Supplementary Information for

Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer.

Christopher E. Barbieri, Sylvan C. Baca, Michael S. Lawrence, Francesca Demichelis, Mirjam Blattner, Jean-Philippe Theurillat, Thomas A. White, Petar Stojanov, Eliezer Van Allen, Nicolas Stransky, Elizabeth Nickerson, Sung-Suk Chae, Gunther Boysen, Daniel Auclair, Robert Onofrio, Kyung Park, Naoki Kitabayashi, Theresa Y. MacDonald, Karen Sheikh, Terry Vuong, Candace Guiducci, Kristian Cibulskis, Andrey Sivachenko, Scott L. Carter, Gordon Saksena, Douglas Voet, Wasay M. Hussain, Alex H. Ramos, Wendy Winckler, Michelle C. Redman, Kristin Ardlie, Ashutosh K. Tewari, Juan Miguel Mosquera, Niels Rupp, Peter J. Wild, Holger Moch, Colm Morrissey, Peter S. Nelson, Philip W. Kantoff, Stacey B. Gabriel, Todd R. Golub, Matthew Meyerson, Eric S. Lander, Gad Getz, Mark A. Rubin* and Levi A. Garraway*

* To whom correspondence should be addressed: rubinma@med.cornell.edu (M.A.R.); levi_garraway@dfci.harvard.edu (L.A.G.)

This PDF file includes:

Supplementary Note
Supplementary Figures 1 through 18
Supplementary References
Supplementary Tables 1, 4 through 8, 10 through 12

Other Supplementary Information for this manuscript includes the following:

Supplementary Tables 2, 3, 9 (.xls or .seg files)

Nature Genetics: doi:10.1038/ng.2279

Supplementary Note

Description of prostate tumor cohorts

Clinically localized primary prostate cancers were selected for exome- and transcriptome-sequencing from two cohorts: Weill Cornell Medical College (WCMC; New York, NY) (Esgueva *et al.*, in press) and Uropath (Perth, Australia), a commercial supplier of banked urological tissues. Participating patients had received no prior treatment for prostate cancer. We selected tumors based on suitability for sequencing, choosing specimens with cancerous foci relatively free from admixed normal tissue that yielded high-quality DNA.

Tumors from the WCMC cohort were collected by the Institutional Biobank from patients undergoing radical prostatectomy by one surgeon (A.K.T.) for clinically localized prostate cancer. Patient-matched normal DNA was obtained from whole blood samples for this cohort.

Tumors from the Uropath cohort were obtained from men undergoing radical prostatectomy for clinically localized prostate cancer across multiple medical centers in Western Australia between 2000 and 2010. Samples from both cohorts were stored at -80°C. Paired normal DNA was sequenced from benign prostate tissue with no histological evidence of neoplasia.

For both cohorts, Hematoxylin and Eosin (H&E)-stained tissue sections were centrally reviewed by J-M. M., K.P. and M.A.R. to verify Gleason score and to determine the percentage of Gleason pattern 4 and 5 histology at the site selected for DNA extraction. To characterize the ethnic composition of the cohorts, we analyzed high-density SNP array data by principal component analysis in combination with data from cohorts of known ethnicity from the HapMap database (CEU, YRI, CHB/JPT; http://hapmap.ncbi.nlm.nih.gov/) (Supplementary Fig. 16). All but five individuals chosen for exome sequencing clustered with CEU HapMap samples, indicating that patients were predominantly of European descent. Four samples showed mixed or undetermined ethnicity and one clustered clearly with CHB/JPT (Han Chinese in Beijing/Japanese in Tokyo) HapMap samples.

Prostate tumor cohorts from University of Michigan (UM), University of Washington (UW) and University Hospital Zurich (UZH) were used for extension screening for SPOP mutation. Prostate samples from the UM cohort were obtained from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program[54], University of Michigan Prostate Cancer Specialized Program of Research Excellence Tissue Core (Ann Arbor, MI). Tumors from the UW cohort were obtained from the Rapid Autopsy Program, University of Washington and Fred Hutchison Cancer Research Center

University (Seattle, WA). Samples from the UHZ cohort included a series of radical prostatectomy specimens, metastases, and benign prostatic hyperplasia samples. H&E–stained slides of all specimens were reevaluated by two experienced pathologists (P.J.W., H.M.) to identify representative areas. Tumor stage and Gleason score of the Zurich cohort were assigned according to the International Union Against Cancer and World Health Organization/International Society of Urological Pathology criteria.

All samples were collected with informed consent of the patients and prior approval of the institutional review boards (IRB) of respective institutions. Additionally, the sequencing and data release of all exome- and transcriptome-sequenced samples was reviewed and approved by local IRB.

Correlation of mutation rates and pathological features

We investigated whether pathological features corresponded to different mutational spectra. Pathologic stage pT3 tumors contained more mutations than pT2 tumors (P = 0.0012, rank sum test) despite equivalent tumor purity between these classes (Supplementary Figs. 17, 18). Substitutions in *PTEN* and *PIK3CA* were enriched in pT3 tumors (P = 0.011, Fisher's exact test) (Supplementary Table 11), suggesting that these mutations may play a role in disease progression. Consistent with this possibility, activation of the PI3-Kinase pathway in mouse models accelerates the progression of prostate cancer[61,62]. This finding will need to be extended to larger panels of tumors due to the relatively small number of *PTEN* and *PIK3CA* mutations reported here. Interestingly, the base mutation rate showed no correlation with Gleason score (a histological measure of disease risk) (Supplementary Fig. 18), indicating that mutational burden does not track uniformly with disease aggressiveness.

Distinct mutational characteristics of *TMPRSS2-ERG* fusion-positive tumors

We determined whether the mutational spectrum varied between prostate tumors harboring the *TMPRSS2-ERG* fusion and fusion-negative tumors. *TMPRSS2-ERG* fusion-positive tumors showed an increased proportion of CpG to T transitions (P = 0.0002, Supplementary Fig. 18) but did not harbor more mutations overall. Since CpG to T transitions can arise from deamination of methylcytosine in cancer, this trend may reflect the differential methylation of DNA between ETS fusion-positive and fusion-negative tumors that was recently reported[63] or may indicate a distinct mutagenic process in fusion-positive tumors.

Distinct genomic lesions co-occurred with the *TMPRSS2-ERG* fusion. *PTEN* mutation or deletion was significantly associated with *ERG* rearrangement (P = 0.00042) (Fig. 4), consistent with reports of collaboration between these two oncogenic events in mouse models[64,65]. Tumors with *ERG* rearrangement were also enriched for p53 lesions (P = 0.000025) (Fig. 4). This relationship is consistent with other primary prostate cancer datasets (P = 0.0044)[12,33].

Somatic copy number alteration at the *SPOP* locus

Previous reports indicate that SPOP may function as an oncogene based on genomic amplifications in other cancers[17] and protein overexpression in clear cell renal cell carcinomas[29]. However, minimal somatic copy number aberrations are seen in the *SPOP* locus in primary prostate cancers from multiple cohorts, with no evidence of deletions in tumors with *SPOP* mutations (Fig. 1, Supplementary Fig 10)[12]. In addition, *SPOP* mRNA expression was not up-regulated in prostate cancer samples from multiple cohorts, but was more frequently down-regulated compared to benign controls (Supplementary Fig. 11). We therefore conclude that the biological effect of *SPOP* mutations is not recapitulated by somatic copy number alterations in primary prostate cancer.

Mutual exclusivity of *SPOP* and *ERG* alteration within a single prostate

A single tumor was originally classified as positive for both *SPOP* mutation and *TMPRSS2-ERG* fusion. Interestingly, re-analysis of this case revealed two morphologically distinct tumors: one focus with wild-type *SPOP* and *ERG* rearrangement, and a second with the $SPOP^{F133V}$ mutation and no *ERG* rearrangement (Supplementary Fig. 13). Therefore, the mutual exclusivity of these events is recapitulated in two tumor foci from a single prostate.

