# Supplementary Figures Captions

## Supplementary Figure\_S1

The DNA/RNA processing used both single sample and multi-sample analyses as part of the processing. The right pane shows that the two tumor biopsies and 1 germline samples were processed together (BAM realigned and joint somating calling). The Copy number calling did not used paired samples, but rather matched each samples against a collection of reference samples so that the variance for each tumor log2 ratio is dominated by the tumor noise. The left pane shows that the RNASeq samples were processed individually, but that the

## Supplementary Figure\_S2

Example for 1 sample of the plots used to center and rescale of log2 ratio for CNV so that all CNV data log2 ratio corresponds to levels expected in a pure tumor. A. Raw segmentation of gene-averaged CNV log2 ratio. B. Histogram of raw gene-average CNV log2 ratio before recentering and rescaling. C. Smoothed histogram of raw gene-averaged CNV log2 ratio before recentering and rescaling needed to identify log2 levels corresponding to discrete copy levels (0,1,2,3,4,5,..etc copies). D. Sorted plots of raw gene-averaged CNV log2 ratio before used to identify levels together with (C.) E. Purity-adjusted and recentered segmentation of gene-averaged CNV log2 ratio. D. Histogram gene-average CNV log2 ratio after recentering and rescaling. E. Smoothed histogram of raw gene-averaged CNV log2 ratio after recentering. Main peak centered on 0.0 is 4 copies for this sample (tetraploid tumor). F. Sorted plots of recentered and rescaled gene-averaged CNV log2 ratio. Most genes are at 4 copies for this tetraploid tumor.

## Supplementary Figure\_S3

Purity does not appear to differ for Responders and Non-responders. A) Boxplot of Copy Number derived Sample Purity (passing 15% CNV QC cutoff) vs visit, limited to samples in the lowest quartile of TTTC (non-responders) and the upper quartile of TTTC (responders) B) Boxplot of Copy Number derived Sample Purity (passing 5% Somatic Mutations QC cutoff) vs visit, limited to samples in the lowest quartile of TTTC (non-responders) and the upper quartile of TTTC (responders)

## Supplementary Figure S4:

Plots of the gene level (median of the probe level) rescaled/recentered log2 ratio (CNV) around the UGT2B17 gene which is frequently deleted in the germline. The error bars represent ±1 standard deviation around the median value. Panel A, a rare Homozygous Gain. Panel B) A rare Heterozygous deletion of the UGT2B15 gene. Panel B) A Homozygous deletion (data truncated). Panel D) The UGT2B17 gene is not deleted (or gain) and this is detectable even in this low purity student. Note that a +1 gain would be at log2 of 0.57. This low purity sample exemplifies why the Copy Number data is resegmented at the gene level.

## Supplementary Figure S5:

A measure of the noisiness of CNV calls for each samples was obtained by computing the robust Standard Deviation (mad) of the rescaled gene-level calls minus the segmented levels for chromosome 9. This shows that the variance around the segmentation values is independent of the purity of the sample, even though low-purity samples underwent rescaling which could have potentially increased the relative noise.

## Supplementary Figure\_S6

A) Purity vs Log10(TTTC V2) shows that purity is not associated with the phenotype. B) The Immunity Score is strongly associated with Purity, which makes sense since lower purity samples have more normal cells or Immune cell infiltrates.

## Supplementary Figure\_S7

QC plot for RNA-seq showing the dfArray and stress metrics for all samples. Outliers have a median dfArray >= 1.5 and/or median stress > 1.0.

## Supplementary Figure\_S8

QC plot for RNA-seq showing the dfArray and stress metrics for high quality samples only, after removing the outliers. This shows no large batch effect based on visit or tissue type.

## Supplementary Figure\_S9

Heatmap of the GSVA gene sets with the highest median absolute deviation (mad). Below the heatmap, there is an annotation tab showing various QC metrics described in supplementary methods(IQR=1 flags samples that are outliers according to IQR, ARreads flags samples with low AR mapping reads (outliers according to AR reads IQR), whether the dfarray metric is too high (>1 or >2), a LOWPURITY indicator or low purity samples, two purity estimates, indicators of degradation(TIN/RIN), and mapping percentages) and data features that we expected to be present in many tumors (e.g. TMPRSS2-ERG fusion reads at various cutoffs). Samples on the left were not prostate cancer samples, so samples that fell in this heatmap cluster were deemed to be of low quality and removed from analysis. The next cluster (colored purple in the clustering tree at the top) of samples shows a lot of poor quality metrics, but was retained and labelled “SECONDARY”. Details of the QC metrics can be found in the Supplementary Methods.

