Supplementary Methods

## Abbreviations for table headings of Supplementary results.

TTTC : Time to Treatment Change

Resp or R : Responders, samples upper quartile of TTTC

NonResp or NR : Non-Responders, samples in lower quartile of TTTC

P: P-value

V1: Visit 1 biopsy (pre-treatment)

V2: Visit 2 biopsy (12 weeks post-treatment)

vs: Versus. Typically analyses will be Non-Responders versus Responders. Some are V2 versus V1.

wt: Wild-type, used to indicate status of variant in germline

Mut: Somatic Mutation

GS: Gene set analysis

RNA: RNA-seq results

CNV: Somatic Copy Number results

Somatic: Somatic Mutation results

OR: Odds Ratio

RR: Relative Risk

OS: Overall Survival

logFC: log2 fold change

logCPM: log2 of Counts per Million (CPM)

LR: Likelyhood ratio

FDR: False Discovery rate

FWER: Family Wise Error Rate

DFE: Differentially expressed

## Sample Data Processing.

## Sample Quality Control

### DNA

The tumor DNA and blood DNA from the same patients were sequenced via Whole Exome Sequencing (WES) and processed as previously described [1] to produce BAM file alignment, germline VCF, somatic VCF, annotated and filtered somatic VCF, and count summary for CNV calling following the procedure in Figure S1 (right pane). In brief, Germline, Visit1, and Visit2 Bam files were realigned together to facilitate cross-visit somatic mutation calling (V1 vs germline and V2 vs Germline) and increase sensitivity and repeatability across time points. CNV were called against a germline reference panel to lower the noise in log2 ratios.

Whole exome sequencing (WES) in the post-treatment samples yielded an average of 320 million 101mer, paired-end reads for each sample **(Supplementary Table S2).** The DNA QC standard was set based on our ability to compute the purity (also known as cellularity) and ploidy of tumor biopsies **(Supplementary Table S3)**. The median cellularity of the 158 DNA exomes included in the study was 26% before QC for both visits. After QC, 99 samples with median purity of 40% were retained for copy number variation (CNA) analysis, and 128 samples with median purity of 35% for both visits were retained for somatic mutation analysis. The purity was lower at V2 **(Supplementary Figures S3)** but was not associated with responder status (Kolmogorov-Smirnov test *p*=1).

For CNV calling, the outputs of the workflow were probe level log2 ratios that were GC-corrected and normalized using a collection of reference blood normal samples by PatternCNV[2]. The capture kit used had extra capture baits tiling the AR gene, but one sample from the early part of the study did not have enough DNA to be resequenced with the AR kit (**Supplementary Table S**2). Thus, two different capture kits were used. and two different germline set of samples were used. Using a reference versus paired reduced the CNV noise for the log2 ratio by almost a factor of 2.

To guard against a reference biased by region of common germline CNV in the data, the reference building algorithm detects regions with multiple (generally two) dominant peaks in the reference normalized count distribution and picked the most frequent peak. For the UGT2B17 gene, a gene involved in the steroid metabolism, the deletion minor allelic frequency(maf) of the CNV was high (38% in our germline reference panel). As a result the frequency of diploid carriers versus heterozygous deletion was lower in our diploid reference (30 diploid, 31 heterozygous deletions, and 8 homozygous deletions), comfounding our diploid finding algorithm. We therefore had to adjust the final log2 ratio for that gene as the reference was picked to be the heterozygous site. (See **Supplementary Figure S4** to see the adjusted gene-level values). Note that the TTTC is not associated with the germline genotype of UGT2B17 (p=0.309). Values for tumor UGT2B17, tumor UGT2B15, and germline UGT2B17 for samples usables in CNV analysis are in **Supplementary Table S3.**

### CNV Normalization and Purity-ploidy Estimation.

Because the biopsies were a mix of tumor and normal DNA, and because the tumor deletions and gains are not balanced, the raw normalized log2 ratios for tumors are not centered against either the diploid or the dominant ploidy of the tumor and the following procedure was used to further recenter the log2 ratios around the dominant ploidy as well as rescale the log2 values so that all **log2 ratios are equivalent to that of pure tumors** for all samples.

To identify the ploidy and purity of each biopsy, four plots were generated and examined to identify the log2 levels corresponding to the dominant ploidy level (2,4,6) and deletions and gains relative to that. **Supplementary Figure S2A-S2H** shows the process for 1 sample. Plots A-D are from the raw gene-level log 2 rations and panels E-H are similar plots after rescaling log2 signals to that of a pure tumor. Panel A shows the gene-level segmented data genome-wide. Next, the probe level signals were median aggregated at the gene level to create a histogram (panel B) of gene level signals (second plot per sample), which was then smoothed (Panel C). The gene level signals were segmented using DNAcopy [3] to generate the Panel C, which plots all the segments levels sorted by log2 ratio, with horizontal length proportional to the genome size. Plateaus in this plot correspond to discrete Copy number states. A fifth plot was also generated but not used: the B-allele Frequency (BAF) (ratio of reads carrying the alternate allele to the total number of reads at each site identified to be a germline variant in the normal blood DNA sample) versus genome coordinates. The BAF plots were too noisy to help detect ploidy below a purity of about ~ 60%. The four plots were examined to determine ploidy levels of gains and levels of deletions. These gains/deletions were then used to estimate the purity. If the model is consistent, the purity of the gains and deletions is similar and predicts the levels in the fourth plot and location of peaks in first plot. Purity from CNV was lower at Visit 2 than Visit 1, but not associated with TTTC or Responder/Non-Responder status (**Supplementary Figure S3A** and **Supplementary Figure S6**).

The log2 ratios of a sample with N copies of a chromosome, with “x” percent of tumor cells, “1-x” percent of normal cells, and ploidy P is described by following equation (note many normalization factors cancel out and are not included)

|  |  |
| --- | --- |
| $$log2 ratio=log2(\frac{^{Tumor Counts per region}/\_{Total Tumor Counts}}{^{Germline Counts per region}/\_{Total Germline counts}})=\frac{\left[N\*x+2\*(1-x)\right]}{P+x+2\*(1-x)}$$ | (1) |

Solving that equation, enables us to use the CNV-based purity and ploidy, to convert the log2 values for each sample to what is expected for pure tumor samples. Using those equations yield plots E-H that are similar to A-D, except that they use the rescaled and recentered data. These transformed log2 ratios of the copy number were then converted to absolute copy number of each tumor using the defined ploidy. Gains and deletions were then called relative to the baseline copy; gains were defined as >0.732 extra copies and deletions as 0.732 fewer copies than the baseline.

Note that we tried using the program ABSOLUTE [4] for automated purity/ploidy estimation, and it was not able to give us answers for the majority of low purity tumors and often predicted the ploidy incorrectly, which is why we used DNACopy + Equation (1) and manual review of results.

The final values used for analysis were the rescaled/segmented values. Because the high polymorphism rate of the UGT2B17 gene affected the segmentation process with the neighboring gene (UGT2B15), the copy number calls in those regions were manually reviewed and either made to match the adjoining regions values or kept at the computed value (the review took both the V1 and V2 data into account).

**Supplementary Figure S5** shows the median absolute deviation (MAD) computed versus purity. On average most samples have very tight CNV log2 ratios around the segmented values, with a mode of 0.06.

### DNA CNV/Purity Quality Control

A secondary purity estimate was computed using the relative ratio of reads carrying a somatic mutation allele to total reads, the mutant allele frequency (MAF). Best practices normally omit somatic mutations located in deletion or gain regions. However, because about half of the tumors were tetraploid, we could not omit mutations in gain regions for tetraploid tumors. Furthermore, even in a tetraploid tumor, mutations in the four copy state could not be assumed to carry somatic mutations reflecting the tumor purity if the somatic mutations occurred after the genome duplication event. Because of our lack of knowledge on the timing of mutations relative to CNV events, we chose to simply compute the purity as 2\* median(MAF), and only used this as a secondary purity estimate [4] as a quality control check of the CNV-based purity. The purity computed this way was also lower at V2, but not associated with TTTC or responder/Non-responder status (**Supplementary Figure S3B**).

The frequency of recurrent Amplification and Deletions in the TCGA for prostate Cancer was downloaded **from Supplementary Table S4** of [5] an included as a column in **Supplementary Tables S13** (Visit 1 CNV) and **Supplementary Table S14** (Visit 2 CNV). They counted a sample as having a gain if the segmented log2 ratio downloaded for the TCGA cloud was > 0.25 and a deletion if it was lower than < -0.25. The TCGA data is not centered on the dominant ploidy, but rather on the state with an average number of deletions/gains. Those copy number are not rescaled to take into account purity and ploidy. For a pure diploid tumor a +1 gain is a log2 ratio of 0.57 and -1 del is a log2 ratio of -1. Lower purity will reduce those signals, and miscalling the ploidy, will divide the expected levels further, much smaller than a 1 copy gain (log2=0.57 or 1 copy deletion, log2=-1). For example, using equation 1, a 50% tetraploid tumor would have a log2 ratio of 0.16 for a gain and -0.19 for a deletion if treated as diploid.