Low-frequency mutations in cancer-associated genes

Multiple genes with established roles in other cancers were mutated at low frequency, including *IDH1*, *AKT1* and *HRAS* (Supplementary Table 4). An analysis of predicted “damaging” mutations (nonsense substitutions, frame-shift indels and splice site alterations) in genes expressed in prostate tumors identified mutations in *APC*, *PIK3R1* and *EPHA7* (Supplementary Table 4). In addition, several chromatin-modifying enzymes harbored low-frequency damaging mutations, including *MLL1*, *MLL2, MLL3*, *ARID1A, NCOR1* and the histone demethylase gene *KDM6A* (*UTX*). Two *KDM6A* mutations involved residues situated within the catalytic Jumonji domain (I1209 and G1212), while a third introduced a frame-shift deletion directly N-terminal to this region (Supplementary Table 4). These findings underscore the emerging importance of chromatin-modifying genes in prostate cancer[66]. Notably, *AR* was not mutated in any primary tumor analyzed, consistent with prior studies suggesting that mutations in this gene are restricted to metastatic or castration-resistant disease[7,12].

List of Tables and Figures

Supplementary Figure 1. Depth and breadth of exome sequencing coverage.

Supplementary Figure 2. Overlap of sample profiling across platforms.

Supplementary Figure 3. Rates of somatic substitutions in prostate exomes.

Supplementary Figure 4. Expression levels of select mutated genes.

Supplementary Figure 5. Laser capture microdissection and sequencing of *MED12*.

Supplementary Figure 6. Mutations in *SPOP* in RNA-seq data and Sanger sequencing of genomic tumor DNA.

Supplementary Figure 7. Multiple Sequence Alignment of the MATH domain of SPOP across species.

Supplementary Figure 8. Multiple independent siRNAs targeting *SPOP* have similar effects in prostate cell lines.

Supplementary Figure 9. Transfection with SPOP siRNA or SPOP mutant does not affect cell growth or viability.

Supplementary Figure 10. Minimal changes in *SPOP* copy number in primary prostate cancer.

Supplementary Figure 11. *SPOP* is not upregulated in prostate cancer.

Supplementary Figure 12. Tumors with *SPOP* mutation lack *ETS* rearrangements.

Supplementary Figure 13. Separate foci of prostate adenocarcinoma with mutually exclusive *ERG*-rearranged and *SPOP*-mutated status.

Supplementary Figure 14. Detection of *SPOP* mutation in high-grade prostatic intraepithelial neoplasia (HGPIN).

Supplementary Figure 15. Tumors with *SPOP* mutation lack *PTEN* deletion in primary but not metastatic prostate cancer.

Supplementary Figure 16. Ethnicity analysis of exome-sequenced DNA.

Supplementary Figure 17. Relative ability to detect mutations in subgroups of tumors.

Supplementary Figure 18. Mutational landscape across a spectrum of primary prostate cancers.

Supplementary Table 1. Clinical Characteristics of Exome-Sequenced Primary Prostate Cancers

Supplementary Table 2. Exome Sequencing Statistic Summaries

Supplementary Table 3. Somatic Mutations in 112 Primary Prostate Tumor-Normal Pairs

Supplementary Table 4. Mutations in Significantly-Mutated Genes and other Selected Genes

Supplementary Table 5. Significantly Mutated Gene Sets

Supplementary Table 6. Systematic Sequencing Studies Including *SPOP*

Supplementary Table 7. *SPOP* Mutations in Multiple Cohorts

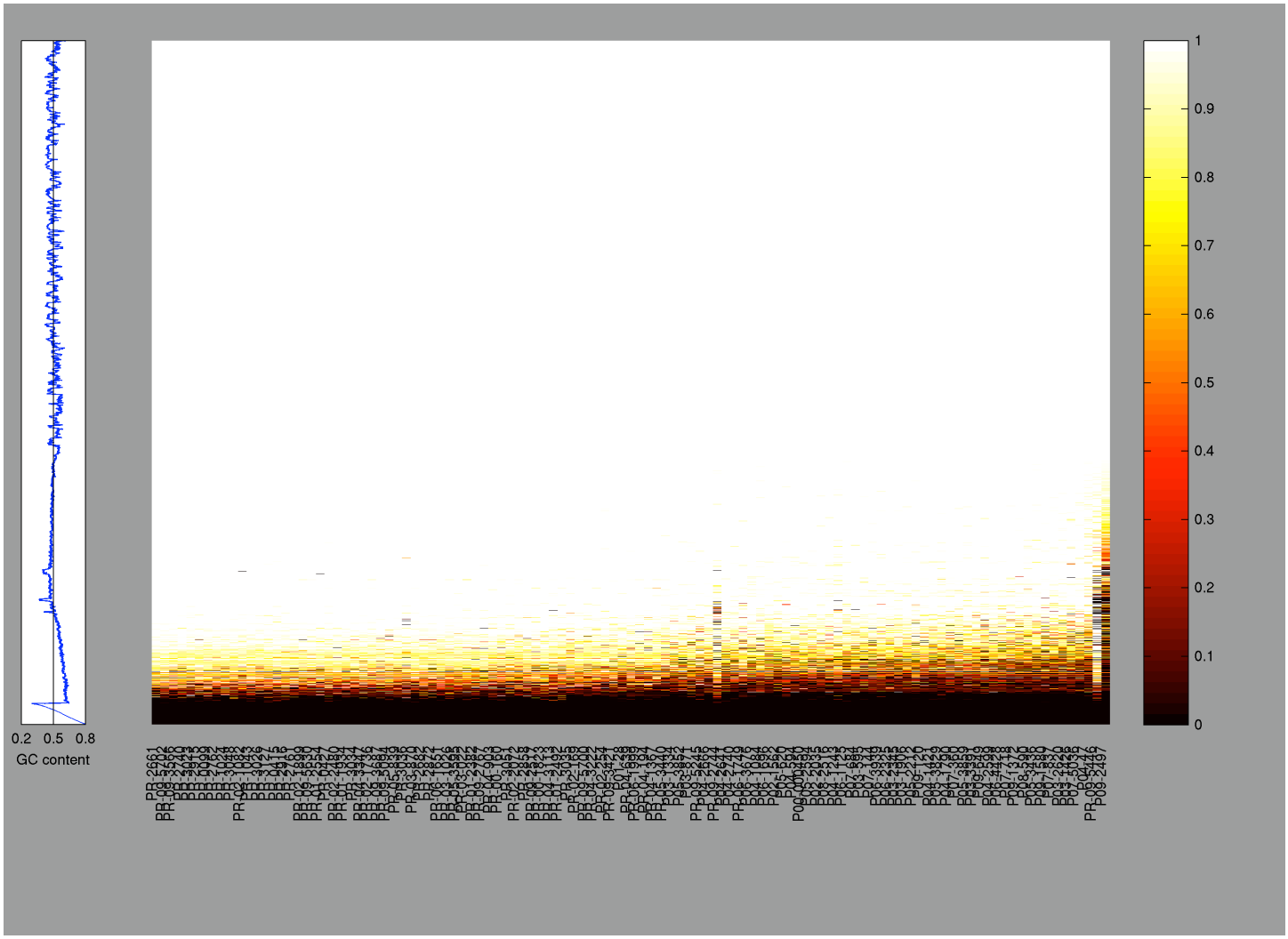
Supplementary Table 8. Somatic Copy Number Alterations Associated with *SPOP* Mutation

