Supplemental Methods

Sample QC: Of the original 128 tumor samples sequenced, we first removed samples due to contamination estimates which resulted in a subset of 110 samples. Tumor samples with a ContEst score of greater than 5% were excluded from our analysis (1). We then sought to identify samples with low tumor content and exclude them from analysis. Low tumor content candidate samples were identified on the basis of three criteria: few discernible copy number variants (CNVs), low mutation rate, and low purity estimates returned by ABSOLUTE (2). Global mutation rates were computed by MutSigCV, and samples with mutation rates < 0.5 Mb^{-1} were flagged (3). With respect to CNVs, hierarchical clustering of 110 samples (see Methods below) revealed a distinct cluster with few to no CNVs, suggesting stromal contamination. Samples in this cluster were flagged. Finally, ABSOLUTE was run on all samples, and those with purity estimates < 20% tumor content were flagged. Samples flagged according to all three criteria (8 total) were removed from analysis for a final set of 102.

Methods for targeted panel and exome analysis: For the extension set, of the 172 sequenced tumor-normal pairs, 90 pairs were selected for analysis based on a cutoff of at least 50x coverage on more than 85% of target bases for tumor samples, and 60% of target bases for normal samples. This resulted in a mean target coverage of 335x for the 90 tumor samples and 348x for the 90 normal samples. For the targeted panel and exome data, detection and calling of somatic single-nucleotide

variants (SSNVs), germline variants, small insertions and deletions (indels) were performed through the Firehose (http://www.broadinstitute.org/cancer/cga/Firehose) environment at the Broad Institute using The Genome Analysis Toolkit and other previously published algorithms, including Mutect, Indelocator (http://www.broadinstitute.org/cancer/cga/indelocator), and HaplotypeCaller.(4-7) Artifacts of C>T base substitutions formed in FFPE samples due to cytosine deamination to uracil were removed as previously described (8). Called SSNVs and small indels were then annotated for variant type, codon change, amino acid change and other relevant information using Oncotator (9). DNA copy number alterations were identified from whole exome data using CapSeg.

Tumor purity analysis:

Tumor purity in the whole exome sequencing study was estimated using ABSOLUTE (2).

AAPC global heatmap: GISTIC2.0 was run on 102 AAPC samples to generate gene-level normalized log2 copy ratio data for every sample (10). Cytoband-level copy ratios were inferred from gene-level data by selecting the median copy ratio among all genes in each cytoband. Once cytoband-level data was obtained for each sample, the cytobands were arranged in genomic order and a heatmap was generated using the pheatmaps (https://cran.r-project.org/web/packages/pheatmap/index.html) R

package. Unsupervised heirarchical clustering of samples was done using Ward's method.

Somatic Copy Number Alteration cohort comparisons: GISTIC2.0 was run on both AAPC and TCGA cohorts to generate gene-level copy number calls on a five-point scale: -2 (deep deletion), -1 (shallow deletion), 0 (no event), 1 (low amplification), 2 (high amplification). In the GISTIC analysis, we used a threshold of log2 copy number ratio >0.3 and < -0.3. A curated list of genes with recurrent SCNAs in primary PCa was selected and CNV data from both cohorts was probed for the frequency of events in these genes.(11) Comparisons were done both cohort-wide as well as between subsets with the same Gleason score. Gleason-specific comparisons were done for Gleason 6, 7 (3+4 and 4+3 individually as well as total Gleason 7), and 8+. P-values were computed by Fisher's exact test. Fraction of the copy number altered genome was calculated by dividing the number of bases with estimated log2 copy number ratio > 0.1 and < -0.1 by the total number of bases sequenced. As described above, using the threshold of log2 copy number ratio > 0.3 and < -0.3 the fraction of copy number altered genome was also calculated for samples in the AAPC cohort.

