration: (-0.3, 0.3)[11,17](https://paperpile.com/c/SisDfj/GwuQS%2BcR8ze). We define a further category for mutant *BRCA2*, *ATM*, and *CDK12*—non-deletion loss-of-heterozygosity (LOH)—as encompassing both mutant-allele specific gains and copy-neutral LOH across the mutant allele (i.e. somatically-acquired uniparental disomy). Detection of non-deletion LOH required a log-ratio of >-0.3 and evidence of heterozygous single nucleotide polymorphism (SNP) imbalance.

Gene copy number log-ratio and minimum ctDNA thresholds were derived empirically. Specifically, we studied the distribution of coverage log-ratios and heterozygous SNP VAFs across metacohort cfDNA samples (constituting both ctDNA-positive and negative samples), as well as a separate cohort of healthy controls[11](https://paperpile.com/c/SisDfj/GwuQS). To reduce the false negative rate in cohort-level copy number alteration (CNA) frequency summaries or comparisons, we additionally require samples to have ≥20% ctDNA purity for their identified CNAs to be included. This minimum ctDNA heuristic was gleaned from the observation that—given a minimum required log-ratio of -0.3 for deletions—detection of homozygous deletions at 100% cancer-cell fraction is theoretically bounded by a lower-limit of 18.8% ctDNA purity (assuming a diploid tumor genome model). We relaxed this minimum ctDNA requirement to 5% for *AR* amplifications, since multi-copy gains across chromosome X have a high signal-to-noise ratio.

Absolute *AR* copy-number was estimated using ctDNA fraction and copy log-ratio[10](https://paperpile.com/c/SisDfj/XoI1n). Denoting X as the average number of *AR* copies in a population of tumor cells, the average number of AR gene copies released into circulation from both tumor and normal cells is (F × X + (1 - F) × 1), where F represents ctDNA fraction. Coverage log ratio (denoted as R) represents the log-transformed number of gene copies in an average cell releasing DNA into circulation divided by the expected number of copies in noncancerous cells at the same genomic coordinates. Therefore, coverage-log ratio can be linked to the number of *AR* copies as R = log2((F × X + (1 - F) × 1) / 1) and rearranged to solve for X such that X = 1 + (2R  - 1) / F. We restricted application of this formula to samples with both ctDNA fraction ≥5% and evidence of copy-number change via coverage log-ratio analysis alone.

To precisify classification of *BRCA2*, *ATM*, and *CDK12* copy number loss, we leveraged the VAFs of germline single-nucleotide polymorphism (SNPs) detected in cfDNA. Specifically, we tracked the median deviation from 50% heterozygosity in all SNPs detected within these genes. We first calculated the background noise in median SNP VAF (relative to 50% heterozygosity) by measuring its standard deviation 𝜎 in *BRCA2*, *ATM*, and *CDK12* in 195 metacohort samples without detectable ctDNA and sequenced using the same methodology. To identify non-deletion loss-of-heterozygosity (LOH) and corroborate LOH via monoallelic deletion in ctDNA-positive samples in our study cohort, we required a median SNP VAF deviation of ≥3𝜎 (±10% from 50% heterozygosity), in addition to the aforementioned log-ratio thresholds. Genes with a measured log-ratio between (-0.4, -1) and SNP VAF deviation <3𝜎 were additionally classified as harboring putative homozygous loss. Similarly, assignment of non-deletion LOH required a SNP VAF deviation ≥3𝜎 (non-deletion LOH was unevaluable in patients where intragenic germline SNPs were not identified). The former criteria of SNP deviation (for detection of deletion LOH and homozygous loss) was omitted for patients without detected germline SNPs.