Supplementary Table 9. Primer Sequences

Supplementary Table 10. BAC Probes Used for FISH

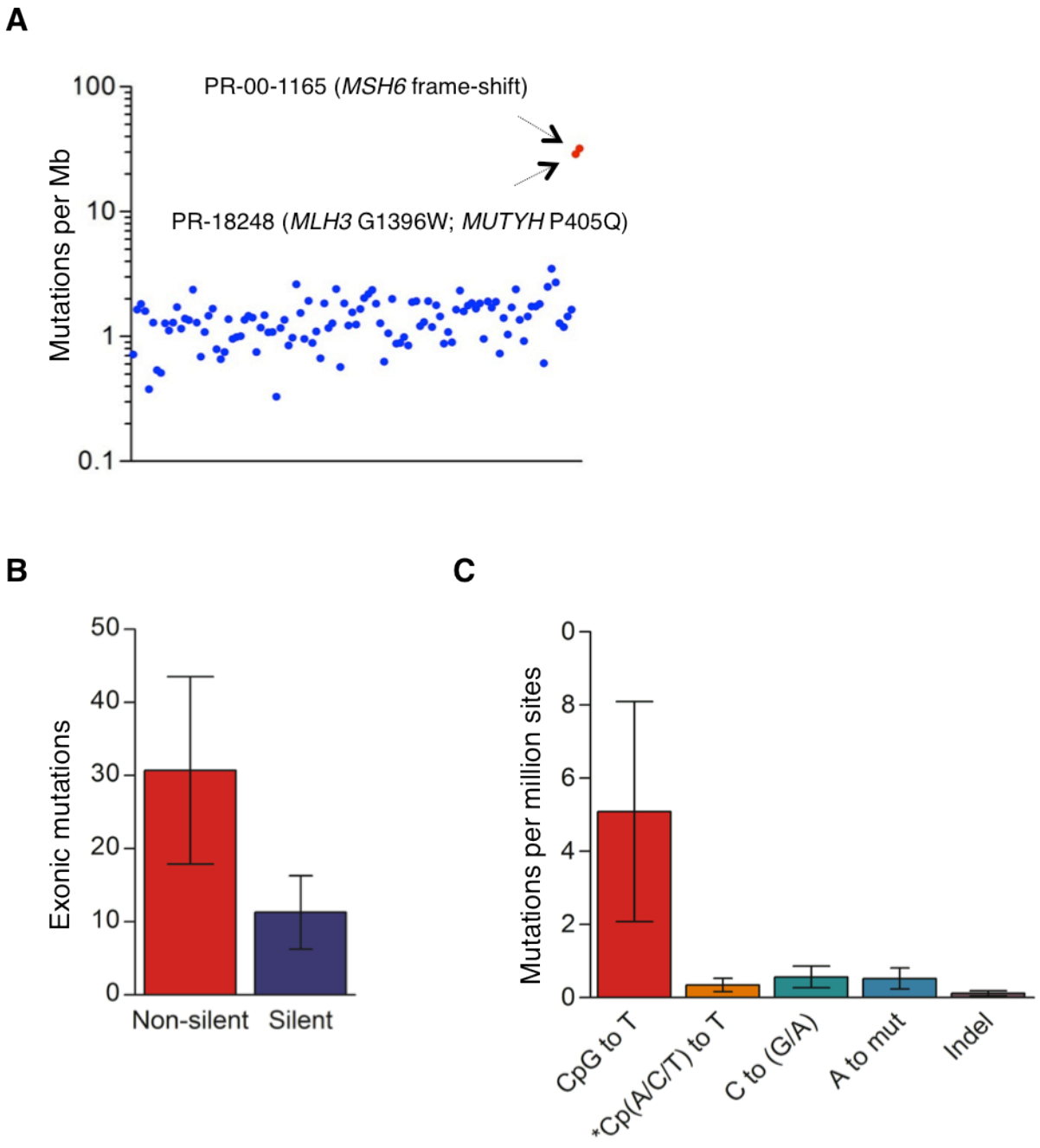
Supplementary Table 11. Mutation of *PIK3CA* and *PTEN* is Enriched in Locally Advanced Tumors

Supplementary Table 12. Copy Number Alterations in 112 Exome-Sequenced Prostate Tumors (seg file format)



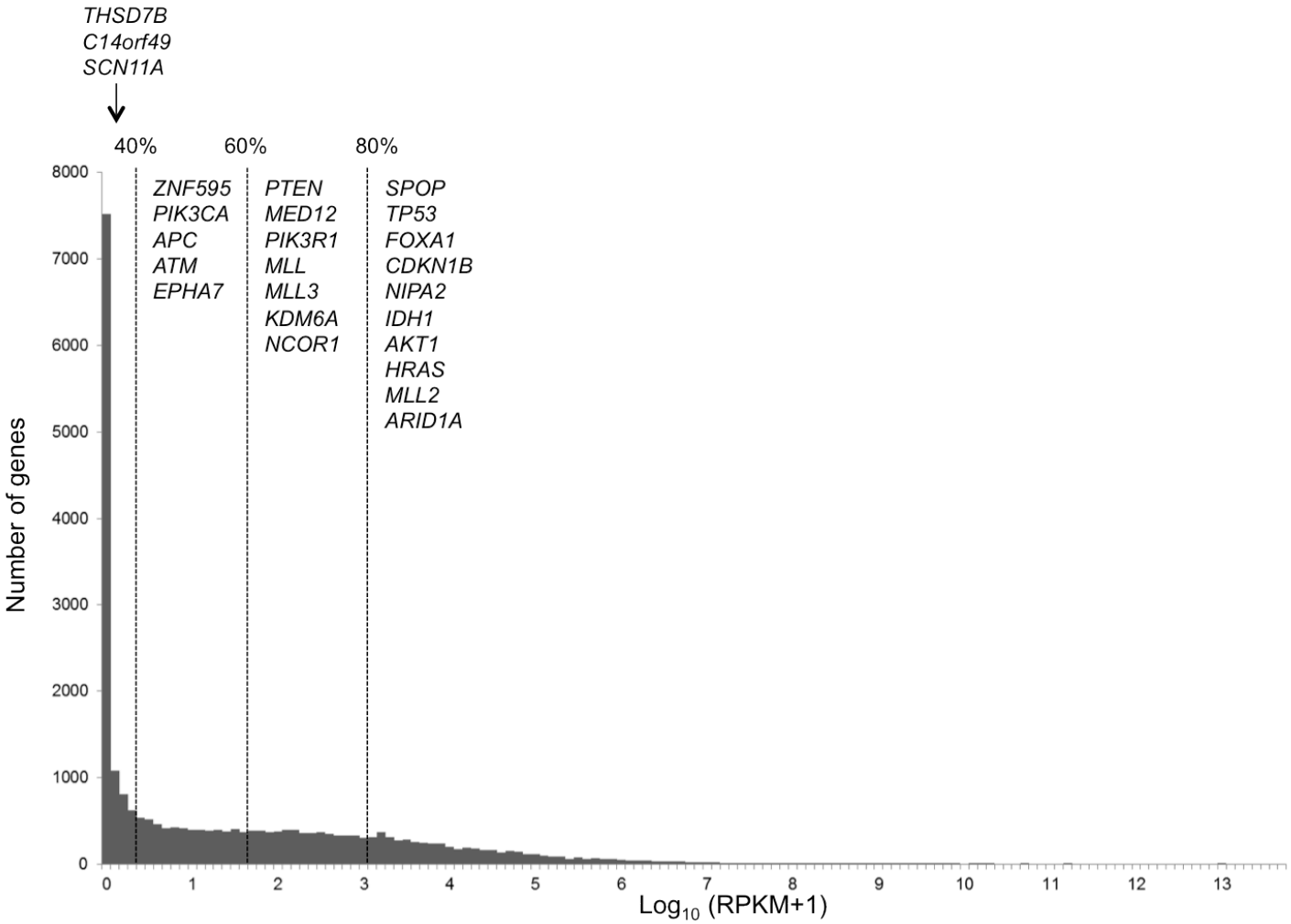
Supplementary Figure 1. Depth and Breadth of Exome Sequencing Coverage. (Center) Sequencing coverage across all sites targeted by hybrid capture. Each row represents a targeted exonic site; each column represents a tumor-normal pair. Coloring reflects the depth of sequencing coverage. White coloring indicates a minimum of 14 reads in the tumor and 8 reads in the normal. (Left) GC content across targeted sites (GC content is equal to the number of C or G nucleotides divided by the total number of nucleotides).

Supplementary Figure 2. Overlap of sample profiling across platforms. Exome sequencing was conducted on 112 tumor-normal pairs. A single highly-mutated tumor (PR-00-1165) was excluded from subsequent analyses, except where otherwise indicated, leaving 111 pairs. RNA-sequencing was performed on 22 of the exome-sequenced tumors and 41 independent tumors. All but four of the 112 exome-sequenced tumors, plus an additional 61 tumors, were analyzed for copy number alteration by high-density SNP array (169 total).