Our gene-level frequency was highly correlated at V1. V1 Gain (Adjusted R^2 =0.7, but our frequency is 2 times more common. V1 Deletions were also fairly correlated (adjusted R^2 0.46), but our intercept start at 0.45, so our deletion frequency are much more common), the V2 data shows a similar trend. We thus expect to be calling more gains (our better sensitivity for low purity/tetraploid tumors) . We may be overcalling deletions by calling events relative to the dominant ploidy depending on the relative timing of genome doubling events relative to copy loss. This overcalling should not affect our association analyses.

### DNA Somatic Mutation Calls

Somatic mutations were identified using 3 somatic callers as previously described [1].

Pathogenicity of selected variants were estimated by two primary bioinformatics tools, the cancer-specific OncoKB database [6] and the Variant Effect Scoring Tool (VEST)[7, 8] The OncoKB database was accessed via the OncoKB-annotator program. The VEST statistics and other annotations were obtained via the CRAVAT server[9] .

### PTEN Deletions, AR Gain/Deletion Calls:

Our method for calling CNV is less sentive to single-gene focal deletions/amplifications, so we manually reviewed the copy number in regions for AR, PTEN, UGT2B17, and UGRT2B15 by reviewing the log2 ratio data and Somatic Variant Allele frequency for these regions for each samples. Values for PTEN are shown in **Supplementary Table S3**.

### TMPRSS2-ERG fusions:

Reads with one end mapping in TMPRSS2 and the other end mapping in the ERG gene were counted in the RNA-Seq (see fusion detection below). We also manually reviewed the DNA CNV and found a number of samples with deletion of the characteristic region of this fusion. In all cases where the deletion was detected, the fusion reads were also detected. Some of the fusions are expected to be balanced and were not detectable by the copy number data. Values are shown in **Supplementary Table S**3. TMPRSS2 erg fusions were not associated with TTTC.

### RNA-seq Preprocessing and Quality Control Metrics.

RNA sequencing was performed as previously described [1] and normalized using conditional quantile normalization (cqn) [10] on the combined 161 Visit 1 and Visit 2 bone and tissue expression samples. We then computed the stress and dfArray metrics[11]. **Supplementary Figure S7** shows the stress and dfArray metrics as well as standard exclusion cutoff lines; **Supplementary Figure S8** shows a zoomed in view after removing the outliers and shows no correlation between the quality metrics and Visit 1 or Visit2 or biopsy source (Bone or Tissue).

**Supplementary Figure S4** shows a heatmap of the gene set scores for a number of gene sets specifically associated with androgen response and prostate cancer. Below the heatmap are a number of QC metrics (described below). We see that the poor performing samples (by various metrics) were all clustered on the left. The leftmost cluster included a set of out-group RNA-seq samples from the blood. Tumor samples that clustered with the blood samples tended to have the LOWPURITY flag activated and likely contained no tumor. We chose to exclude from analysis all samples that clustered with blood, as this cluster contained most of the samples that failed the dfArray and stress QC. We retained the samples in the purple cluster, but labelled them as second tier. The following metrics are provided in **Supplementary Table S4**:

 RNA-seq Purity: The RNA-seq purity was computed as in the Supplementary Methods of [12]. In brief, this method makes use of the high accuracy purity obtainable from the DNA using the CNV. The DNA purity was then multiplied by the robust ratio of the somatic mutant allele fraction in RNA to that of the allele fraction in DNA. The robust averaging across multiple variants mitigates the RNA allele counts bias stemming from allele-specific expression bias.

 ESTIMATE purity: We used the v2.0 ESTIMATE package [13], downloaded from ‘http://r-forge.r-project.org/R/?group\_id=2237’, to estimate the stromal fraction for the RNA-seq. Since this estimator is not calibrated for RNA-seq, the estimate can only be expected to be monotonically related to the purity. We found that ESTIMATE purity was correlated with CNV/Somatic Purity (not shown).

RIN: The RIN numbers [14] were provided to us by the Mayo Clinic Genome Sequencing Facility. The RIN is a number between 0 and 1, indicative of the level of integrity (non-degradation) of the samples. High integrity data is usually above RIN>8.

 TIN: The Transcript Integrity Number [15] is another measure of the degradation; it quantifies the uniformity of the reads across the span of the transcript and correlates with the RIN, but is more informative of the uniformity of transcripts at low RIN numbers.

 IQR : We used the InterQuartile Range (IQR) of the RNA-seq data to identify outliers as follows:

* Obtained raw counts data using RSEM
* Retained those genes that had non zero counts in at least 1/3rd of total samples (44 out of 131 samples)
* Normalized the raw counts as log 2 (counts per million)
* Estimated InterQuartile Range (IQR) for all samples, and the third quartile (Q3)
* From a boxplot of these IQRs, identified outliers that were Q3+1.5\*IQR in that boxplot;

QC threshold was set as above 10.61 or below 3.18

 ARreads: The number of reads mapping to the AR gene region.

pctMapped: Percent of mapped reads.

For RNA sequencing, we computed several indicators of RNA quality and tumor content **(Supplementary Table S4)** to identify samples with abnormal RNA QC profiles **(Supplementary Figures S7 and S8)**. We clustered the samples to identify those that had poor cellularity or poor quality due to tissue necrosis **(Supplementary Figure S9)**. Samples in the leftmost cluster of **Supplementary Figure S9** were enriched in annotations indicative of poor quality or absence of tumor, and these samples did not have alterations commonly found in CRPC tumors: PTEN deletions TMPRSS2-ERG fusions, sequencing reads supporting full-length AR or AR isoforms. Therefore, samples in the leftmost cluster were excluded from the analysis. After removing duplicates based on lower sequencing depth or lower purity, 131 samples from both visits were included in subsequent analyses.

We performed burden tests to evaluate whether any of the differences that we observed between responders and non-responders or pre-treatment and post-treatment samples might be due to differences in the number of somatic mutations per sample (burden) (either because of lower purity of post-therapy samples or because of the acquisition of a hypermutator phenotype). Using Fisher exact tests, we tested the difference in count of “non-synonymous or worse” (Frameshift, stop gain, stop loss, splice-site distrupting) somatic mutations per sample between pre-treatment and post-treatment samples (test1) or between responders and non-responders either pre (test2) or post-therapy (test3). All three tests were not significant (t-test *p*-values=0.16, 0.1, and 0.5 respectively), indicating that the differences between these groups were likely, not due to excess of mutation burden. In non-responders, two samples had an extremely high burden with possible mutations that may serve as explanations. Non-responders (at V2) had higher mutation burden (p=0.12).

We also tested whether purity was associated with TTTC at V2. This is not unlikely because our strategy of biopsying the same site should lead to more tumors – hence higher purity – in non-responders. The p-value of Purity vs TTTCV2 is 0.96, while the P-value of Purity vs log10(TTTCV2) is 0.18 (**Supplementary Figure S5**).

**Fusion Calls:**

The tophat aligner we used for RNASeq also made fusion calls. To account for the sensitivity trade-offs of fusion callers, we made a list of all unique fused gene pairs and genotyped all these fusions across all samples by aligned the assembled fusions across samples. No significant association was found and those calls are not reported.

**AR isoforms:**

AR isoforms were called by counting read junction signatures of various AR isoforms (**Supplementary Table S33**)

## Gene fusion

The count of reads supporting *TRMPRSS2-ERG* fusions is included in **Supplementary Table S4**. Due to sample degradation , our polyA targeting RNA-seq kit often to inadequate coverage to call the *TRMPRSS2-ERG* fusion that normally occur near the 5’ end of the ERG gene, so we were not able to find a difference between responders and non-responders. Other fusions were too rare to detect a difference between responders and non-responders.

# STATISTICAL METHODS:

 The outcome measure used for the analyses was the time to treatment change (TTTC). We also discretized the TTTC into quartiles and considered the analysis of patients in the lower quartile of TTTC (non-responders) and upper quartile of TTTC (responders). In the RNA-seq, the log2 ratio had the lower quartile of TTTC patients in the numerator (i.e., non-responders are the affected patients/cases in analyses).

Statistical analyses were conducted with R version 3.4 and Bioconductor version 3.5.

