Supplementary Information Titles

Please list each supplementary item and its title or caption, in the order shown below.

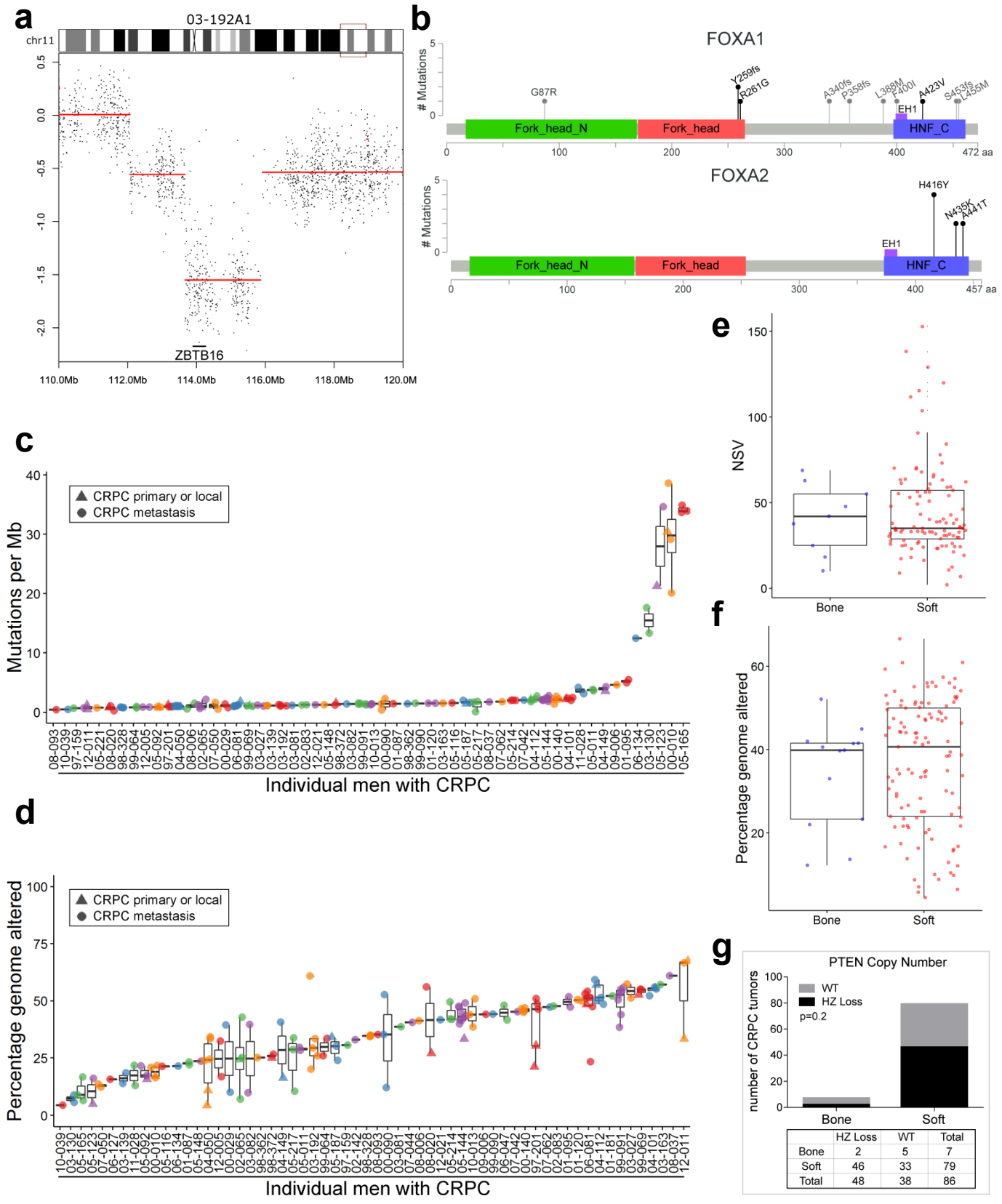
Note that we do NOT copy edit or otherwise change supplementary information, and minor (nonfactual) errors in these documents cannot be corrected after publication. Please submit document(s) exactly as you want them to appear, with all text, images, legends and references in the desired order, and check carefully for errors.