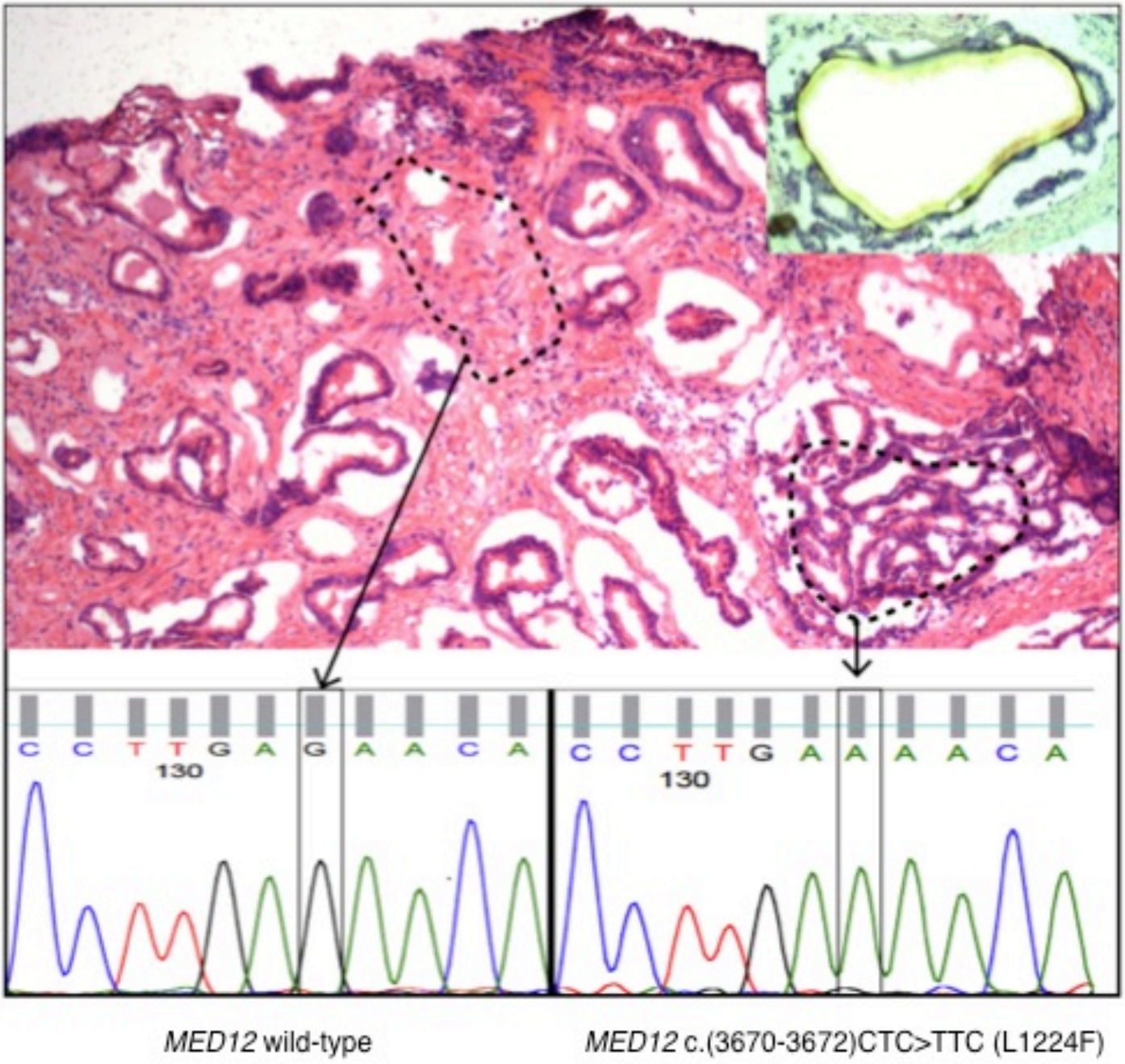


Supplementary Figure 3. Rates of somatic substitutions in prostate exomes. (A) Number of somatic mutations per Mb sequenced across the cohort of tumors. A single primary tumor (PR-00-1165) harbored a large excess of mutations compared to other tumors (32.1 per Mb versus 1.4 per Mb median in the remaining primary tumors, red). A prostate cancer metastasis sequenced but not reported here showed a similar extent of mutation (PR-18248; 29.0 mutations per Mb, red). The two highly mutated tumors contained the indicated alterations in DNA mismatch repair genes. (B) Median number of non-synonymous and synonymous mutations across 111 exomes (the single hyper-mutated primary tumor PR-00-1165, with 997 mutations, is excluded). (C) Mutations per million sites sequenced for the most frequent mutation categories in the dataset. *CpG to T, C to T transversion at a CpG dinucleotide; *CpG to A/C/T, C to T transversion not in the context of a CpG dinucleotide; C to (G/A), mutation of C to G or A; A to mut, mutation of A; Indel, small insertion or deletion. Error bars indicate standard deviation.