In our study, a large fraction of the paired samples did not have a matched sample passing quality control at the other visit (purity too low in specific biopsy), so, to compare V1 and V2 counts, we used a test that takes into account both paired and unpaired samples. The Derrick [16] Proportion test Z8 , which optionally merged the paired counts and unpaired counts, was shown to have the maximum power while controlling the false positive. We also report the results for tests that only use the paired samples: McNemar test (paired data) with continuity correction, and the conditional logistic regression (for paired data). All of the paired tests are designed for the asymptotic cases andmay be inflated when the counts are very low (especially when one count is 0 or 1), so we also used the Odds Ratio for paired data and Risk Ratio for filtering significant results. Odds ratio for paired data: a) only use paired data, and b) only use pairs in which the results at both visits were not the same.

The Fisher Exact test or logistic regression test were used when the data was not paired or when we collapsed the pairs, e.g. Patients with Gains from V1 to V2 versus No Gains.

The TTTC phenotype was also analyzed using Survival (cox models in R using the survival package).

## Gene sets:

Gene sets were selected from: a) all of the NCI-Nature Pathways gene sets starting with a numeric code; b) subset of MSigDB [17] from keyword searches: Androgen, prostate, CRPC, castrate, PD1, PD-1, PDL1, AR, Wnt, hippo, PTEN, cell cycle, steroid, hormone, AKT, ERG, TMPRSS2, Notch; c) genes between TMPRSS2 and ERG1, deleted during imbalanced fusions (as opposed to just moved another place in the genome); d) two gene sets computationally created by the Janssen team using a large collection of RNA-seq that discriminate prostate cancer from other types of samples; and e) gene sets curated from publications or other references identified by literature or web searches. References and descriptions of the gene sets are in Table S32.

Gene set analyses were performed with the GSVA package for the CNV and RNA-seq. The GSVA Signals (per sample, per gene set) were then used in regression models. V2 (Lower vs Upper TTTC = Non-responders vs Responders), V1 (Lower vs Upper TTTC = Non-responders vs Responders), and V2-V1 (Difference of the GSVA Signals). For CNV analysis, the GSVA score was derived from either the V2 log2 ratios (relative to the major ploidy) or the difference of the log2-ratios for the V2-V1 analysis. For RNA-seq, the cqn normalized data were used as input to GSVA.

### ANALYSES

### Somatic Mutations

After Quality Control, we had several samples with missing data at one visit, so, for paired analysis with both V1 and V2 data, we were using a test that takes into account both the paired counts and the unpaired samples either from V1 or V2, test Z8 [16]. We also computed (lower power) tests for paired data, either the McNemar test or the logistic regression. Finally, we also computed the odds ratio for the paired samples only, ignoring samples that had the same mutation status at both levels.

We also used counted “Gain of Mutation” or “Loss of Mutation”, and with the data combined this way, no adjustment was necessary for the paired samples.

1. V2 versus V1 analyses, done by default with test Z8 [16], which uses both paired and partially paired (e.g., one sample missing RNA at one visit due to poor quality) samples. In **Supplementary Table S8**:
	1. Gene Symbol
	2. pV1V2: Partially paired comparison using test Z8 [16], of number of samples with mutations in that gene at V1 versus V2. This test has high power, but can suffer from inflated type II error with low counts
	3. pPaired: McNemar test for Paired data, more conservative
	4. Relative Risk
	5. Counts of mutated samples
	6. Paired and unpaired counts of samples with wild type (wt) or mutation (mut).
2. V2 vs V1 analysis comparing non-responders (lower quartile of TTTC) and responders (upper quartile of TTTC) in gene sets. In **Supplementary Table S9:**
	1. Gene Symbol
	2. pGS.Gain.TTTC: Conditional Logistic regression of counts for paired V1 vs V2 for gain of mutations from V1 and V2 in each geneset versus TTTC.
	3. pGS.Loss.TTTC: Conditional Logistic regression of counts for paired V1 vs V2 for loss of mutations from V1 and V2 in each geneset versus TTTC.
	4. pCox.Gain: Cox model of counts of mutations gained from V1 and V2 vs TTTC.
	5. pCox.loss: Cox model of counts of mutations lost from V1 to V2 vs TTTC.
	6. NGain.NR: Number of samples that gained mutations in non-responders for genes in the gene set
	7. NGain.R: Number of samples that gained mutations in responders for genes in the gene set
	8. NLoss.NR: Number of samples that lost mutations in non-responders for genes in the gene set
	9. NLoss.R: Number of samples that lost mutations in responders for genes in the gene set
	10. Gain.genes.R: Genes with gain of mutations in responders for genes in the gene set
	11. Gain.genes.NR: Genes with gain of mutation in non-responders for genes in the gene set
	12. Loss.genes.R: Gene with loss of mutations in responders for genes in the gene set
	13. Loss.genes.NR: Gene with loss of mutations in non-responders for genes in the gene set
	14. Description of gene set
3. Gene set Test in (V1 vs V2 in non-responders only) **Supplementary Tables S10**
	1. Gene Set name
	2. Gene Set Test V1 vs V2, using test Z8 ; counting samples with at least 1 gene mutated in non-responders only.
	3. P-value of cox-model of GeneSet Counts vs TTTC
	4. Odds Ratio for paired data only (only using samples without same mutation at both visits)
	5. Relative Risk for Geneset.
	6. NV1.Sample: Number of V1 samples with a mutation in the geneset
	7. NV2.Sample: Number of V2 samples with a mutation in the geneset
	8. V1mutgenes: Genes mutated in at least 1 sample at V1, Number in parentheses is number of samples with a mutation in that gene.
	9. V2mutgenes: Genes mutated in at least 1 sample at V2, Number in parentheses is number of samples with a mutation in that gene.
4. Gene set test at V2 (non-responders vs responders) – **Supplementary Table S11**
	1. Gene set name
	2. pGS: Gene set test (Fisher exact test for Responders vs Non-Responders)
	3. pGS\_Surv: Gene set test with counts of mutations versus TTTC with all V2 samples using Cox Model
	4. Odds Ratio for the GeneSet for responders vs Non-responders
	5. Relative Risk of the GeneSet for responders vs Non-responders
	6. Number of non-responders(lower quartile of TTTC) with mutations in the geneset at V2
	7. Number of responders (upper quartile of TTTC) with mutations int the geneset at V2
	8. Absolute Difference in number of mutated samples between non-responders and responders in the geneset at V2.
	9. pV1vsV2 (from as **Supplementary Table S7**)
	10. pV1vsV2\_NR: test Z8 for counts of mutation in geneset between V1 and V2 samples, limited to non-responder samples
	11. pV1vsV2\_R: test Z8 for counts of mutation in geneset between V1 and V2 samples, limited to responder samples
	12. Significant genes in this gene set (and cox p-value and counts of non-responders/responders)
	13. Description of gene set
5. Gene Level Somatic vs Response at V2 – **Supplementary Table S12**
	1. Gene symbol
	2. P-value for count of mutated samples in non-responders vs responders (Fisher Exact Test)
	3. Cox model p-value for all V2 samples – wether the sample is mutated vs TTTC (relative to second visit) per gene
	4. Number of samples with (at least) 1 mutation in that genes
	5. q-value of Cox model
	6. Relative Risk for Non-Responders vs Responders
	7. Odds Ratio for Non-responders vs Responders
	8. Mutated samples in Non-Responders at V2
	9. Mutated Samples in Responders at V2
	10. Number of Non-responders at V2
	11. Number of Responders at V2

### CNV Analyses:

 For CNV Calls, Copy Number gains and deletions were called relative to the dominant ploidy of the tumor. Analyses used discretized values of the CNV (gains, deletions), as described in section “CNV Normalization and Purity-ploidy Estimation.” Gains and deletions were then analyzed separately and combined at the gene level. The reported Pgene P-value is the minimum of the gains P-value or deletions P-value. Gene set tests were performed using logistic regression of the GSVA signal.

1. V1 Lower TTTC vs Higher TTTC quartile (non-responders vs responders), Fisher Exact Test
	1. pTTTC: Lowest of the next two p-values(min (b.,c.))
	2. pTTTC gains: Gains vs No Gains Fisher Exact Test(**Supplementary Table S13**)
	3. pTTTC dels: Losses vs No Losses Fisher Exact Test (**Supplementary Table S13**)
2. V2 Lower TTTC vs Higher TTTC quartile (non-responders vs responders), Fisher Exact Test
	1. Gains vs No Gains (**Supplementary Table S14**)
	2. Losses vs No Losses (**Supplementary Table S14**)
	3. The gene p-value in **Supplementary Table S14 and S17** is the min of a. and b.
3. V2-V1 Lower TTTC vs Higher TTTC (non-responders vs responders), Fisher Exact Test
	1. Gain at V2 vs No Gain at V1 (**Supplementary Table S15**)
	2. Deletion at V2 vs No Deletion at V1 (**Supplementary Table S15**)
	3. The gene p-value in **Supplementary Table S15 and S21** is the min of a. and b.