## Supplementary Figure\_S10

Heatmap of the CNV GSVA signal at V2 for gene sets significantly different between non-responders and responders. Samples are ordered (left to right) according to TTTC. Annotations below the heatmap indicate AR Amplifications at V2 (ARV2=Amp), AR Amplifications at V1 (ARV1=Amp), and PTEN deletions at V1 or V2, with 1 copy deletion being indicated by “-1” (PTENV1=-1, PTENV2=-1) and two copy deletions being indicated by “-2” (PTENV1=-2, PTENV2=-2). For AR and PTEN features, grey indicates no DNA passing QC was available for V1 DNA samples matching the V2 or was unable to call the AR gene. Also included are annotations (black coloring) for samples with at least 2 TMPRSS2-ERG fusion supporting reads at V1 or V2, with grey indicating RNA-seq sample matching the V1 DNA sample.

## Supplementary Figure\_S11

Heatmap of the CNV GSVA signal at V1 for gene sets significantly different between nonresponders and responders. Samples are ordered (left to right) according to TTTC (annotated at bottom of plot). Annotations below the heatmap indicate AR Amplifications at V2 (ARV2=Amp), AR Amplifications at V1 (ARV1=Amp), and PTEN deletions at V1 or V2, with 1 copy deletion being indicated by “-1” (PTENV1=-1, PTENV2=-1) and two copy deletions being indicated by “-2” (PTENV1=-2, PTENV2=-2). For AR and PTEN features, grey indicates no DNA passing QC was available for V2 DNA samples matching the V1. Also included are annotations (black coloring) for samples with at least 2 TMPRSS2-ERG fusion supporting reads at V1 or V2, with grey indicating RNA-seq sample matching the V1 DNA sample.

## Supplementary Figure\_S12

Heatmap of the CNV GSVA signal at V2 minus the signal at V1 (paired for same patient) for gene sets significantly different between nonresponders and responders. Samples are ordered (left to right) according to TTTC (annotated at bottom of plot). Annotations below the heatmap indicate AR Amplifications at V2 (ARV2=Amp), AR Amplifications at V1 (ARV1=Amp), and PTEN deletions at V1 or V2, with 1 copy deletion being indicated by “-1” (PTENV1=-1, PTENV2=-1) and two copy deletions being indicated by “-2” (PTENV1=-2, PTENV2=-2). For AR and PTEN features, grey indicates no DNA passing QC was available for V2 DNA samples matching the V1. Also included are annotations (black coloring) for samples with at least 2 TMPRSS2-ERG fusion supporting reads at V1 or V2, with grey indicating RNA-seq sample matching the V1 DNA sample.

## Supplementary Figure\_S13

1. Violin plot of the VDR gene in responders vs non-responders showing increased expression in responders.
2. Violin plot of the RXRG gene in responders vs non-responders showing increased expression in responders.

## Supplementary Figure\_S14

1. Normalized counts of Splice Junction supporting AR isoforms V1 V3 V3 V4 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
2. Normalized counts of Splice Junction supporting AR isoform AR\_V8 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
3. Normalized counts of Splice Junction supporting AR isoform AR\_V9 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
4. Normalized counts of Splice Junction supporting AR isoform AR\_23 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
5. Normalized counts of Splice Junction supporting AR isoform AR\_V3 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
6. Normalized counts of Splice Junction supporting AR isoforms AR\_V3 and AR\_V4 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
7. Normalized counts of Splice Junction supporting AR isoform AR\_45 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
8. Normalized counts of Splice Junction supporting AR isoform AR\_V7 and AR\_V8 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
9. Normalized counts of Splice Junction supporting AR isoform AR\_V7 , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.
10. Normalized counts of Splice Junction supporting Full Length AR , showing non-responder’s V1 and V2 distribution and responder’s V1 and V2 distribution.