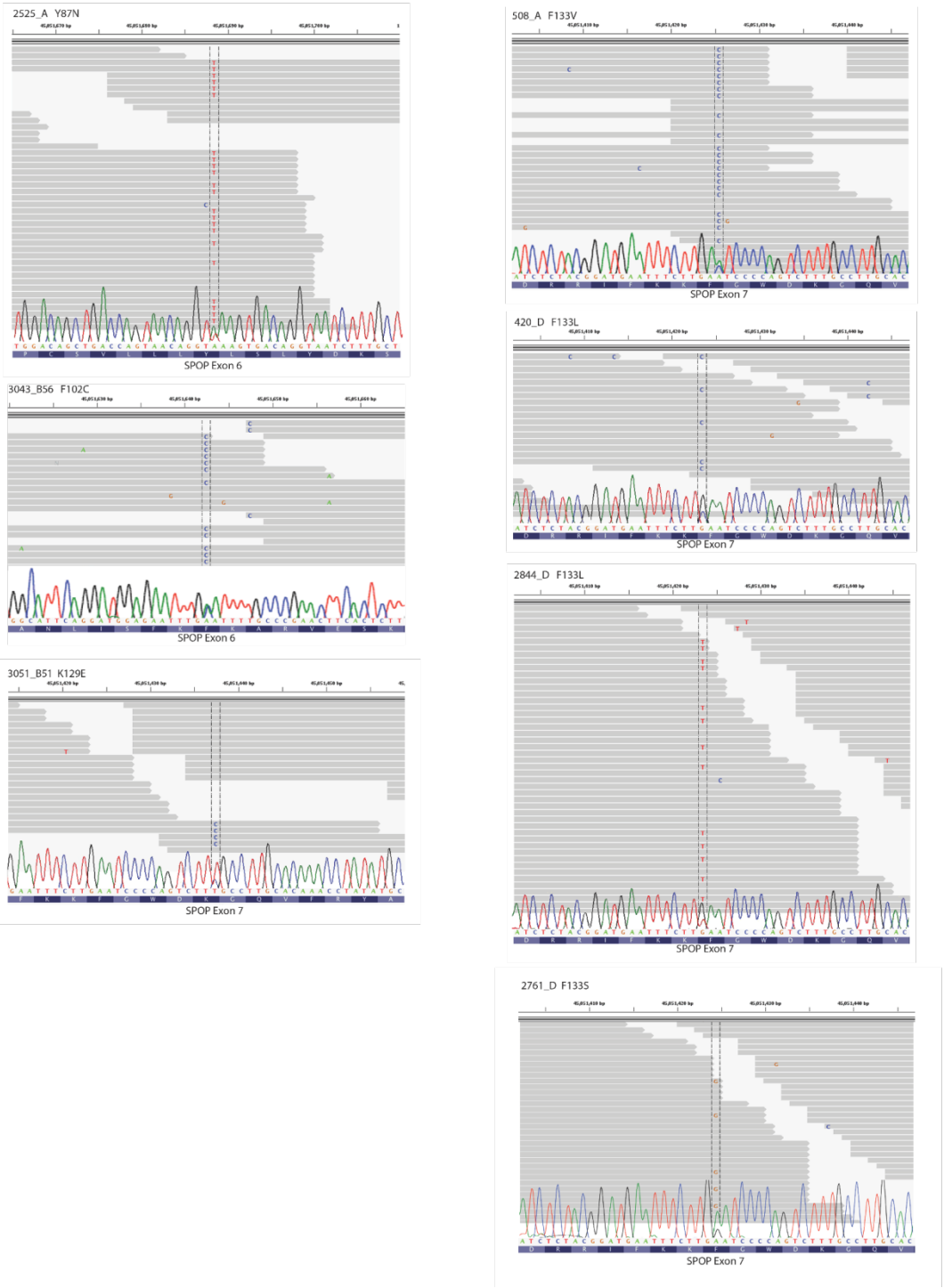
Nature Genetics: doi:10.1038/ng.2279



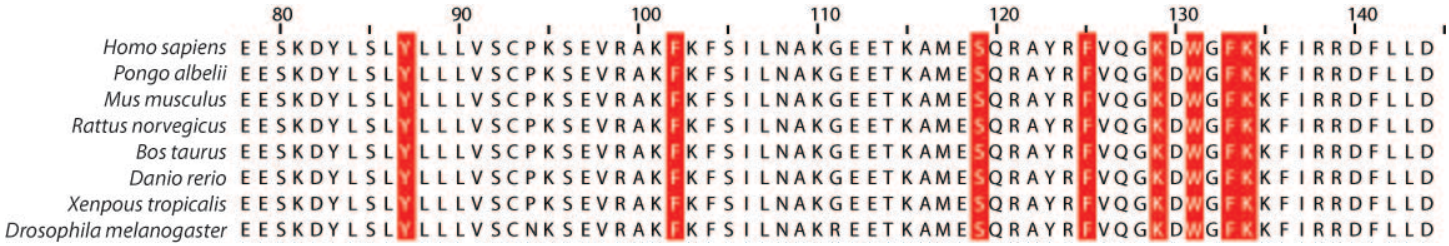
Supplementary Figure 4. Expression levels of select mutated genes. Significantly-mutated genes and selected genes listed in Supplementary Table 4 were analyzed for level of transcript expression in the RNA-seq dataset. The histogram shows the number of transcripts with a given value of $\log_{10}$ (RPKM+1) (where RPKM is the number of reads per kilobase of exon per million mapped sequence reads), binned by increments of 0.1. The RPKM provides an estimate of the relative expression of transcripts. Vertical lines indicate the percentile of $\log_{10}$(RPKM+1) among all transcripts. Listed genes are grouped based on their percentile of $\log_{10}$(RPKM+1) value: <40%, 40-60%, 60-80% and >80%. Values and percentiles are listed in Supplementary Table 4.



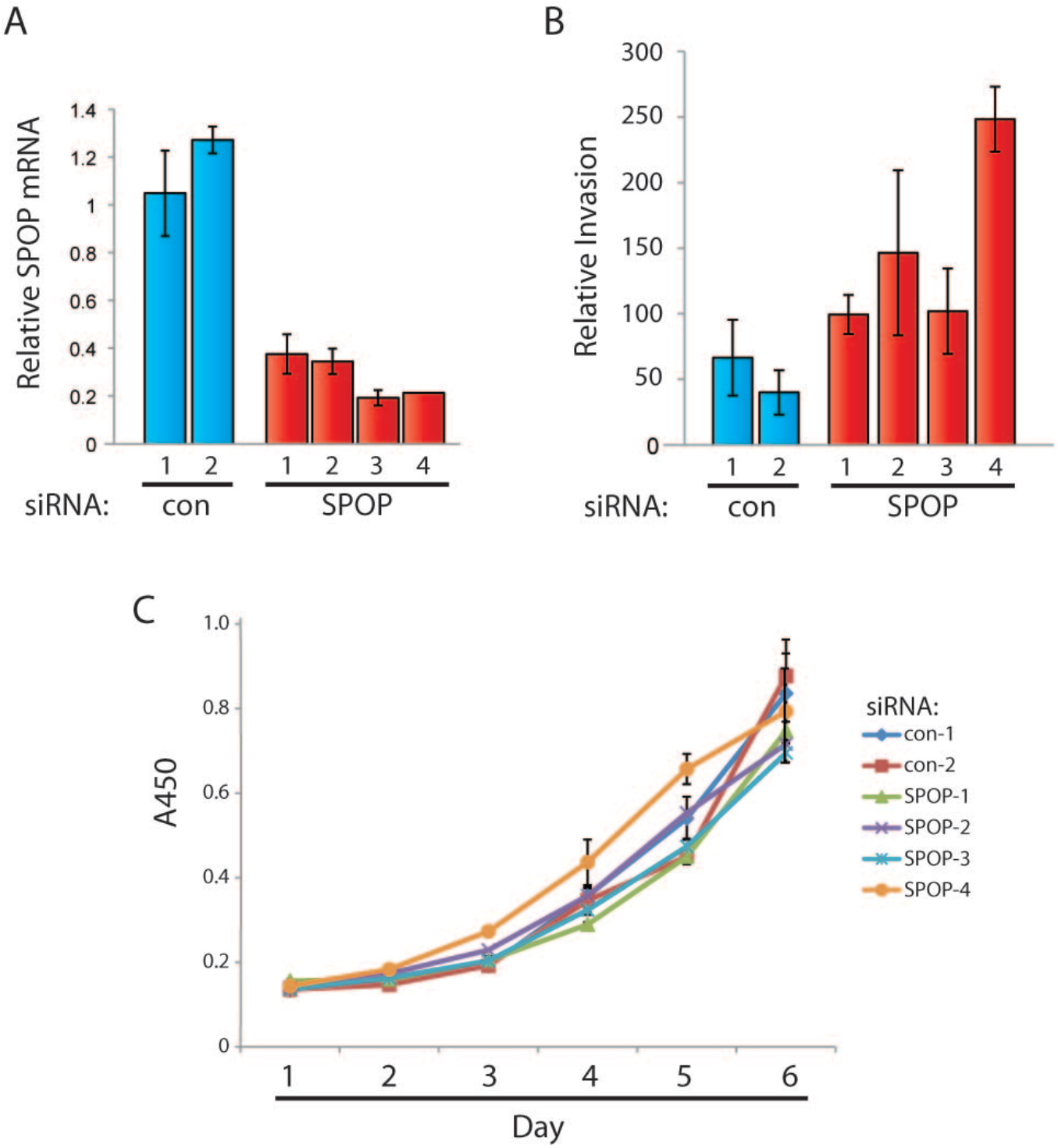
Supplementary Figure 5. Laser capture microdissection and sequencing of *MED12*. (Top) Laser capture micro-dissection and Sanger sequencing was performed on *MED12*-mutant tumors to determine whether the mutations were present in epithelial or stromal cells. H&E slide of frozen tissue from a *MED12*-mutant tumor (PR-3026) showing adenocarcinoma and surrounding mixed stroma. Exome sequence reads demonstrated an L1224F mutation in exon 26 of *MED12*. Laser capture micro-dissection was performed to separate epithelium from stroma (inset). (Bottom) The selected stromal area (dashed line, left) demonstrates wild-type *MED12* sequence by Sanger sequencing, while the dissected tumor gland (dashed line, right) exhibits the L1224F mutation.