Journal: Nature Medicine

Article Title:	**Substantial inter-individual and limited intra-individual genomic diversity among tumors from men with metastatic prostate cancer**
Corresponding Author:	**Peter S. Nelson**

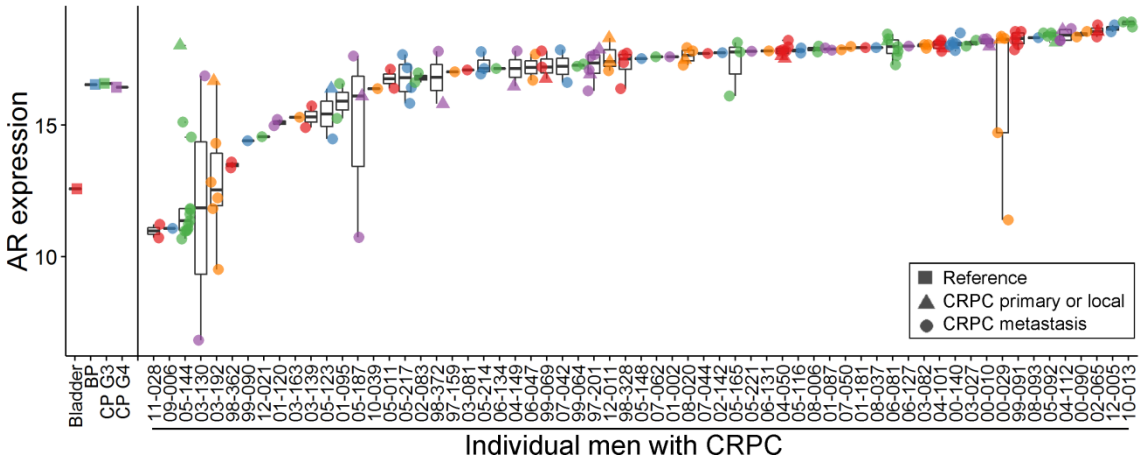
Supplementary Item & Number **(add rows as necessary)**	Title or Caption
Supplementary Figure 1	Genomic aberrations in metastatic CRPC
Supplementary Figure 2	Androgen receptor and neuroendocrine gene expression in CRPC metastasis.
Supplementary Figure 3	Assessments of gene expression pathways and putative somatic driver aberrations are consistent across metastasis within an individual: Patient 04-101.
Supplementary Figure 4	Assessments of gene expression pathways and putative somatic driver aberrations are consistent across metastasis within an individual: Patient 05-144.
Supplementary Figure 5	Unsupervised clustering of tumors based on gene expression or nonsynonymous mutations and relationships between cell proliferation, genomic aberrations, and AR.
Supplementary Figure 6	Fanconi Anemia complex genes, E2F1, and prostate cancer proliferation.
Supplementary Table 1	CRPC Patient Clinical Data
Supplementary Table 2	Patients, Tumor Sites, and Molecular Profiling Assays.
Supplementary Table 3	TMPRSS2-ERG Fusion by CGH or FISH
Supplementary Table 4	Target sequences for siRNAs
Supplementary Table 5	Primer sequences for QRT-PCR