Gene-Set tests were computed using the GSVA package to create a unique score per sample per gene set. Statistical analyses were performed using logistic regression vs TTTC. We also included the median CNV in non-responders (lower quartile of TTTC) and responders (upper quartile of TTTC) to facilitate interpretation of the logistic regression results (with samples from all quartiles).

### RNA-seq Analyses

Differential expression analysis was performed comparing expression for two groups of patients determined as cases and controls: the lower (25th percentile) and upper (75th percentile) quartile, respectively, of the time to treatment change (TTTC) distribution of 84 PROMOTE patients.  Two differential expression analyses were performed using edgeR [18] as described [19], on 21303 genes with sufficient expression in the normalized expression set.

1. Pre-treatment (V1) case versus control set (15 lower quartile (assigned to case) and 14 upper quartile (control)) – **Supplementary Table S23**
2. Post-treatment (V2) case versus control set (18 lower quartile (case) and 11 upper quartile (control)) – **Supplementary Table S22**
3. V2-V1: Defining the post- and pre-treatment samples as repeated measures on the same patients, we tested for differential expression of case and control groups with complete data across both visits, which was 8 V2 controls (upper quartile) from 12 V2 cases (lower quartile), as described in the edgeR user manual [edgeR, section 3.5,page 36]. – **Supplementary Table S24**

### RNA-seq Gene set Analyses

Gene set analyses were performed using the cqn normalized data (described in **RNA-seq Preprocessing and Quality Control Metrics** section). Significant gene sets were reported together with genes significant by the gene analyses (False Discovery Rate (FDR)<=0.05), with a “(-)“ for downregulated in non-responders and “(+)” for upregulated in non-responders.

1. V1: **Supplementary Table S26**
2. V2: **Supplementary Table S25**
3. V2 – V1: **Supplementary Table S27**

## Pathogenicity of Selected Somatic Variants of the PROMOTE study

**Supplementary Table S10** shows pathogenicity annotation of the most common Somatic Variants. The selected variants were queried by two primary bioinformatics tools, the cancer-specific OncoKB database [6] and the Variant Effect Scoring Tool (VEST) [7, 8]. The OncoKB database was accessed via the OncoKB-annotator program. The VEST statistics and other annotations were obtained via the CRAVAT server [9].

## SIGNATURES

Gene Signatures were computed using a z-score method relative to the distribution in the whole study (all quartiles) following the method of Liguo Wang [20]. Gene symbols in the signatures were updated to match the annotation version of our RNASeq processing using searches using genecards [21] . We used 3 AR activity scores, the first one from Hieronymous [22] is a 27 gene signature. Later publications by Kumar and Wang[20, 23] limited to 20 genes and defined a core set of genes.

#### AR\_activity\_hieronymus

**ABCC4**,ACSL3,ADAM7,AR,C1orf116,CD200,CENPN,**EAF2**,ELL2,**FKBP5**,GLRA2,GNMT,HERC3,KLK2,KLK3,MAF,MAN1A1,MAPRE2,MED28,MPHOSPH9,NKX3-1,NNMT,PIP4K2B,PMEPA1,PTGER4,SRP72,TBC1D9B,TMPRSS2,TNK1,ZBTB10

#### AR\_activity\_kumar\_liguo

**ABCC4**,ACSL3,ADAM7,C1orf116,CENPN,**EAF2**,ELL2,**FKBP5**,GNMT,HERC3,KLK2,KLK3,MAF,MED28,MPHOSPH9,NKX3-1,NNMT,PMEPA1,PTGER4,TMPRSS2,ZBTB10

#### AR\_activity\_core\_kumar\_liguo

AR,KLK2,KLK3,NKX3-1,PMEPA1,TMPRSS2

We used two sources for Signatures of NeuroEndocrine subtype features. The first set were taken from [24] , except that we excluded the TMPRSS2-ERG fusion, which is not well captured in our RNAseq data. We separated the signature in up and downregulated genes to support the z-score approach of scoring the signature.

#### NE\_UP\_KUMAR2019

SYP,CHGA,CHGB,AURKA,MYCN,EZH2,ENO2,CALCA,SCG2,SCG3,VIP,GRP,

NKX2-1,NKX22,NCAM1,FOXA2,WNT11,POU3F2,SRRM4,SOX2,SOX11,CEACAM5,

ASH1L,ASCL1,PEG10,CDKN2A,DLL3

#### NE\_DOWN\_KUMAR2019

AR,KLK3,TP53,RB1,FOXA1,PTEN,REST,CYLD,SPDEF,CCND1

key genes for NE activity were taken from Figure S2 of [23]

#### NE\_key\_kumar

CHGA,CHGB,NKX2-1,ENO2,ELAVL4,SCG3,PCSK1,SCN3A,CHRNB2

The other groups of signatures for Neuro Endocrine features was developed using 15 Prostate Cancer patients with neuro-endocrine pathology by Tsai [25]

#### NE\_UP\_TSAI

APLP1,ASCL1,CCDC88A,CDC25B,CRMP1,DLL3,DNMT1,ELAVL3,ELAVL4,ENO2,

FAM161A,FANCL,FGF9,IGFBPL1,INA,INSM1,KCNC1,MIAT,NKX2-1,NPPA,

NPTX1,PCSK1,PCSK2,PHF19,RNF183,RUNDC3A,SEZ6,SH3GL2,SNAP25,SOX2,SRRM4,

STMN1,TMEM145,TOX,TUBB2B,UNC13A

#### NE\_DOWN\_TSAI

AR,AIM1,ADRB2,SPDEF,STEAP1,STEAP2,C1orf116

DNA repair genes signatures were also taken from Kumar [23], though they are not strictly speaking a signature because inactivation/deletion of only a few genes are necessary to trigger the status of DNA repair defective. They required any two of the following to classify a tumor as DNA repair defective: deletion or mutation of ATM at either somatic or germline level, germline mutation in BRCA2, or “homozygous loss of any of the 15 FA-complex-associated genes FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, BRIP1, FANCL, FANCM, PALB2, SLX4”. Note that FANCI, FANCC, USP1 are regulated by AR

#### FA\_DNArepair\_kumar

BRCA1,FANCD2,FANCI,FANCA,BRIP1,BRCA2,FANCB,FANCG,FANCM,FANCE,

FANCC,FANCL,PALB2,SLX4,FANCF,CDK12

NMYC regulated genes were identified by ChIP-seq in [26]

#### NMYC\_regulated

PARP1,PARP2,BRCA1,RMI2,TOPBP1

A subset of DNA repair genes that have AR response elements occupied under conditions of DHT administration were identified by Polkinghorn [27] .

#### DNA\_repair32\_polkinghorn

POLE2,MAD2L1,FANCI,RFC3,POLA2,RAD54B,MCM7,RFC4,RAD18,RAD51C, CHEK1,POLA1,FANCC,TOPBP1,CCNH,MRE11A,MSH6,RAD21,XRCC4,PARP1,ATR, USP1,RFC1,HUS1,MSH2,XRCC5,LIG3,NBN,SHFM1,ALKBH1,TDP1,WRN

#### NHEJ\_AR\_polkinghorn

XRCC4,XRCC5

#### HR\_AR\_polkinghorn

MSH2,MSH6

#### BER\_AR\_polkinghorn

PARP1,LIG3

#### DNARepair\_down\_polkinghorn

MRE11A,FANCI,RAD18,MAD2L1,MCM7,TDP1,MSH6

#### DNARepair\_up\_polkinghorn

HUS1,RAD51C

Genes in the Biosynthesis pathway are taken from Figure 6 of [28]

#### androgen.biosynthesis

AKR1C3,HSD17B3,HSD3B1,HSD3B2,SRD5A1,DHRS9,RDH5,SRD5A2,RDH16,HSD17B10,HNF4A,UGT2B15,UGT2B17,AKR1C2

The genes that are preferably activated by ARV isoform variants and the genes indicative of Full Length AR activity were taken from [20]

#### ARV\_pref\_targets\_wang

ANKRD32,ANO6,ASPM,ATL2,CA8,CDK1,CEP128,CHAC2,CPS1,CROT,

DIAPH3,DSCAM,DTL,E2F7,ENDOG,GCNT1,GJA1,GMPR,GRIN3A,IL1R2,

INMT,INMT-FAM188B,INSC,INSIG1,ISL1,KBTBD8,KCNC4,KCNMB2,KRT19,

LPL,LRFN2,MAD2L1,MAF,MESP1,MESP2,MID1,NCAPG2,NDST4,NUF2,

NUP93,PARL,PHTF2,PREX2,PRIM2,RAB33A,RARB,RDH10,RECQL,RGS1,

RGS2,SGOL1,SLC31A2,SLCO2A1,SMC2,STOX1,TIFA,TMEM169,TMEM241,TMEM97,

TMPO,TRIM36,TTK,UCHL5,WNT16,ZBTB10,ZNF367

In this list, we retained three genes (ASPM, CDK1and DTL) that were removed from Wang’s analysis due to overlapping with the cell cycle progression list.