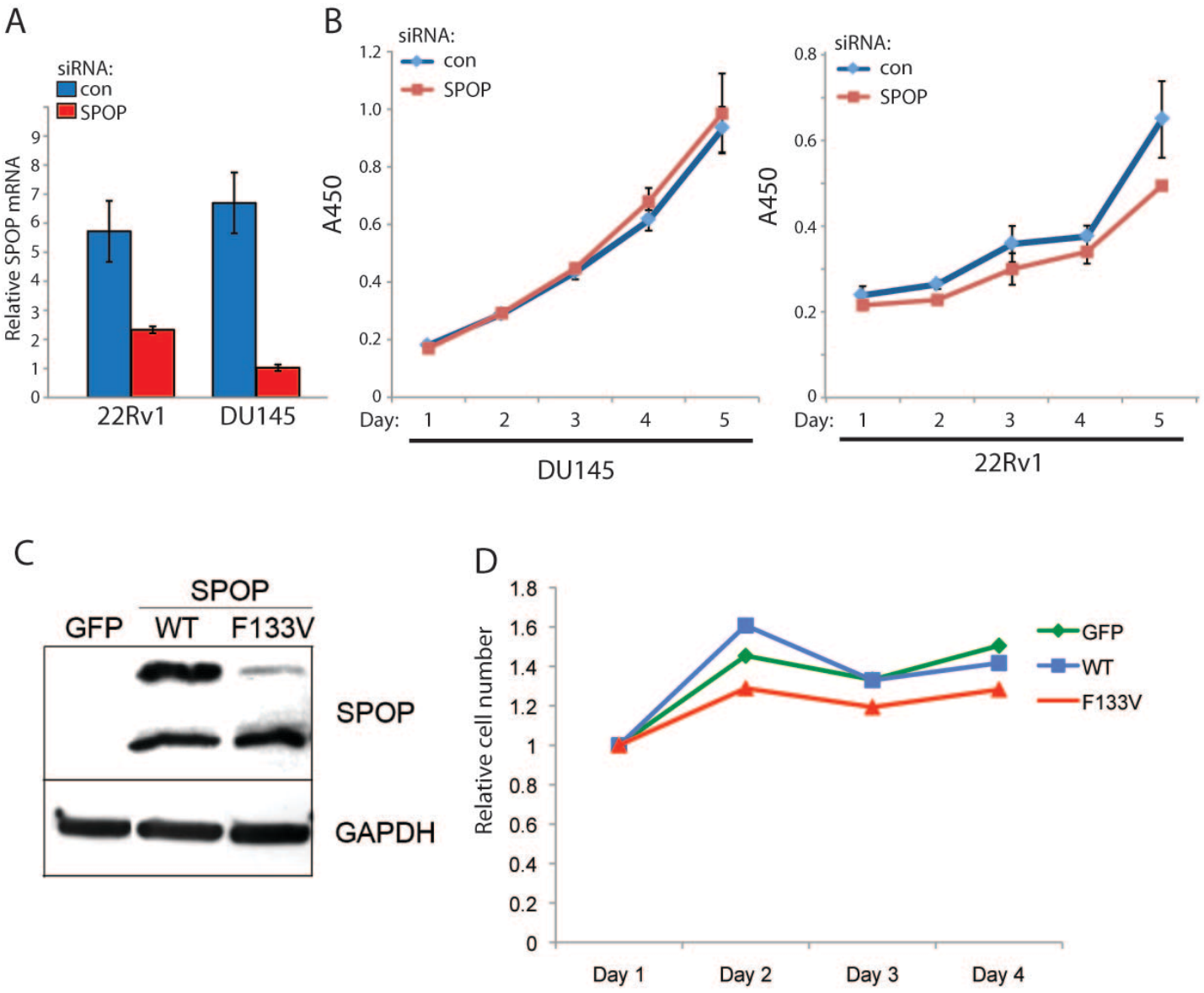
Supplementary Figure 6. Mutations in *SPOP* in RNA-seq data and Sanger sequencing of genomic tumor DNA. In each panel, RNA-seq reads mapping to SPOP Exon 6 or 7 from the indicated sample are shown. Coordinates (hg18) on chromosome 17 are at the top of each panel, and the reference genome (hg18) and wild-type SPOP amino acids are displayed at the bottom. Each horizontal gray bar represents one read. Nucleotide mismatches with respect to the reference genome in each read are highlighted by displaying the mismatched base. The Sanger tracing of genomic DNA from the same tumor focus is overlayed below the RNA-seq reads.



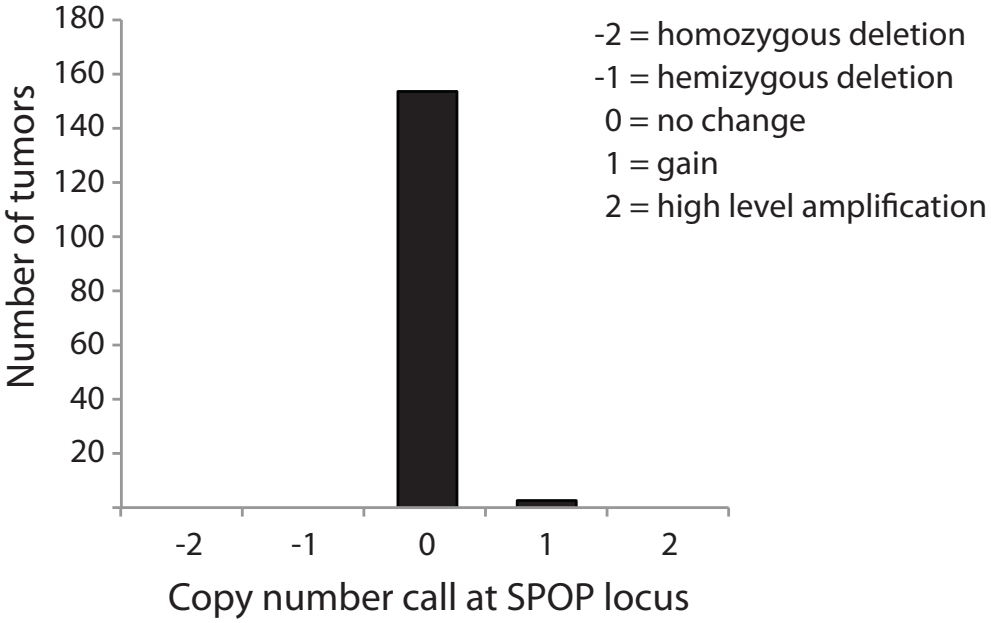
Supplementary Figure 7. Multiple Sequence Alignment of the MATH domain of SPOP across species. Multiple sequence alignment was performed with ClustalW2 and visualized using Jalview. Residues mutated in prostate cancer (Y87, F102, S119, F125, K129, W131, F133, K134) are highlighted.



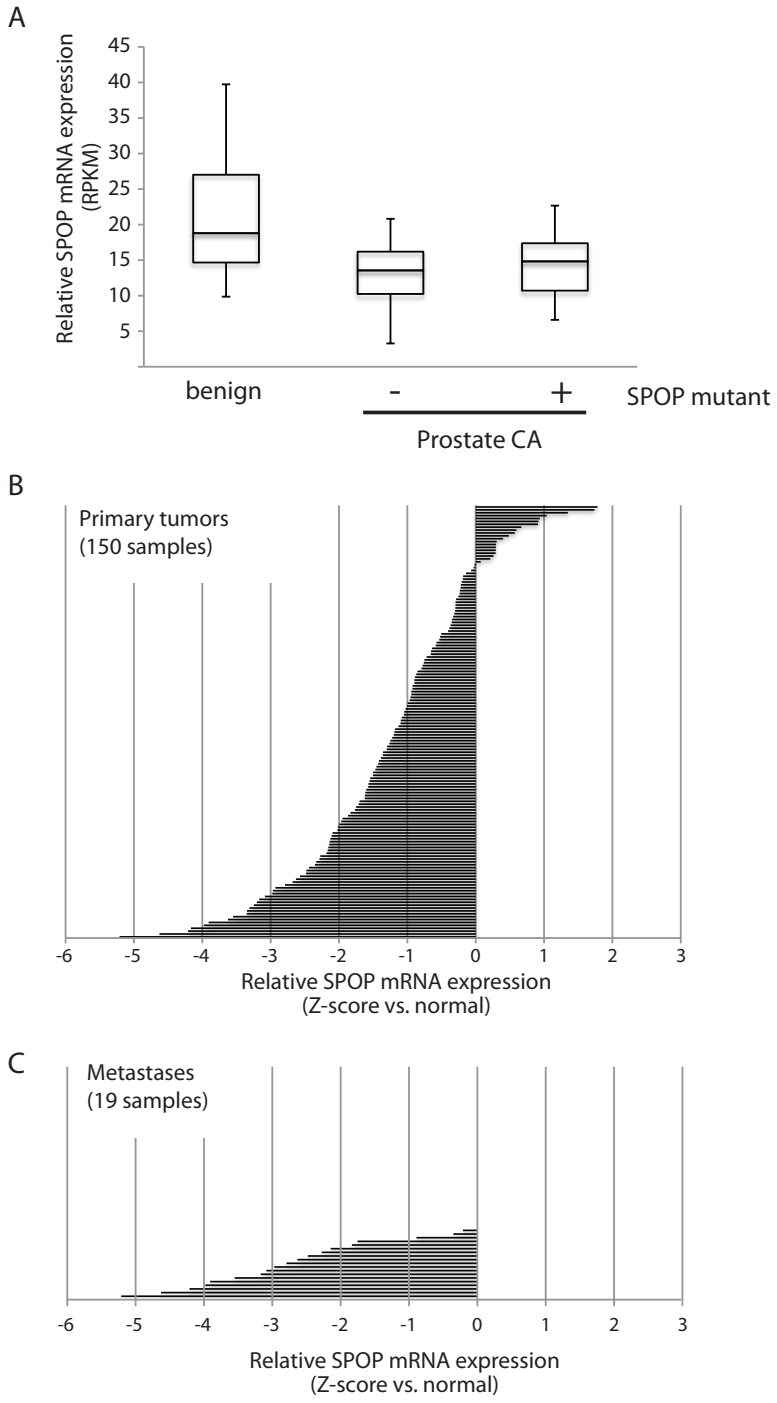
Supplementary Figure 8. Multiple independent siRNAs targeting *SPOP* have similar effects in prostate cell lines. (A) Expression of *SPOP* mRNA in DU145 cells transfected with 2 different control siRNAs and 4 different *SPOP* siRNAs, normalized to *GAPDH* expression, by real-time RT-PCR. (B) Quantitation of invaded DU145 cells transfected with control and SPOP siRNAs in Matrigel invasion assays. (C). Growth curves of DU145 cells transfected with control and *SPOP* siRNAs, measured with WST-1 assay.