Supplementary Figure 1

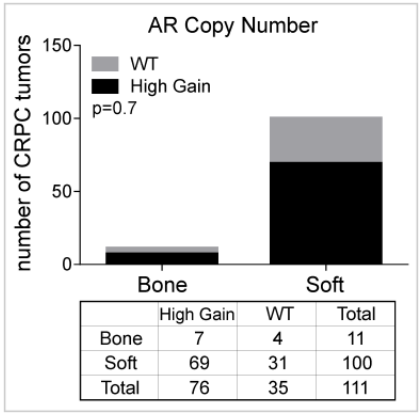


Supplementary Figure 2

a

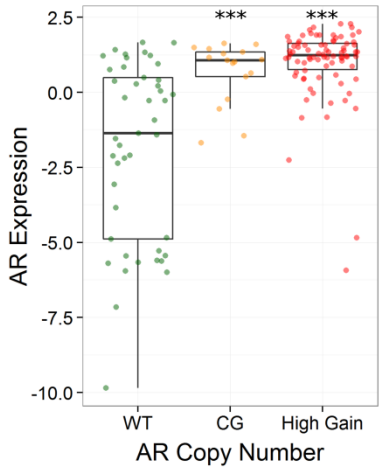


b

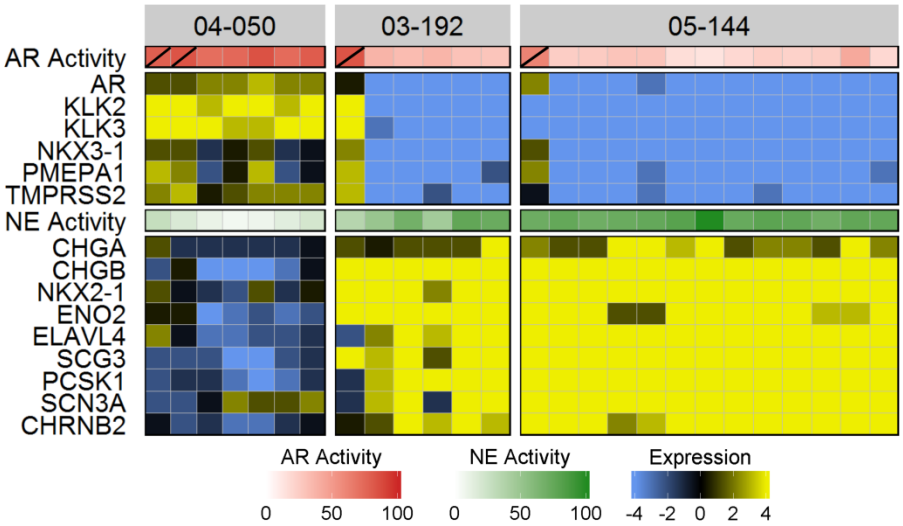


	High Gain	WT	Total
Bone	7	4	11
Soft	69	31	100
Total	76	35	111

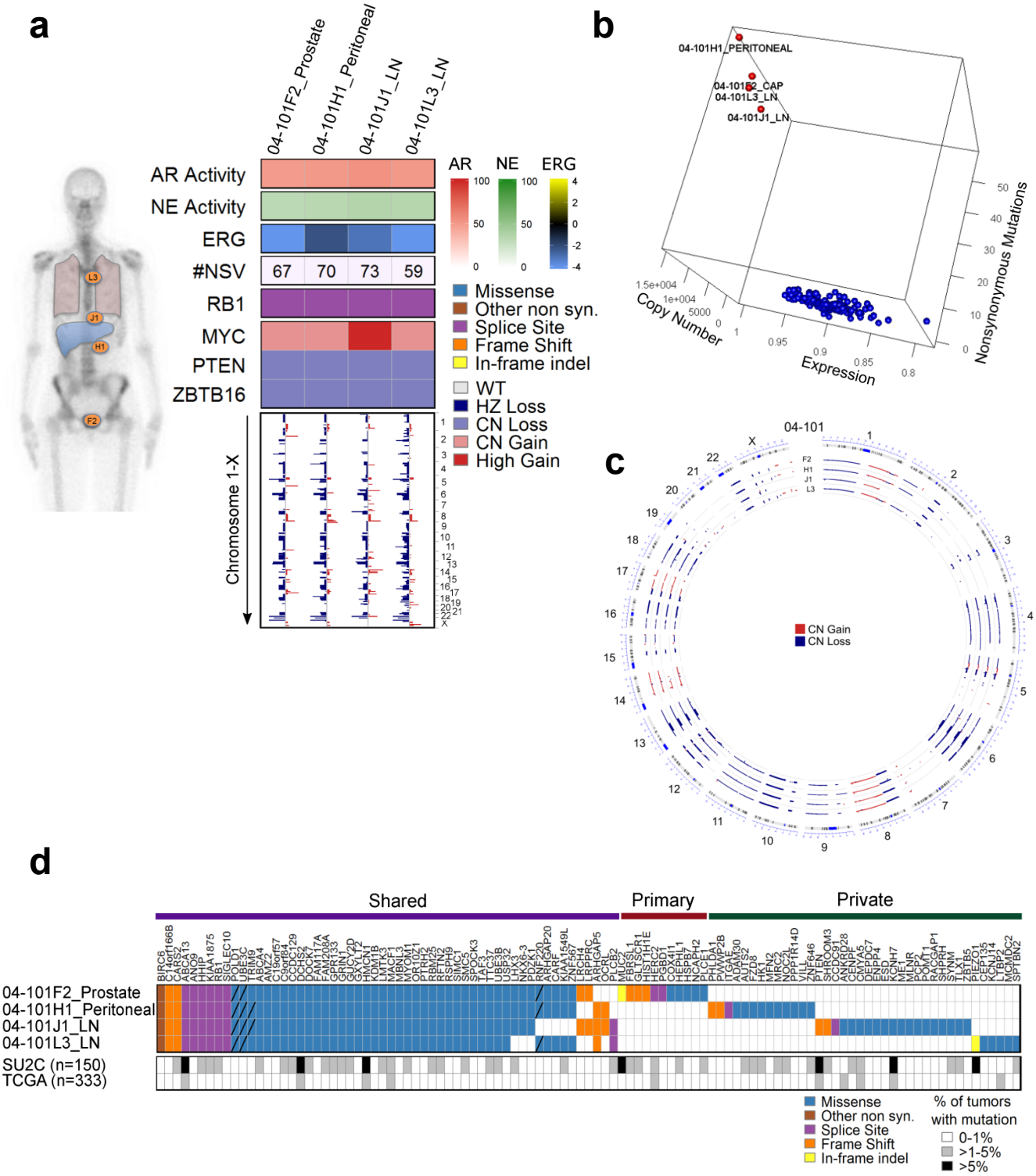
c



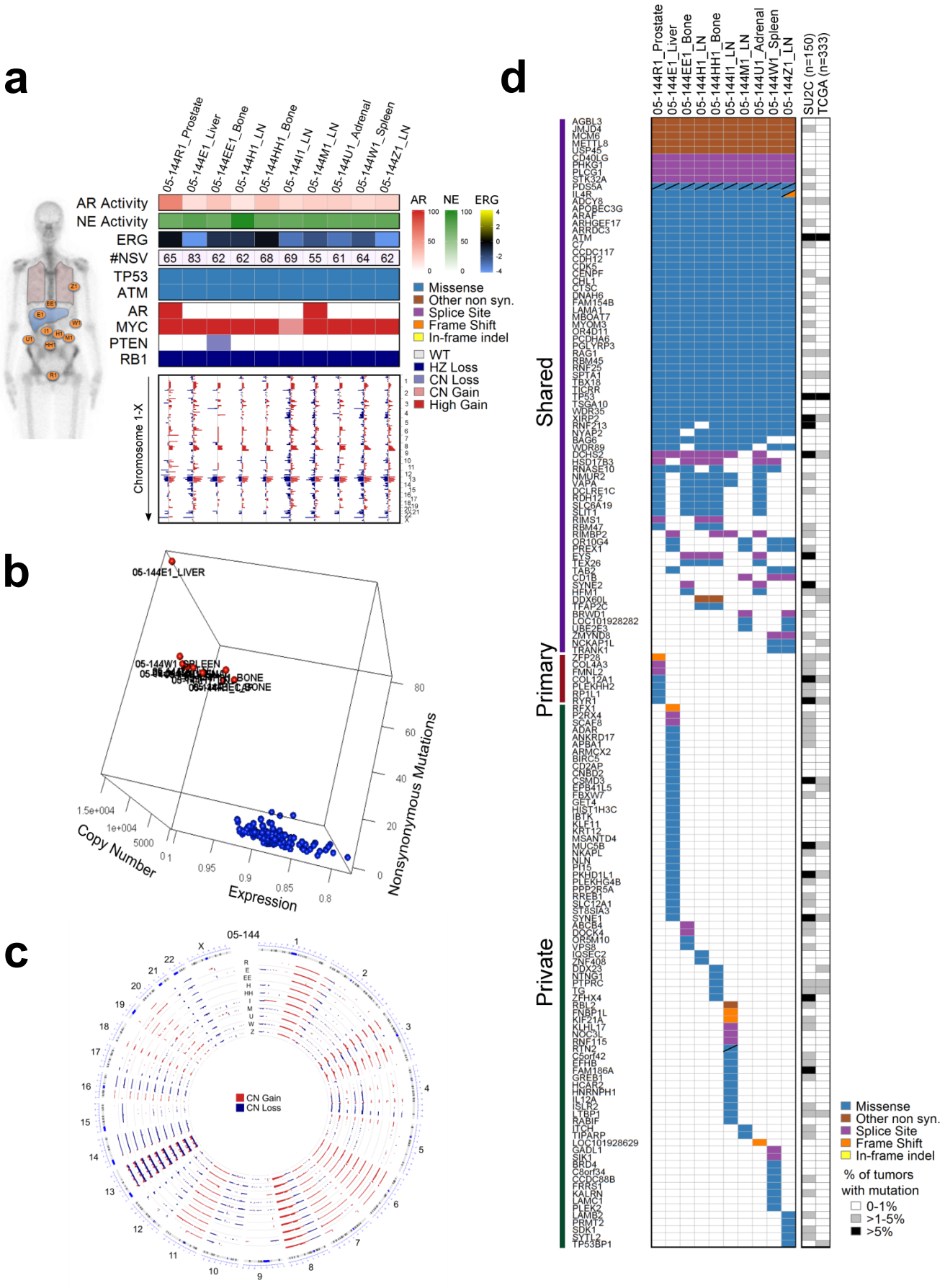
d



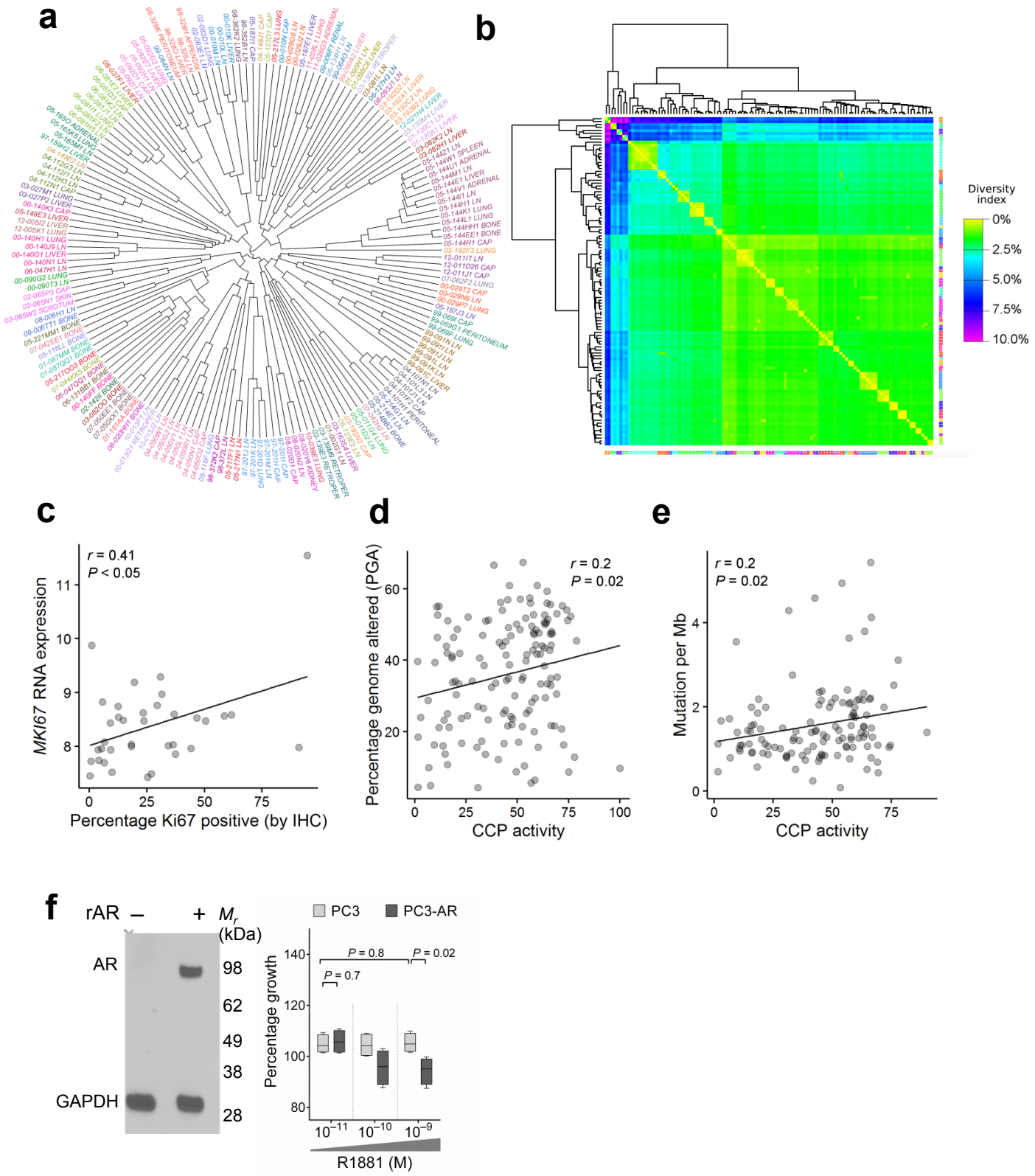
Supplementary Figure 3



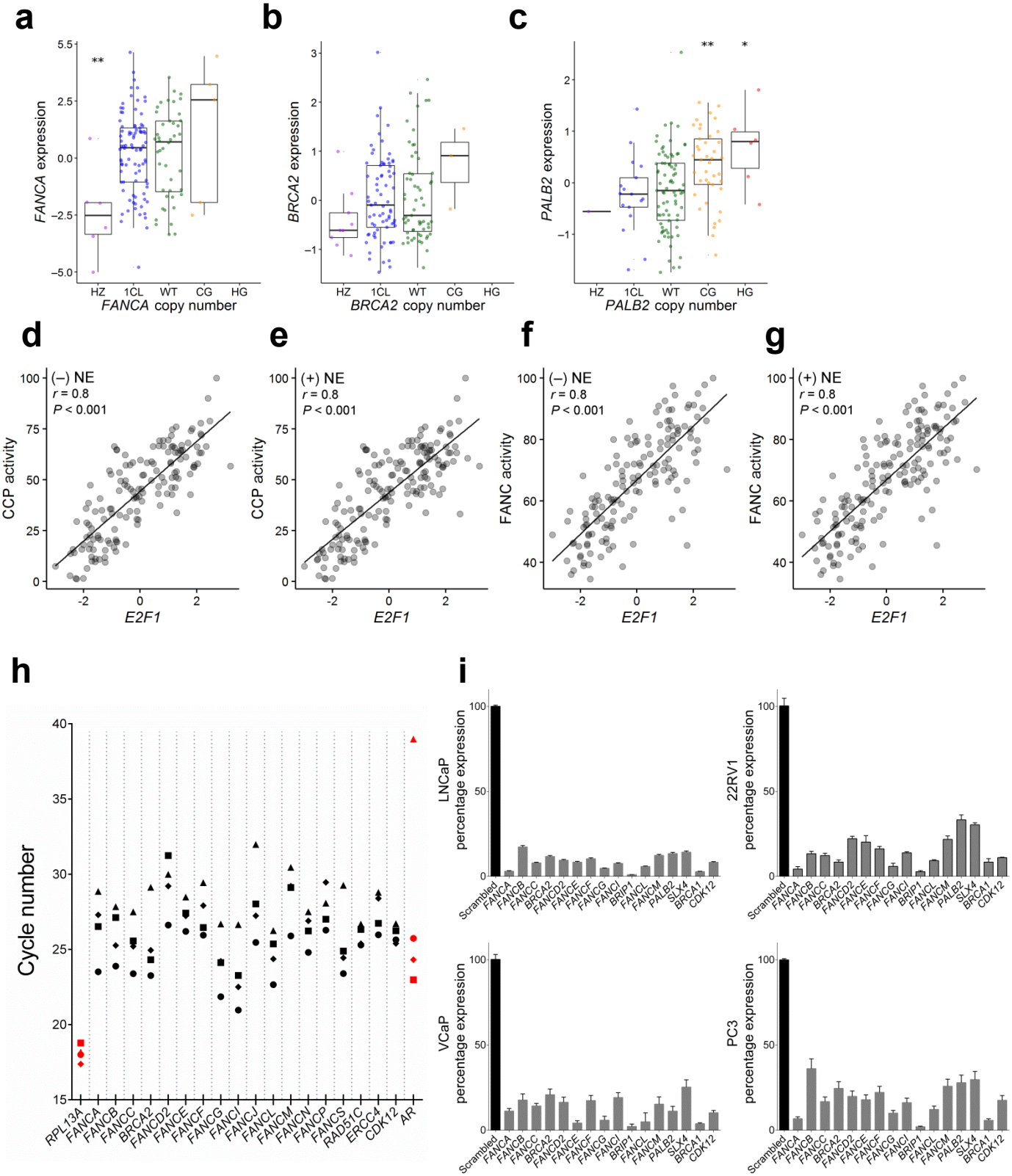
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6