#### ARFL\_activity

ADAM9,SLC50A1,RAB4A,PHF8,PARVA,NRARP,NFIX,MYO1B,MMP20,LRIG3,

KLF5,KCNJ1,GLI3,G3BP2,FAM96B,ERN1,CUZD1,CLDN8,CDK2AP2,C15orf23,

TIGD6,GPR65,GALNTL4,FOXD4,C17orf81,DNM1L,ABCA1,CKB,CEACAM16,CAB39L,

ACPL2,ABCA5,AGAP9

The Cell Cycle Gene score is from [29] . It was originally published with a housekeeping gene normalization. However, when we tried the housekeeping gene normalization, we found that it correlated exceedingly well with the z-score signature (Pearson correlation coefficient of 0.98), so we used the z-score approach for consistency.

CCPgenes
FOXM1,CDC20,CDKN3,CDK1,KIF11,KIAA0101,NUSAP1,CENPF,ASPM,BUB1B,RRM2,DLGAP5,BIRC5,KIF20A,PLK1,TOP2A,TK1,PBK,ASF1B,SKA1,RAD54L,PTTG1,CDCA3,MCM10,PRC1,DTL,CEP55,RAD51,CENPM,CDCA8,ORC6

### Gene Sets:

The gene sets used for the analyses are included in **Supplementary Table S36**.

## CNV ANNOTATIONS:

The gene-level CNV results are annotated as follows:

Known: Genes known to be related to Castrate Resistant Prostate Cancer (CRPC) [30, 31], [32], and TMPRSS2-ERG interstitial genes. See **Supplementary Table S37** for complete list of known CNV.

AR\_targets : Genes identified to be regulated by AR, based on chromatin immunoprecipitation sequencing (ChIP-seq) experiments [5]. **See Supplementary Table S38** for complete list of genes that are AR targets.

AR regulated: Genes identified to be regulated by AR, according to review of multiple publications [33]. **Supplementary Table S39** shows the genes regulated by AR.

TCGA: Genes with recurrent gains/deletions in The Cancer Genome Atlas Prostate Cancer paper [5]. The regions from that paper are listed in **Supplementary Table S40** and the mapping between genes and those regions is in **Supplementary Table S41**

## Literature search for the most significantly differentially expressed genes in post AA/p samples between non-responder and responders.

AT V2, there were 93 genes that were upregulated and with FWER<=0.05 and 10 genes downregulated at that same stringency level. We reviewed half of the 93 genes and the 10 downregulated genes to identify pathways associated with non-response or long term response.

In this list, the gene symbol is preceded by the rank of the gene (from the V2 differential expression analysis).

Two lists are provided: Genes elevated in non-responders and genes elevated in responders. Genes are grouped together if they are related, so they are not necessarily in rank order.

1. CDC20 was reported to stabilize beta-catenin (WNT signaling) [34] and silencing it suppresses CRPC and sensitizes to docetaxel [35]

2. CKS2 is a target gene of the Polycomb-independent activation of EZH2 [36]

3. NEK2 stabilizes both EZH2 [37]and beta-catenin [38]. This gene is regulated by AR-V7, but not AR-FL[39].

21. ZWINT is a kinetochore protein, that is regulated by AR-V7, but not AR-FL[39]. Was found to be upregulated in PC (together with FEN1(#67)) [40]

32. DSN1 is a mitotic kinetochore complex that may interact with ZWINT [41]

4. FOXM1 has been reported to be a driver of one subtype of prostate cancer [42], to upregulate PSA expression and be androgen-independent[43]. FOXM1 is upregulated in cMYC driven prostate cancer[44], where shRNA depletion of FOXM1 reduced expression of cell cycle regulatory genes in cMYC cell lines [44]. Regulates KIF4A, which stabilizes AR and AR-V7 against degradation.[45]

17. KIF4A binds AR and AR-N7 and prevents degradation while AR binds the promoter of KIF4A[46]. In Hepatocellular carcinoma, KIF4A is regulated by FOXM1 [45].

6. UBE2T is a ubiquitin-conjugating enzyme E2T whose overexpression has been reported to promote prostate cancer cell proliferation [14] significantly.

7. CDK1 can phosphorylate AR at S81 to provide androgen-independent AR activation [47]

8. TROAP correlates with WNT3 and survivin (BIRC5) expression, and it's knockdown inhibited cell proliferation [48]

9. CENPA Together with CENPE (#46 , FWER=0.0015) and BIRC5 (#40) are cell cycle metaphases checkpoint genes they are upregulated in CRPC by the co-binding or AR and KDM1A(LSD1)[49, 50] to their promoters. The latter two are responsive to high levels of androgen-independent AR [49],[39], and since KDM1A is not differentially expressed in our data, it must be elevated because of AR-V isoforms.

46. CENPE regulates AURKB, PLK1, and E2F1 [49].

15. HJURP centromeric protein upregulated in CRPC (but down in Neuroendocrine) [51]. In the absence of LSD1, can be recruited to be the CENPA chaperone [50]

34. MCM4 This gene is negatively regulated by AR in the presence of LSD1/KDM1A activity and hypophosphorylated retinoblastoma protein (Rb) recruitment to AR binding sites, but positively in the absence of RB[49, 52] or RB hyperphosphorylation.

It is worth noting that both RB1 and KDM1A/LDSD1 are highly expressed and are similar in the responders and non-responders. The underphosphorylated RB1 represses E2F1, and we found that E2F1 is highly overexpressed in responders, suggesting that RB1 is either inactive (not so by expression) or hyperphosphorylated and therefore not repressing E2F1 or AR target genes with LDS1 and Rb1 sites.

16. ASF1B a cell cycle gene/chromatin structure gene, in the same cluster of expression as HJURP [51],CDC6, EZH2, TOP2A, AURKA, and RAD54L

23. TOP2A is regulated by AR-V7 [53], upregulated in CRPC and down in NEPC [51], and together with EZH2 defined an aggressive cancer subgroup in [54]. Of note is that TOPBP1 (p=0.008, FDR=0.11) is also significant, which is important since it is required for TOP2A recruitment during mitosis[55], while TOP2B is not, consistent with it being non-essential for progression through the mitosis [55] and TOP2A being a better target for cancer therapy [56]

10. CDKN3 is associated with progression in prostate cancer, and its knockdown lowers the level of cell cycle and DNA replication-related proteins.[57]

11. KIF23 is regulated by AR-V7 [53] and forms a complex with Aurora B [58]

12. E2F1 is elevated in non-responders and has been associated with inhibition of AR transcription, but also cooperates with EZH2 to up activate genes involved in CRPC progression.[59]

27. MYBL2 has binding sites alongside FOXA1 for Cyclin E2 (CCNE1 (FDR=0.013)) and E2F1 [60], explaining that FOXA1 axis genes can still be deregulated in the absence of FOXA1 difference.

5. NUSAP1 is involved in microtubule binding. This gene is regulated by E2F1 [61] (E2F1 regulates EZH2 [62]) and AR-V7, but not AR-FL[39].

13. TMEM132D is up in responders. It has a strong EZH2 transcription factor ChIP-seq signature in the promoter using the ENCODE 3 TF ChIP Track [63]

14. SPAG5 is downregulated in PC cells treated with enzalutamide (targeting AR) or PI3K-AKT-mTOR pathway inhibitors (dasatanib, temsirolimus, docetaxel, and others) [64]

18. RMI2 DNA repair gene, up in non-responders, with an N-MYC site in its promoter[26]. Is involved in the MYC-PARP-DNA damage response pathway.

28. RAD54L is a 2nd DNA repair gene elevated in non-responders in CRPC and controlled by AR [65].

(451) BRCA1 is a 3rd DNA repair gene elevated in non-responder.