SU2C + TCGA + AAPC stickplot: Point mutations and indels were called in three separate cohorts: Stand Up To Cancer metastatic CRPC (n= 303), TCGA primary prostate cancer (n=492) and AAPC (n=102). Mutations and indels were called using MuTect and Indelocator, respectively (7). Nonsense mutations, missense mutations,

and indels called in *ERF* in the three cohorts were overlaid onto a linear representation of the gene.

SU2C mutation plot: Mutations and indels were called in the Stand Up To Cancer (SU2C) metastatic CRPC cohort (n=303) using MuTect and Indelocator (unpublished). CNVs were called by ReCapSeg (unpublished). Samples with nonsynonymous mutations, indels, and deletions (both hemizygous and biallelic) in *ERF* were marked and colored as labeled. A total of 269 individual samples were used after accounting for patient duplicate samples and quality for copy number calls. All other SU2C samples (*ERF* wt) are shown for context and colored grey.

ERF overexpression: Constructs containing the open reading frame of ERF in the plx304 expression vector were obtained from the Genomics Perturbation Platform (Broad). Site-directed mutagenesis was performed using the Quikchange XL Lightning (Agilent) system. Primers are available upon request. Constructs were sequence verified using Sanger sequencing.

Proliferation Assay: Cells were seeded into tissue culture-treated 96-well black plates (3904; Corning) at a density of 1,000 (LN-CaP, PC3) or 2,000 (RWPE) cells per well, suspended in 100 µL of culture medium supplemented with 10% FBS. After 5 days in culture (37°C in 5% CO_2), cell viability was measured with the CellTiter-Glo (Promega) luminescent assay per manufacturer's instructions. The mean and standard deviation of six replicates per condition were used for analysis.

Growth in low attachment (GILA) assay: Cells were seeded into 96-well ultra-low attachment plates (3474; Corning) at a density of 2,000 (RWPE) or 5,000 (DU-145) cells per well suspended in 100 µL of culture medium supplemented with 10% FBS. After 5 days of incubation at 37°C, 5% CO_2, the CellTiter-Glo (Promega) assay was used to measure viability. After CTG addition and incubation according to manufacturer's instructions, the cells were transferred to 96-well black plates (3904; Corning) and luminescence was read using the EnVision Plate Reader (PerkinElmer). The mean and standard deviation of six replicates per condition were used for analysis.

Invasion Assay: Cells stably expressing shRNAs were trypsinized, counted, and re-suspended in serum free medium. $2x10^5$ cells suspended in 400 µL of serum free medium were seeded in triplicate in the upper chamber of BD BioCoat Matrigel Inserts (24-well plates, 8-µm PET Membrane; Becton Dickinson, San Jose, CA). 300 µL of culture medium containing 10% FBS were added to the bottom of the wells (except for two control wells in which 300 µL of serum free medium were added for analysis purposes). Cells were incubated for 48 hours at 37°C in 5% CO_2. After incubation, cells in the upper chamber were removed with a cotton swab and cells that trespassed the membrane were fixed in 1 mL of methanol, stained with 0.5% crystal violet, and washed with distilled water. Quantification of crystal violet uptake was performed by submerging the bottom chamber in 500 µL of 10% acetic acid, followed by shaking for one hour. Subsequently, three aliquots of 100 µL were

transferred (for each of the three replicates per condition) to a 96-well clear plate and absorbance was measured at 595 nm on the SpectraMax 190 instrument.

Mouse tumor xenografts: All animal experiments were approved by the Dana-Farber Cancer Institute Institutional Animal Care and Use Committee and were performed in accordance with institutional and national guidelines.

$1x10^6$ PC-3 cells in 50% Matrigel (BD Corning) stably expressing control shRNA or shRNAs against *ERF* were subcutaneously injected into both flanks of IcrTac:ICR-*Prkdc*^{scid} mice (Taconic). Tumor growth was observed within a week of injections due to the rapid growth of PC-3 cells *in vivo*. We measured tumor growth every 2-4 days for 3 weeks until tumor sizes reached ~2cm diameter. Average tumor volume growth was calculated in comparison to first measurable tumor volumes that occurred on day 7 after injection.