Kumar et al NMED A76952

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. **Genomic aberrations in metastatic CRPC**.

a. Physical map of the focal somatic deletion on chr11 in three patients exhibiting ZBTB16/PLZF homozygous deletion. One representative tumor (03-192A1) is shown. The tumor vs. normal ratios of individual CGH probes in the region are plotted with horizontal bars indicating segmented copy number values. Genomic regions with a single copy loss are indicated by a single blue bar and double thickness blue bars represent regions with homozygous loss

b. Mutations in FOXA1 and FOXA2. Shown at the top is a schematic of the FOXA1 genes with seven nonsynonymous mutations identified in prostate tumors and cell lines by Grasso et al. shown in grey and those found in three patients in this study shown in black. The FOXA2 gene is shown in the bottom panel with nonsynonymous somatic mutations identified in three patients from this study clustered in the Hepatocyte Nuclear Factor 3 C terminal domain (HNF_C) which includes an engrailed homology 1 (EH1) domain known to exhibit repressor activity. FH=Forkhead Winged Helix DNA binding domain.

c. Intra- and inter-individual relationships of somatic mutation frequency in tumors from men with CRPC. Nonsynonymous mutations: individual patients are ordered on the x-axis by the median mutations/MB with each data-point representing an individual tumor. Circles are metastases and triangles are primary tumors. The 141 tumors from 56 patients with exome sequencing data are shown.