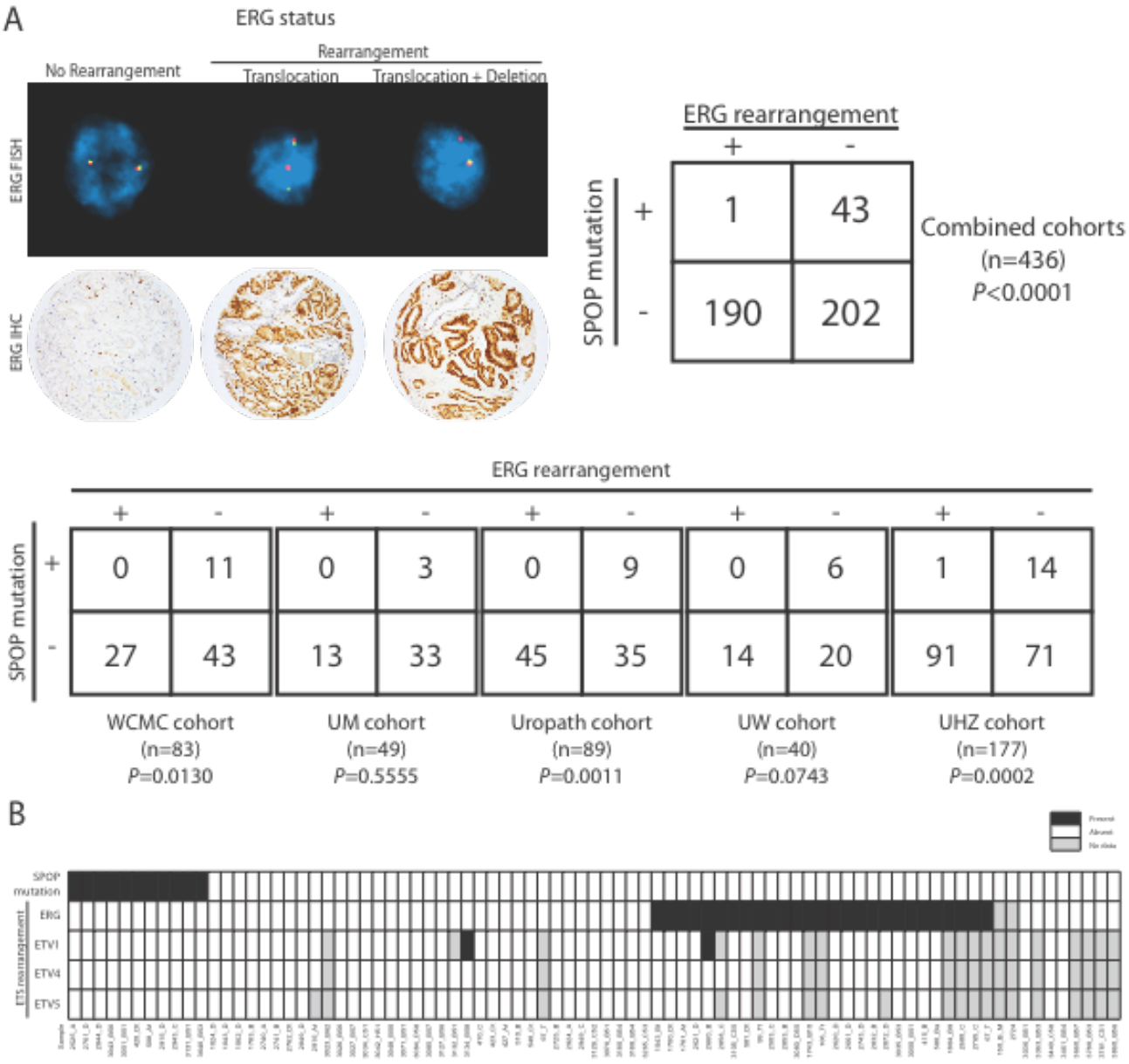
Supplementary Figure 9. Transfection with SPOP siRNA or SPOP mutant does not affect cell growth or viability. (A) Expression of SPOP mRNA in 22Rv1 and DU145 cells transfected with control and SPOP siRNA, normalized to GAPDH expression, by real-time RT-PCR. (B) Growth curves of 22Rv1 and DU145 cells transfected with control and *SPOP* siRNA, measured with WST-1 assay. (C) Western blot showing SPOP expression in DU145 cells transfected with SPOP wt and F133V. (D) Growth curves of DU145 cells transfected with SPOP wt and F133V, measured with WST-1 assay.



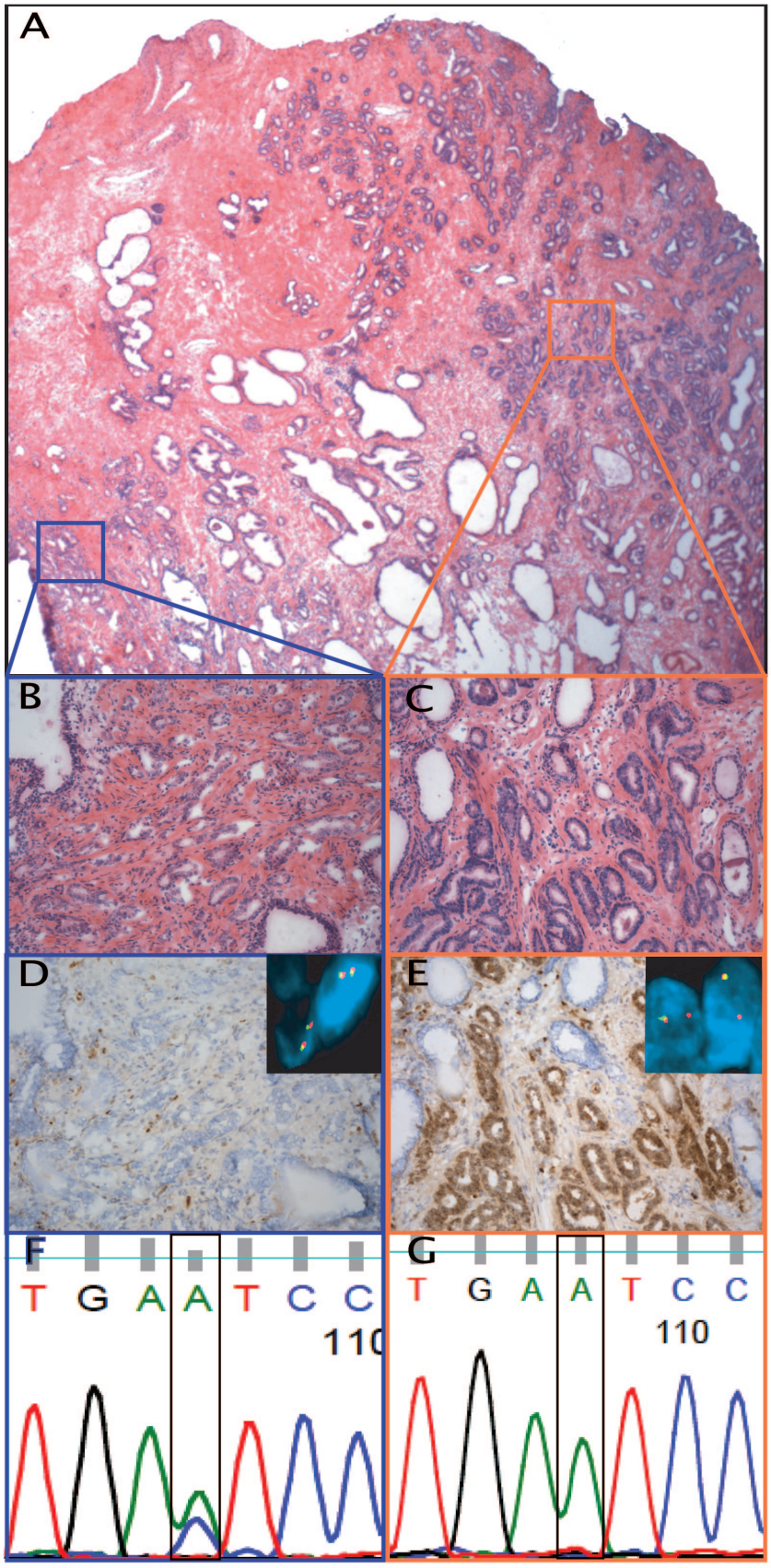
Supplementary Figure 10. Minimal changes in *SPOP* copy number in primary prostate cancer. Discretized copy number calls at SPOP locus in 157 primary prostate cancers, from a publicly available dataset (www.cbioportal.org/cgx/).[12]