GSEA signature enrichment profiles. To generate sample per sample Gene Set Enrichment Analysis (GSEA) profiles we used the single-sample GSEA method introduced in (Barbie et al. 2009) and available as part of the GenePattern environment (www.genepattern.org) (12). This procedure was used in the following signatures:

- The shERF signatures described in Methods.
- An ERG signature from the c2 sub-collection of MSigDB. This signature contains 67 up-regulated genes in prostate cancers with the TMPRSS2-ERG

fusion: “SETLUR_PROSTATE_CANCER_TMPRSS2_ERG_FUSION_UP” (Setlur et al. 2008).(13)

- Androgen Receptor I: 150-gene gene set of AR-regulated genes from (Sharma et al. 2013): “Sharma_AR_UP.”(14)
- Androgen Receptor II: 21-gene gene set of AR-mediated regulated genes from (Hieronymus et al. 2006): “AR_Hieronymus_at_al.”(15)

Association between signatures and genomic features. In order to estimate the association between the different transcriptional signatures, genomic features and phenotypes we used the Information Coefficient (IC).(16) The IC is a normalized version of the mutual information defined as,

$$IC(x,y) = sign(\rho(x,y))\sqrt{\left(1-\exp(2I(x,y))\right)}, \qquad (1)$$

where x and y are e.g. signature profiles or genomic features, used for this task is a normalized version of the mutual information $I(x,y)$ is the differential mutual information and the *sign* of the correlation coefficient $\rho(x,y)$is used to provide directionality to the coefficient (16-18).

Mutation analysis of combined TCGA and AAPC cohorts:

We combined the mutation list from the 102 AAPC exome cohort samples and aggregated them with the mutation list from prostate TCGA cohort (n=457). TCGA samples were passed through the same workflow as the AAPC cohort (see methods above). The aggregated calls were then run through MutsigCV (3). Genes with a FDR

q value less than 0.1 were considered significantly mutated.

1. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics.* 2011;27(18):2601-2602.
2. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012;30(5):413-421.
3. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-218.
4. McKenna A, Hanna M Fau - Banks E, Banks E Fau - Sivachenko A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. (1549-5469 (Electronic)).
5. DePristo MA, Banks E Fau - Poplin R, Poplin R Fau - Garimella KV, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. (1546-1718 (Electronic)).
6. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. (1934-340X (Electronic)).
7. Cibulskis K, Lawrence Ms Fau - Carter SL, Carter Sl Fau - Sivachenko A, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. (1546-1696 (Electronic)).
8. Giannakis M, Mu XJ, Shukla SA, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. LID - S2211-1247(16)30364-3 [pii] LID - 10.1016/j.celrep.2016.03.075 [doi]. (2211-1247 (Electronic)).
9. Ramos AH, Lichtenstein L Fau - Gupta M, Gupta M Fau - Lawrence MS, et al. Oncotator: cancer variant annotation tool. (1098-1004 (Electronic)).
10. Mermel CH, Schumacher Se Fau - Hill B, Hill B Fau - Meyerson ML, Meyerson Ml Fau - Beroukhim R, Beroukhim R Fau - Getz G, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. (1474-760X (Electronic)).
11. Williams JL, Greer PA, Squire JA. Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data. *Cancer genetics.* 2014;207(10-12):474-488.
12. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* 2009;462(7269):108-112.
13. Setlur SR, Mertz KD, Hoshida Y, et al. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst.* 2008;100(11):815-825.
14. Sharma NL, Massie CE, Ramos-Montoya A, et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer Cell.* 2013;23(1):35-47.

15. Hieronymus H, Lamb J, Ross KN, et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell.* 2006;10(4):321-330.
16. Kim JW, Botvinnik OB, Abudayyeh O, et al. Characterizing genomic alterations in cancer by complementary functional associations. *Nat Biotechnol.* 2016.
17. Joe H. Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association.* 1989;84(405):157-164.
18. Linfoot EH. An informational measure of correlation. *Information and Control.* 1957;1(3):85-89.