### *BRCA2*-reversion variant analysis

We queried PubMed for previously reported prostate cancer *BRCA2* reversion mutations following treatment with PARP inhibitors or platinum-based chemotherapy. UCSC LiftOver was used to harmonize reference builds to GRCh38[18](https://paperpile.com/c/SisDfj/xman). NCBI reference sequences NC\_000013.11 and NP\_000050.3 were used to plot *BRCA2* exonic/intronic regions and protein domains, respectively. Variants were considered to restore the native BRCA2 protein reading frame if the net coding base-pair change (including the initial frameshifting variant if preserved) was a multiple of three. In circumstances of compound or complex reversion alterations, read-phaseable indels were concatenated to produce single stop/start positions. Variants that spanned genomic intron/exon boundaries were clipped at the terminal amino-acid positions in protein space. All reversion variants were plotted relative to a wild-type reference sequence.

### Control cohort

The control cohort for clinical outcomes and genomic alteration frequency comparisons is the subset of the metacohort (n=879) who were: 1) enrolled in a recent phase II prospective trial (NCT02125357), and 2) did not have evidence of deleterious somatic or germline mutations (see **study cohort inclusion criteria**) or homozygous loss in *BRCA2*, *ATM*, or *CDK12* as identified through targeted cfDNA sequencing. These 187 patients had comprehensive audited clinical data and outcomes available, and were profiled using the identical targeted sequencing and variant calling methodology. The full metacohort was not used as a comparator cohort due to the logistical and personnel challenge of 1) retrieving and reviewing full clinical annotation for all samples, and 2) curation of all non-DDR related mutations, copy number changes, and structural rearrangements to be consistent with the approach used in the study cohort and clinical trial. Because low ctDNA fraction precludes sensitive detection of some types of somatic variants, mutation frequency comparisons were performed using the 127 patients from this cohort with ≥2% ctDNA. Autosomal copy number comparisons used the 64 patients with ≥20% ctDNA. Comparisons of *AR* amplification frequency required a minimum 5% ctDNA purity (n=105 control patients). In mCRPC, the detection of ctDNA is independently associated with poor prognosis[11,19](https://paperpile.com/c/SisDfj/GwuQS%2BRvDK8). Therefore, we restricted comparison of clinical outcomes in the mCRPC setting to the 127 control patients with >2% ctDNA in at least one cfDNA sample.

### Statistics

Statistical tests and data analyses were conducted in Python 3.7 using pandas 0.25.0, numpy 1.16.4, scipy, statsmodels and Lifelines survival analysis 0.22.9[20–22](https://paperpile.com/c/SisDfj/LsGX5%2BLLc8j%2B1z7II). Survival functions were estimated via the Kaplan-Meier method and compared using the log-rank test. Hazard ratios were calculated via univariate Cox proportional hazards regression. Genomic alteration frequencies between the study and DDR-wildtype cohorts were compared using 2×2 contingency tables and assessed via the Fisher’s Exact Test. Mechanism-specific rates of biallelic loss between gene categories was also evaluated using the Fisher’s Exact Test (or the generalized Fisher-Freeman-Halton test for 2×n comparisons). Mann Whitney U statistics facilitated cross-categorical comparisons of ctDNA fraction and extent of *AR* gene amplification as continuous variables. Mutually exclusive genomic alterations were statistically quantified using the DISCOVER algorithm[23](https://paperpile.com/c/SisDfj/pwy07). All hypothesis tests were two-tailed and required a 5% significance threshold. All visualizations were generated using matplotlib[24](https://paperpile.com/c/SisDfj/fupVC).

### Data availability

Somatic mutation and copy number calls from initial targeted exon sequencing of the study cohort are provided in **Supplementary Tables 4-5**. De-identified sequencing data is deposited at the European Genome-phenome Archive (EGA) under accession number EGAS00001004800.