d. Intra- and inter-individual relationships of genomic copy number aberration burden in tumors from men with CRPC. Individual patients are ordered on the x-axis by the median percentage of the genome altered (PGA) as determined by array CGH with each data-point representing an individual tumor. Circles are metastases and triangles are primary tumors. The 149 tumors from 60 patients with CN data are shown.

e. The number of nonsynonymous mutations (NSV) in each tumor were plotted by metastatic site. There were no significant differences between bone and soft tissue by paired t-test (p=0.6) or paired Wilcoxon rank sum test (p=1).

f. The percentage of the genome altered (PGA) in each tumor was plotted by metastatic site. There were no significant differences between bone and soft tissue by paired t-test (p=0.6) or paired Wilcoxon rank sum test (p=0.4).

g. The proportion of tumors with homozygous CN loss in PTEN were compared by metastatic site. There were no significant differences between bone and soft tissue by Fisher's Exact test (p=0.2).

Supplementary Figure 2. **Androgen receptor and neuroendocrine gene expression in CRPC metastasis.**

a. AR expression in metastatic tumors within and between individuals. Patients are ordered on the x-axis by the mean level of AR expression, shown on the y-axis, measured in all tumors from a given individual. Each data-point represents an individual tumor. Circles = metastases. Triangles = primary tumors.

b. The proportion of tumors with high CN gain in AR were compared by metastatic site. There were no significant differences between bone and soft tissue by Fisher's Exact test (p=0.7).

c. Higher AR transcript expression is associated with gain and high gain of AR copy number (three asterisks indicates p<0.0005 by pairwise t-test and Wilcoxon rank sum test to WT group). WT, wild-type, CG, one copy gain, High Gain, more than one copy gain.

d. Tumors from one patient with conventional adenocarcinoma histology and two patients with small cell/neuroendocrine histology with expression of the AR and AR-regulated genes and genes associated with neuroendocrine differentiation shown in the heatmaps. Hatched boxes for AR-activity denote the primary prostate cancers from these patients removed at the time of rapid autopsy.

Supplementary Figure 3 Assessments of gene expression pathways and putative somatic driver aberrations are consistent across metastasis within an individual: Patient 04-101.

a. Shown are selected features from four tumors including one primary tumor and three metastases from patient 04-101. Activity scores comprise transcripts of genes regulated by the AR (AR activity score), genes expressed in neuroendocrine carcinoma (NE activity score). ERG transcript levels are relative to the mean-centered ratio for the cohort. NSV is the number of somatic non-synonymous nucleotide variants identified. Genome wide copy number losses (blue) and gains (red) are shown for each tumor ordered by chromosome.

b. Relationships of tumors derived from a single patient (red points) relative to tumors from all other individuals (blue points) based on a calculated a sample similarity score that comprised single nucleotide alterations, copy number alterations, and gene expression.

c. Relationships of genomic copy gain and loss across the genome denoted by chromosome number depicted by a Circos plot. Red = copy gain and Blue = copy loss.

d. Distribution of all 105 nonsynonymous point mutations and indels identified in the four tumors from patient 04-101. The table indicates the presence of a mutation (blue/tan/purple/orange/yellow) or its absence (white) in each tumor. The color bars above the table indicate whether the mutations were ubiquitous, shared by primary-tumor regions, shared by metastatic sites, or unique to a given tumor (private). The bottom panel denotes the frequency of somatic mutations found in whole exome sequencing studies of 333 primary prostate cancers (TCGA) or 150 metastatic prostate cancers (SU2C).

Supplementary Figure 4 Assessments of gene expression pathways and putative somatic driver aberrations are consistent across metastasis within an individual: Patient 05-144.

a. Shown are selected features from ten tumors including one primary tumor and nine metastases from patient 05-144. Activity scores comprise transcripts of genes regulated by the AR (AR activity score), genes expressed in neuroendocrine carcinoma (NE activity score). ERG transcript levels are relative to the mean-centered ratio for the cohort. NSV is the number of somatic non-synonymous nucleotide variants identified. Genome wide copy number losses (blue) and gains (red) are shown for each tumor ordered by chromosome.

b. Relationships of tumors derived from a single patient (red points) relative to tumors from all other individuals (blue points) based on a calculated a sample similarity score that comprised single nucleotide alterations, copy number alterations, and gene expression.

c. Relationships of genomic copy aberrations across the genome denoted by chromosome number depicted by a Circos plot. Red = copy gain and Blue = copy loss.

d. Distribution of all 156 nonsynonymous point mutations and indels identified in the four tumors from patient 05-144. The table indicates the presence of a mutation (blue/tan/purple/orange/yellow) or its absence (white) in each tumor. The color bars above the table indicate whether the mutations were ubiquitous, shared by primary-tumor regions, shared by metastatic sites, or unique to a given tumor (private). The bottom panel denotes the frequency of somatic mutations found in whole exome sequencing studies of 333 primary prostate cancers (TCGA) or 150 metastatic prostate cancers (SU2C).