19. UBE2C: Ubiquitin-Conjugating Enzyme 2C is regulated by AR-V7, but not AR-FL[39].[66]. It is involved in the destruction of mitotic cyclins and regulation of the cell cycle progression.

20. SAPCD2 controls cell division [67] and is associated with the development of gastric carcinoma and promotes beta-catenin and p-ERK protein levels [68].

22. TPX2 Interacting Kinetochore Protein: This gene is regulated by AR-V7, but not AR-FL[39].

24. KIFC1 is a centrosome motor protein, regulated by AR elevated in non-responders, previously known to be a biomarker of docetaxel sensitivity in CRPC.[69]

25. EXO1: DNA repair Homologous Recombination (Check other HR genes:  CHEK1, BRCA1, EXO1, BLM, RMI1, RAD54L, RAD51, LIG1, XRCC3 and RMI2, and 1 downregulated HR gene (RPA1)) see <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5855082/>

26. CDC6 is regulated by AR [70], and another gene in the cluster of genes elevated in CRPC and down in NEPC [51]. CDC6 is linked to ATR-CHK1 signaling

29. DEPDC1 is a cell cycle gene, is elevated in TCGA and lower tumor viability if inhibited[71]. High expression activated PI3K/AKT/mTOR signaling in breast cancer [72] via increased phosphorylation of PI3K, AKT, and mTOR.

30. FAM57B (now TLCD3B) regulates adipogenesis through ceramide synthesis.[73] in highly expressed in male tissues (Human Protein Atlas). This gene is predicted to have sphingosine N-acetyltransferase activity (according to genecards) and could be involved in the androgen-deprivation survival mechanism of sphK1 [74].

31. RACGAP1 is a component of the central spindling complex and is a metastatic driver in uterine carcinoma [75] and regulates survivin (BRC5)

39. BIRC5 , also known as survivin, is regulated by AR-V7, but not AR-FL[39]. Survivin has many mechanisms of action, [76] , including regulation of CDK1, which provides an AR-independent mode of action and cell cycle gene CDC25B.

33. KPNA2 This gene imports proteins into the nucleus and regulated by AR-V7, but not AR-FL[39]. It is a predictor of progression from PC to CRPC [77, 78]

35. CCNB1 is regulated by AR-V7 [53] . Note that CCNB2 is rank # 48.

36. CDC25A is elevated in non-responders, consistent with its regulation by FOXM1[43]. Intriguingly, CDC25A has been reported to act an AR co-repressor [79] of PSA, but it appears that other co-factors dominate the expression of PSA.

40. KIF2C is a kinesin, a cell cycle gene elevated in non-responders using the pre-treatment data in our previous publication[20]

41. GINS1 is a cell cycle gene

42. AURKA Aurora Kinase A regulated by AR [80]

43. NMU increased expression is associated with YAP1 (Hippo Pathway) and poor outcome in pancreatic cancer [81]

44. FAM111B upregulated upon DHT-induction (along with KLK2,PSA, and FKBP5) [82] and appears to positively regulate cyclin D1-CDK4 cell cycle progression [83]. This gene is also a target of TP53.

45. ASPM promoted progression via increased in Wnt pathway signaling [84]and is increased in non-responders.

46. CENPE A centromeric binding protein, is regulated by co-binding of LSD1 and AR to it’s promoter, indicating that non-responders have active Androgen Receptor.[49]

47. NUF2 A target gene of AR, that is highly expressed in CRPC [85]

48. CCNB2 Cell cycle gene, co-expressed with PLK1,CCNB1, and AURKA in [86]

49. SKA1 A spindle and kinetochore associated gene , that is a part of a cell cycle progression signature [29] and upregulated in non-responders.

50. BUB1 mitotic checkpoint serine/threonine-protein kinase upregulated in non-responders.

37. CLEC3B is a development and cellular differentiation gene that is Downregulated in non-responders, consistent with its behavior in Androgen deprivation therapy-resistant cell lines.[87]

13. TMEM132D is up in responders. It has a strong EZH2 transcription factor ChIP-seq)signature in the promoter using the ENCODE 3 TF ChIP Track [31]

38. NPAS3 Downregulated in non-responders, and low expression has been associated with low recurrence and survival in prostate cancer [49], with its expression regulated with dna variant rs8004379 (HR=0.61). It has also been reported to be a tumor suppressor in other cancers [50].

**Genes Upregulated in Responders (or down-regulated in non-responders)**

74. TGFBR3 Restoration of this gene function leads to inhibition of invasiveness[88]

75. GLTSCR2 Involved in stabilizing TP53, should increase relative activation of apoptosis in responders [89]

77. GPD1 Reported to be diminished in fast proliferating tumors[90] – such as our non-responders and GPD1 may enhance Metformin efficacy in prostate cancer for responders [91]

83. BOC is involved in regulation of signaling in the HedgeHog pathway via interactions with Patch and SHH, and release of SMO[92]. Because of the complex dynamics involved with Patch/SMO/SHH it remains unclear how elevated levels of BOC supports (or synergizes) response to abiratetone.

91. LOC646576 Also known as HHIP-AS1 suppresses hepatocellular carcinoma via stabilization of HHIP mRNA [93]

99. PLIN1 increase was shown to be a marker of response to Androgen Therapy[94]

101. CFD is part of a humoral immune response gene set downregulated in AR-driven (vs non AR-driven) CRPC bone metastases[95], indicating that non-responders at V2 appear to be still AR-driven, despite the abiratetone treatment, likely via the constitutive AR-V isoform expression.

Bibliography

1. Kohli M, Wang L, Xie F, Sicotte H, Yin P, Dehm SM et al. Mutational Landscapes of Sequential Prostate Metastases and Matched Patient Derived Xenografts during Enzalutamide Therapy. PLoS One 2015; 10: e0145176.

2. Wang C, Evans JM, Bhagwate AV, Prodduturi N, Sarangi V, Middha M et al. PatternCNV: a versatile tool for detecting copy number changes from exome sequencing data. Bioinformatics 2014; 30: 2678-2680.

3. Olshen VESaA. DNAcopy: DNA copy number data analysis. In. 2017.

4. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012; 30: 413-421.

5. Wilson S, Qi J, Filipp FV. Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. Sci Rep 2016; 6: 32611.

6. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol 2017; 2017.

7. Douville C, Masica DL, Stenson PD, Cooper DN, Gygax DM, Kim R et al. Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST-Indel). Hum Mutat 2016; 37: 28-35.

8. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics 2013; 14 Suppl 3: S3.

9. Masica DL, Douville C, Tokheim C, Bhattacharya R, Kim R, Moad K et al. CRAVAT 4: Cancer-Related Analysis of Variants Toolkit. Cancer Res 2017; 77: e35-e38.

10. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 2012; 13: 204-216.

11. Mahoney DW, Therneau TM, Anderson SK, Jen J, Kocher JP, Reinholz MM et al. Quality assessment metrics for whole genome gene expression profiling of paraffin embedded samples. BMC Res Notes 2013; 6: 33.

12. Kohli M, Ho Y, Hillman DW, Van Etten JL, Henzler C, Yang R et al. Androgen Receptor Variant AR-V9 Is Coexpressed with AR-V7 in Prostate Cancer Metastases and Predicts Abiraterone Resistance. Clin Cancer Res 2017; 23: 4704-4715.

13. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.

14. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 2006; 7: 3.

15. Wang L, Nie J, Sicotte H, Li Y, Eckel-Passow JE, Dasari S et al. Measure transcript integrity using RNA-seq data. BMC Bioinformatics 2016; 17: 58.

16. Derrick B, Dobson-Mckittrick, A., Toher, D. and White, P. . est statistics for comparing two proportions with partially overlapping samples. . Journal of Applied Quantitative Methods 2015; 10: 1-14.

17. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27: 1739-1740.

18. Nikolayeva O, Robinson MD. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. Methods Mol Biol 2014; 1150: 45-79.

19. Goetz MP, Kalari KR, Suman VJ, Moyer AM, Yu J, Visscher DW et al. Tumor Sequencing and Patient-Derived Xenografts in the Neoadjuvant Treatment of Breast Cancer. J Natl Cancer Inst 2017; 109.

20. Wang L, Dehm SM, Hillman DW, Sicotte H, Tan W, Gormley M et al. A Prospective Genome-Wide Study of Prostate Cancer Metastases Reveals Association of Wnt Pathway Activation and Increased Cell Cycle Proliferation with Primary Resistance to Abiraterone Acetate-Prednisone. Ann Oncol 2017.

21. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics 2016; 54: 1 30 31-31 30 33.

22. Hieronymus H, Lamb J, Ross KN, Peng XP, Clement C, Rodina A et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell 2006; 10: 321-330.

23. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med 2016; 22: 369-378.

24. Patel GK, Chugh N, Tripathi M. Neuroendocrine Differentiation of Prostate Cancer-An Intriguing Example of Tumor Evolution at Play. Cancers (Basel) 2019; 11.

25. Tsai HK, Lehrer J, Alshalalfa M, Erho N, Davicioni E, Lotan TL. Gene expression signatures of neuroendocrine prostate cancer and primary small cell prostatic carcinoma. BMC Cancer 2017; 17: 759.

26. Zhang W, Liu B, Wu W, Li L, Broom BM, Basourakos SP et al. Targeting the MYCN-PARP-DNA Damage Response Pathway in Neuroendocrine Prostate Cancer. Clin Cancer Res 2018; 24: 696-707.

27. Polkinghorn WR, Parker JS, Lee MX, Kass EM, Spratt DE, Iaquinta PJ et al. Androgen receptor signaling regulates DNA repair in prostate cancers. Cancer Discov 2013; 3: 1245-1253.

28. Mostaghel EA, Marck BT, Plymate SR, Vessella RL, Balk S, Matsumoto AM et al. Resistance to CYP17A1 inhibition with abiraterone in castration-resistant prostate cancer: induction of steroidogenesis and androgen receptor splice variants. Clin Cancer Res 2011; 17: 5913-5925.

29. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol 2011; 12: 245-255.

30. Gupta S, Li J, Kemeny G, Bitting RL, Beaver J, Somarelli JA et al. Whole Genomic Copy Number Alterations in Circulating Tumor Cells from Men with Abiraterone or Enzalutamide-Resistant Metastatic Castration-Resistant Prostate Cancer. Clin Cancer Res 2017; 23: 1346-1357.

31. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 2010; 18: 11-22.

32. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med 2016; 22: 298-305.

33. Jin HJ, Kim J, Yu J. Androgen receptor genomic regulation. Transl Androl Urol 2013; 2: 157-177.

34. Zhang Q, Huang H, Liu A, Li J, Liu C, Sun B et al. Cell division cycle 20 (CDC20) drives prostate cancer progression via stabilization of beta-catenin in cancer stem-like cells. EBioMedicine 2019; 42: 397-407.

35. Li K, Mao Y, Lu L, Hu C, Wang D, Si-Tu J et al. Silencing of CDC20 suppresses metastatic castration-resistant prostate cancer growth and enhances chemosensitivity to docetaxel. Int J Oncol 2016; 49: 1679-1685.

36. Wu C, Jin X, Yang J, Yang Y, He Y, Ding L et al. Inhibition of EZH2 by chemo- and radiotherapy agents and small molecule inhibitors induces cell death in castration-resistant prostate cancer. Oncotarget 2016; 7: 3440-3452.

37. Wang J, Cheng P, Pavlyukov MS, Yu H, Zhang Z, Kim SH et al. Targeting NEK2 attenuates glioblastoma growth and radioresistance by destabilizing histone methyltransferase EZH2. J Clin Invest 2017; 127: 3075-3089.

38. Mbom BC, Siemers KA, Ostrowski MA, Nelson WJ, Barth AI. Nek2 phosphorylates and stabilizes beta-catenin at mitotic centrosomes downstream of Plk1. Mol Biol Cell 2014; 25: 977-991.

39. Hu R, Lu C, Mostaghel EA, Yegnasubramanian S, Gurel M, Tannahill C et al. Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. Cancer Res 2012; 72: 3457-3462.

40. Urbanucci A, Sahu B, Seppala J, Larjo A, Latonen LM, Waltering KK et al. Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer. Oncogene 2012; 31: 2153-2163.

41. Obuse C, Iwasaki O, Kiyomitsu T, Goshima G, Toyoda Y, Yanagida M. A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. Nat Cell Biol 2004; 6: 1135-1141.

42. Ketola K, Munuganti RSN, Davies A, Nip KM, Bishop JL, Zoubeidi A. Targeting Prostate Cancer Subtype 1 by Forkhead Box M1 Pathway Inhibition. Clin Cancer Res 2017; 23: 6923-6933.

43. Liu Y, Liu Y, Yuan B, Yin L, Peng Y, Yu X et al. FOXM1 promotes the progression of prostate cancer by regulating PSA gene transcription. Oncotarget 2017; 8: 17027-17037.

44. Cai C, He HH, Chen S, Coleman I, Wang H, Fang Z et al. Androgen receptor gene expression in prostate cancer is directly suppressed by the androgen receptor through recruitment of lysine-specific demethylase 1. Cancer Cell 2011; 20: 457-471.

45. Hu G, Yan Z, Zhang C, Cheng M, Yan Y, Wang Y et al. FOXM1 promotes hepatocellular carcinoma progression by regulating KIF4A expression. J Exp Clin Cancer Res 2019; 38: 188.

46. Chen K, Cao Q, Song Z, Ruan H, Wang C, Yang X et al. Targeting the KIF4A/AR axis to reverse endocrine therapy resistance in castration-resistant prostate cancer. Clin Cancer Res 2019.

47. Liu X, Gao Y, Ye H, Gerrin S, Ma F, Wu Y et al. Positive feedback loop mediated by protein phosphatase 1alpha mobilization of P-TEFb and basal CDK1 drives androgen receptor in prostate cancer. Nucleic Acids Res 2017; 45: 3738-3751.

48. Ye J, Chu C, Chen M, Shi Z, Gan S, Qu F et al. TROAP regulates prostate cancer progression via the WNT3/survivin signalling pathways. Oncol Rep 2019; 41: 1169-1179.

49. Liang Y, Ahmed M, Guo H, Soares F, Hua JT, Gao S et al. LSD1-Mediated Epigenetic Reprogramming Drives CENPE Expression and Prostate Cancer Progression. Cancer Res 2017; 77: 5479-5490.

50. Bergmann JH, Rodriguez MG, Martins NM, Kimura H, Kelly DA, Masumoto H et al. Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J 2011; 30: 328-340.

51. Maina PK, Shao P, Liu Q, Fazli L, Tyler S, Nasir M et al. c-MYC drives histone demethylase PHF8 during neuroendocrine differentiation and in castration-resistant prostate cancer. Oncotarget 2016; 7: 75585-75602.

52. Gao S, Gao Y, He HH, Han D, Han W, Avery A et al. Androgen Receptor Tumor Suppressor Function Is Mediated by Recruitment of Retinoblastoma Protein. Cell Rep 2016; 17: 966-976.

53. Magani F, Bray ER, Martinez MJ, Zhao N, Copello VA, Heidman L et al. Identification of an oncogenic network with prognostic and therapeutic value in prostate cancer. Mol Syst Biol 2018; 14: e8202.

54. Labbe DP, Sweeney CJ, Brown M, Galbo P, Rosario S, Wadosky KM et al. TOP2A and EZH2 Provide Early Detection of an Aggressive Prostate Cancer Subgroup. Clin Cancer Res 2017; 23: 7072-7083.

55. Broderick R, Nieminuszczy J, Blackford AN, Winczura A, Niedzwiedz W. TOPBP1 recruits TOP2A to ultra-fine anaphase bridges to aid in their resolution. Nat Commun 2015; 6: 6572.

56. Austin CA, Lee KC, Swan RL, Khazeem MM, Manville CM, Cridland P et al. TOP2B: The First Thirty Years. Int J Mol Sci 2018; 19.

57. Yu C, Cao H, He X, Sun P, Feng Y, Chen L et al. Cyclin-dependent kinase inhibitor 3 (CDKN3) plays a critical role in prostate cancer via regulating cell cycle and DNA replication signaling. Biomed Pharmacother 2017; 96: 1109-1118.

58. Hornick JE, Karanjeet K, Collins ES, Hinchcliffe EH. Kinesins to the core: The role of microtubule-based motor proteins in building the mitotic spindle midzone. Semin Cell Dev Biol 2010; 21: 290-299.

59. Xu H, Xu K, He HH, Zang C, Chen CH, Chen Y et al. Integrative Analysis Reveals the Transcriptional Collaboration between EZH2 and E2F1 in the Regulation of Cancer-Related Gene Expression. Mol Cancer Res 2016; 14: 163-172.

60. Zhang C, Wang L, Wu D, Chen H, Chen Z, Thomas-Ahner JM et al. Definition of a FoxA1 Cistrome that is crucial for G1 to S-phase cell-cycle transit in castration-resistant prostate cancer. Cancer Res 2011; 71: 6738-6748.

61. Gulzar ZG, McKenney JK, Brooks JD. Increased expression of NuSAP in recurrent prostate cancer is mediated by E2F1. Oncogene 2013; 32: 70-77.