References for Supplemental Methods

1. [Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–5 (2014).](http://paperpile.com/b/SisDfj/AJeLv)

2. [Wu, Y.-M. *et al.* Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. *Cell* **173**, 1770–1782.e14 (2018).](http://paperpile.com/b/SisDfj/juQgf)

3. [Popova, T. *et al.* Ovarian Cancers Harboring Inactivating Mutations in CDK12 Display a Distinct Genomic Instability Pattern Characterized by Large Tandem Duplications. *Cancer Res.* **76**, 1882–1891 (2016).](http://paperpile.com/b/SisDfj/mDp6g)

4. [Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit7.20 (2013).](http://paperpile.com/b/SisDfj/ieufI)

5. [Chedgy, E. C. *et al.* Biallelic tumour suppressor loss and DNA repair defects in de novo small-cell prostate carcinoma. *J. Pathol.* **246**, 244–253 (2018).](http://paperpile.com/b/SisDfj/XentL)

6. [Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).](http://paperpile.com/b/SisDfj/51ShN)

7. [Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).](http://paperpile.com/b/SisDfj/FepJv)

8. [Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Research* vol. 77 e31–e34 (2017).](http://paperpile.com/b/SisDfj/Z8r5a)

9. [Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).](http://paperpile.com/b/SisDfj/6KiIf)

10. [Herberts, C. *et al.* Activating AKT1 and PIK3CA Mutations in Metastatic Castration-Resistant Prostate Cancer. *Eur. Urol.* (2020) doi:](http://paperpile.com/b/SisDfj/XoI1n)[10.1016/j.eururo.2020.04.058](http://dx.doi.org/10.1016/j.eururo.2020.04.058)[.](http://paperpile.com/b/SisDfj/XoI1n)

11. [Annala, M. *et al.* Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. *Cancer Discov.* **8**, 444–457 (2018).](http://paperpile.com/b/SisDfj/GwuQS)

12. [Wyatt, A. W. *et al.* Concordance of Circulating Tumor DNA and Matched Metastatic Tissue Biopsy in Prostate Cancer. *J. Natl. Cancer Inst.* **109**, (2017).](http://paperpile.com/b/SisDfj/6MbcA)

13. [Glusman, G., Caballero, J., Mauldin, D. E., Hood, L. & Roach, J. C. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* **27**, 3216–3217 (2011).](http://paperpile.com/b/SisDfj/yToaV)

14. [Karczewski, K. J. *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.* **45**, D840–D845 (2017).](http://paperpile.com/b/SisDfj/NZNHi)

15. [Vandekerkhove, G. *et al.* Circulating Tumor DNA Abundance and Potential Utility in De Novo Metastatic Prostate Cancer. *Eur. Urol.* **75**, 667–675 (2019).](http://paperpile.com/b/SisDfj/rdL6g)

16. [Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).](http://paperpile.com/b/SisDfj/Kf6lQ)

17. [Li, J. *et al.* CONTRA: copy number analysis for targeted resequencing. *Bioinformatics* **28**, 1307–1313 (2012).](http://paperpile.com/b/SisDfj/cR8ze)

18. [Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Briefings in Bioinformatics* vol. 14 144–161 (2013).](http://paperpile.com/b/SisDfj/xman)

19. [Romanel, A. *et al.* Plasma AR and abiraterone-resistant prostate cancer. *Sci. Transl. Med.* **7**, 312re10 (2015).](http://paperpile.com/b/SisDfj/RvDK8)

20. [Davidson-Pilon, C. lifelines: survival analysis in Python. *Journal of Open Source Software* vol. 4 1317 (2019).](http://paperpile.com/b/SisDfj/LsGX5)

21. [Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).](http://paperpile.com/b/SisDfj/LLc8j)

22. [Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference* (2010) doi:](http://paperpile.com/b/SisDfj/1z7II)[10.25080/majora-92bf1922-011](http://dx.doi.org/10.25080/majora-92bf1922-011)[.](http://paperpile.com/b/SisDfj/1z7II)

23. [Canisius, S., Martens, J. W. M. & Wessels, L. F. A. A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence. *Genome Biol.* **17**, 261 (2016).](http://paperpile.com/b/SisDfj/pwy07)

24. [Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* vol. 9 90–95 (2007).](http://paperpile.com/b/SisDfj/fupVC)