Supplementary Figure 5. **Unsupervised clustering of tumors based on gene expression or nonsynonymous mutations and relationships between cell proliferation, genomic aberrations, and AR.**

a. Gene expression profiles comprising the most variable 5000 transcripts across the cohort were used to cluster tumors. Tumors from the same individual are given the same color and the tumor site is annotated to the individual patient: CAP, primary tumor; LN, lymph node. Note, bone metastasis are primarily grouped by site rather than patient of origin.

b. Heatmap of the percent of the total nonsynonymous mutations in the entire cohort of tumors that differ in a pairwise analyses of all tumors. The tumors are sorted based on complete linkage dendrograms. The most diverse tumors (blue and magenta) are hypermutated.

c. Correlation between Ki67 protein expression as assessed by IHC and Ki67 transcript expression ($r=0.41$, $p<0.05$).

d. Cell cycle progression scores are associated with the burden of genome copy losses and gains reflected by the percentage of genome altered (PGA). Each datapoint represents an individual tumor.

e. Cell cycle progression scores are associated with the burden of non-synonymous mutations. Each datapoint represents an individual tumor. Hypermutated tumors are not included.

f. AR expression level influences cellular responses to the AR ligand R1881. Western blot of AR and GAPDH protein levels in wild type PC3 cells and PC3-AR cells engineered to overexpress AR (rAR) (left panel). PC3 and PC3-AR cell growth measured 72 hours after exposure to the indicated R1881 concentration (right panel).

Supplementary Figure 6. **Fanconi Anemia complex genes, E2F1, and prostate cancer proliferation**.

a-c. Association of Fanconi anemia complex gene transcripts with genomic copy number. Each point represents an individual tumor. Purple, homozygous loss (HZ); Blue, heterozygous loss (1CL), Green, Diploid (WT); Yellow, single copy gain (CG); Red, high copy gain (HG).

a. Lower FANCA expression is associated with homozygous copy loss (two asterisks indicates $p<0.006$ by pairwise t-test and Wilcoxon rank sum test to WT group.)

b. BRCA2 expression by copy number. No comparisons to WT group were significant.

c. Higher PALB2 expression is associated with copy number gain (one asterisk indicates $p<0.05$ and two asterisks indicate $p<0.005$ by pairwise t-test and Wilcoxon rank sum test to WT group.)

d-g. Associations between E2F1 expression and the tumor proliferation rate as determined by the cell cycle progression score and E2F1 expression and a composite measure of FANC activi-

ty determined by the expression of 11 Fanconi Anemia complex genes. Each datapoint represents a distinct tumor. E2F1 expression is the mean-centered ratio of E2F1 transcripts across the cohort. Plots labeled (+NE) include all tumors. Plots labeled (-NE) have the tumors with high expression of NE markers removed from the analysis. NE, neuroendocrine carcinoma; FANC, Fanconi Anemia Complex. Pearson's correlation coefficient (r) and p-value (p) shown on individual plots.

h. Quantitation of FA gene transcript expression by qRT-PCR in four prostate cancer cell lines: triangle = PC3; circle = LNCaP; diamond= 22RV1; square=VCaP. RPL13A and AR are used as reference control transcripts. Transcripts from all genes in all cell lines were measured above background excepting AR in PC3 cells.

i. Assessment of FA gene expression knockdown by siRNA in four prostate cancer cell lines. Percent expression (y-axis) is the residual transcript level, relative to the level in scrambled siRNA control cells determined by qRT-PCR 3 days after the introduction of siRNAs targeting each FA gene (x-axis) or scrambled control.

Supplementary Table 1. CRPC Patient Clinical Data. Clinical data including age, Androgen Deprivation Therapy (ADT), Secondary AR Pathway Antagonism (ARPA), chemotherapy, and PSA are listed for the 63 patients in the cohort.

Supplementary Table 2. Patients, Tumor Sites, and Molecular Profiling Assays.

Supplementary Table 3. TMPRSS2-ERG Fusion by CGH or FISH. The number of tumors for each patient with CGH and/or FISH data are listed, along with the number determined to be either fusion positive (T2E+) or negative (T2E-). There was 100 percent concordance in all but two patients.

Supplemental Table 4: Target sequences for siRNAs

Supplemental Table 5: Primer sequences for QRT-PCR