62. Lee SR, Roh YG, Kim SK, Lee JS, Seol SY, Lee HH et al. Activation of EZH2 and SUZ12 Regulated by E2F1 Predicts the Disease Progression and Aggressive Characteristics of Bladder Cancer. Clin Cancer Res 2015; 21: 5391-5403.

63. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res 2012; 22: 1798-1812.

64. Ebhardt HA, Root A, Liu Y, Gauthier NP, Sander C, Aebersold R. Systems pharmacology using mass spectrometry identifies critical response nodes in prostate cancer. NPJ Syst Biol Appl 2018; 4: 26.

65. Li L, Karanika S, Yang G, Wang J, Park S, Broom BM et al. Androgen receptor inhibitor-induced "BRCAness" and PARP inhibition are synthetically lethal for castration-resistant prostate cancer. Sci Signal 2017; 10.

66. Lee CH, Ku JY, Ha JM, Bae SS, Lee JZ, Kim CS et al. Transcript Levels of Androgen Receptor Variant 7 and Ubiquitin-Conjugating Enzyme 2C in Hormone Sensitive Prostate Cancer and Castration-Resistant Prostate Cancer. Prostate 2017; 77: 60-71.

67. Chiu CWN, Monat C, Robitaille M, Lacomme M, Daulat AM, Macleod G et al. SAPCD2 Controls Spindle Orientation and Asymmetric Divisions by Negatively Regulating the Galphai-LGN-NuMA Ternary Complex. Dev Cell 2016; 36: 50-62.

68. Cao WJ, Du WQ, Mao LL, Zheng JN, Pei DS. Overexpression of p42.3 promotes cell proliferation, migration, and invasion in human gastric cancer cells. Tumour Biol 2016; 37: 12805-12812.

69. Sekino Y, Oue N, Koike Y, Shigematsu Y, Sakamoto N, Sentani K et al. KIFC1 Inhibitor CW069 Induces Apoptosis and Reverses Resistance to Docetaxel in Prostate Cancer. J Clin Med 2019; 8.

70. Thompson TC, Li L. Connecting androgen receptor signaling and the DNA damage response: Development of new therapies for advanced prostate cancer. Mol Cell Oncol 2017; 4: e1321167.

71. Ramalho-Carvalho J, Martins JB, Cekaite L, Sveen A, Torres-Ferreira J, Graca I et al. Epigenetic disruption of miR-130a promotes prostate cancer by targeting SEC23B and DEPDC1. Cancer Lett 2017; 385: 150-159.

72. Zhao H, Yu M, Sui L, Gong B, Zhou B, Chen J et al. High Expression of DEPDC1 Promotes Malignant Phenotypes of Breast Cancer Cells and Predicts Poor Prognosis in Patients With Breast Cancer. Front Oncol 2019; 9: 262.

73. Yamashita-Sugahara Y, Tokuzawa Y, Nakachi Y, Kanesaki-Yatsuka Y, Matsumoto M, Mizuno Y et al. Fam57b (family with sequence similarity 57, member B), a novel peroxisome proliferator-activated receptor gamma target gene that regulates adipogenesis through ceramide synthesis. J Biol Chem 2013; 288: 4522-4537.

74. Dayon A, Brizuela L, Martin C, Mazerolles C, Pirot N, Doumerc N et al. Sphingosine kinase-1 is central to androgen-regulated prostate cancer growth and survival. PLoS One 2009; 4: e8048.

75. Mi S, Lin M, Brouwer-Visser J, Heim J, Smotkin D, Hebert T et al. RNA-seq Identification of RACGAP1 as a Metastatic Driver in Uterine Carcinosarcoma. Clin Cancer Res 2016; 22: 4676-4686.

76. Li F, Aljahdali I, Ling X. Cancer therapeutics using survivin BIRC5 as a target: what can we do after over two decades of study? J Exp Clin Cancer Res 2019; 38: 368.

77. Mortezavi A, Hermanns T, Seifert HH, Baumgartner MK, Provenzano M, Sulser T et al. KPNA2 expression is an independent adverse predictor of biochemical recurrence after radical prostatectomy. Clin Cancer Res 2011; 17: 1111-1121.

78. Tzelepi V, Zhang J, Lu JF, Kleb B, Wu G, Wan X et al. Modeling a lethal prostate cancer variant with small-cell carcinoma features. Clin Cancer Res 2012; 18: 666-677.

79. Chiu YT, Han HY, Leung SC, Yuen HF, Chau CW, Guo Z et al. CDC25A functions as a novel Ar corepressor in prostate cancer cells. J Mol Biol 2009; 385: 446-456.

80. Kivinummi K, Urbanucci A, Leinonen K, Tammela TLJ, Annala M, Isaacs WB et al. The expression of AURKA is androgen regulated in castration-resistant prostate cancer. Sci Rep 2017; 7: 17978.

81. Yoo W, Lee J, Jun E, Noh KH, Lee S, Jung D et al. The YAP1-NMU Axis Is Associated with Pancreatic Cancer Progression and Poor Outcome: Identification of a Novel Diagnostic Biomarker and Therapeutic Target. Cancers (Basel) 2019; 11.

82. Chen ST, Okada M, Nakato R, Izumi K, Bando M, Shirahige K. The Deubiquitinating Enzyme USP7 Regulates Androgen Receptor Activity by Modulating Its Binding to Chromatin. J Biol Chem 2015; 290: 21713-21723.

83. Kawasaki K, Nojima S, Hijiki S, Tahara S, Ohshima K, Matsui T et al. FAM111B enhances proliferation of KRAS-driven lung adenocarcinoma by degrading p16. Cancer Sci 2020; 111: 2635-2646.

84. Pai VC, Hsu CC, Chan TS, Liao WY, Chuu CP, Chen WY et al. ASPM promotes prostate cancer stemness and progression by augmenting Wnt-Dvl-3-beta-catenin signaling. Oncogene 2019; 38: 1340-1353.

85. Yamamoto S, Takayama KI, Obinata D, Fujiwara K, Ashikari D, Takahashi S et al. Identification of new octamer transcription factor 1-target genes upregulated in castration-resistant prostate cancer. Cancer Sci 2019; 110: 3476-3485.

86. Huang CG, Li FX, Pan S, Xu CB, Dai JQ, Zhao XH. Identification of genes associated with castrationresistant prostate cancer by gene expression profile analysis. Mol Med Rep 2017; 16: 6803-6813.

87. Marques RB, Dits NF, Erkens-Schulze S, van Weerden WM, Jenster G. Bypass mechanisms of the androgen receptor pathway in therapy-resistant prostate cancer cell models. PLoS One 2010; 5: e13500.

88. Cao Z, Kyprianou N. Mechanisms navigating the TGF-beta pathway in prostate cancer. Asian J Urol 2015; 2: 11-18.

89. Lee S, Kim JY, Kim YJ, Seok KO, Kim JH, Chang YJ et al. Nucleolar protein GLTSCR2 stabilizes p53 in response to ribosomal stresses. Cell Death Differ 2012; 19: 1613-1622.

90. Singh G. Mitochondrial FAD-linked Glycerol-3-phosphate Dehydrogenase: A Target for Cancer Therapeutics. Pharmaceuticals (Basel) 2014; 7: 192-206.

91. Xie J, Ye J, Cai Z, Luo Y, Zhu X, Deng Y et al. GPD1 Enhances the Anticancer Effects of Metformin by Synergistically Increasing Total Cellular Glycerol-3-Phosphate. Cancer Res 2020; 80: 2150-2162.

92. Kim Y, Lee J, Seppala M, Cobourne MT, Kim SH. Ptch2/Gas1 and Ptch1/Boc differentially regulate Hedgehog signalling in murine primordial germ cell migration. Nat Commun 2020; 11: 1994.

93. Bo C, Li X, He L, Zhang S, Li N, An Y. A novel long noncoding RNA HHIP-AS1 suppresses hepatocellular carcinoma progression through stabilizing HHIP mRNA. Biochem Biophys Res Commun 2019; 520: 333-340.

94. Tousignant KD, Rockstroh A, Poad BLJ, Talebi A, Young RSE, Taherian Fard A et al. Therapy-induced lipid uptake and remodeling underpin ferroptosis hypersensitivity in prostate cancer. Cancer Metab 2020; 8: 11.

95. Ylitalo EB, Thysell E, Jernberg E, Lundholm M, Crnalic S, Egevad L et al. Subgroups of Castration-resistant Prostate Cancer Bone Metastases Defined Through an Inverse Relationship Between Androgen Receptor Activity and Immune Response. Eur Urol 2017; 71: 776-787.