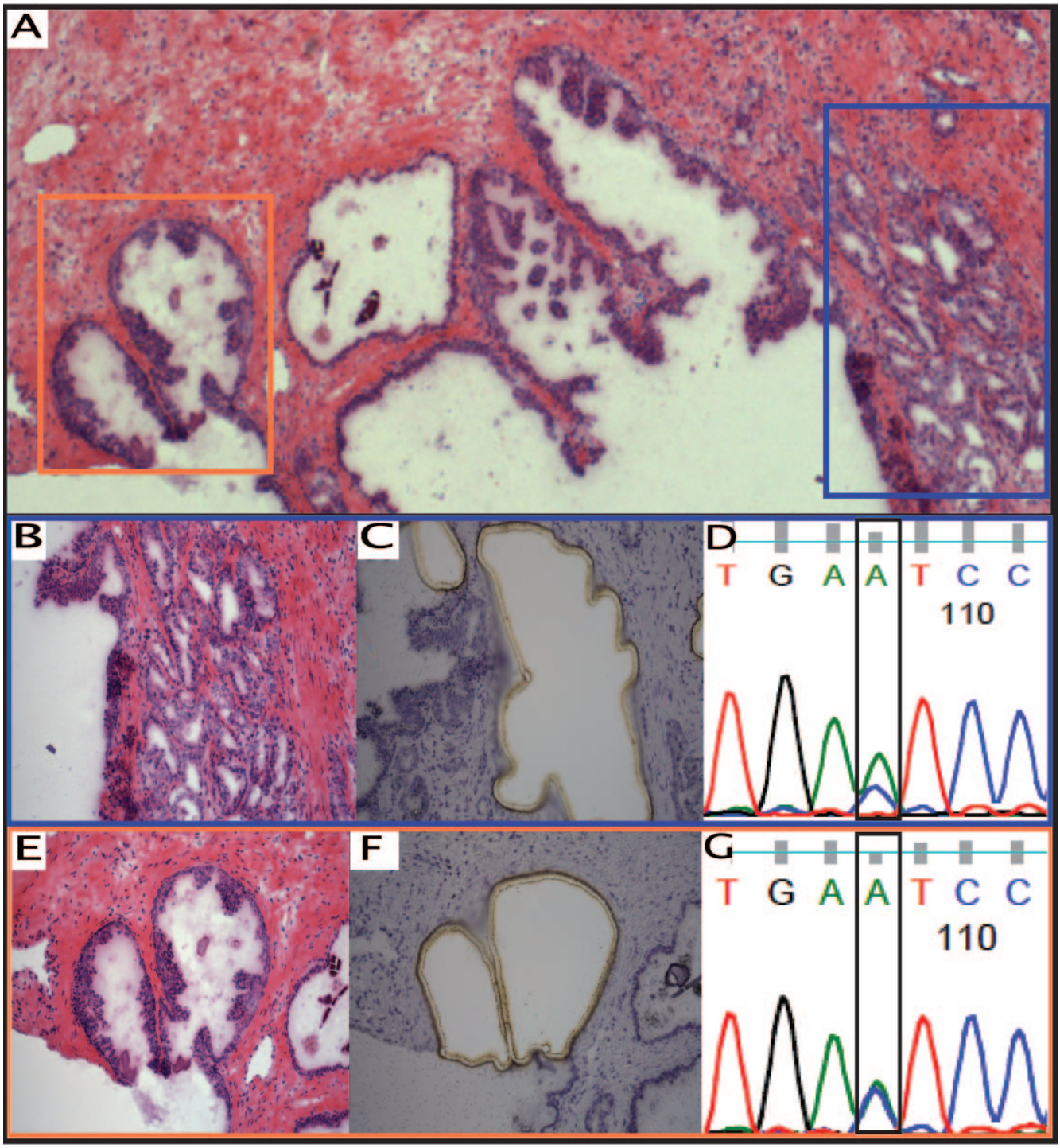
Supplementary Figure 11. *SPOP* is not upregulated in prostate cancer. (A) *SPOP* mRNA expression measured by RNA-seq in 6 benign prostate samples and 53 prostate cancers (7 *SPOP* mutant, 46 *SPOP* wt). Relative expression is displayed as reads Per kilobase per million mapped reads (RPKM). (B,C) *SPOP* mRNA expression from a publicly available dataset (www.cbioportal.org/cgx/)[12] in 150 primary prostate cancers (B) and 19 metastases (C). Relative expression is displayed as Z-score versus matched normal; positive = increased expression, negative = decreased expression.



Supplementary Figure 12. Tumors with *SPOP* mutation lack *ETS* rearrangements. (A) Relationship of *SPOP* mutation and *ERG* rearrangement. *ERG* rearrangement was determined by FISH and IHC. (B) Heatmap showing *SPOP* mutation status and rearrangement status of *ETS* genes in WCMC cohort.



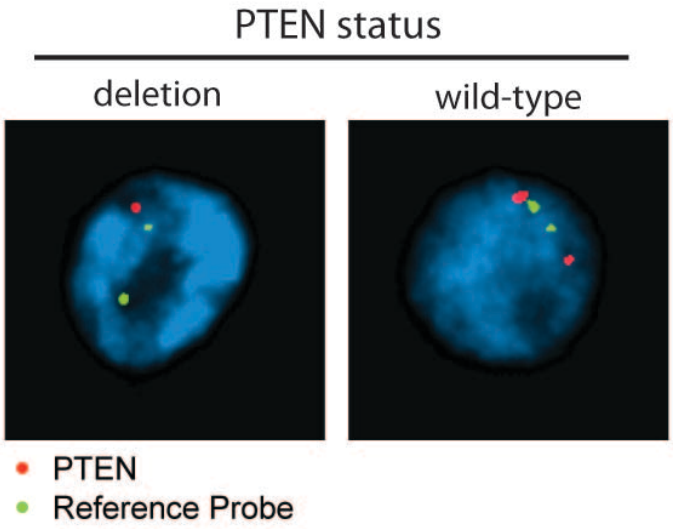
Supplementary Figure 13. Separate foci of prostate adenocarcinoma with mutually exclusive ERG-rearranged and SPOP-mutated status. (A) Low power view of two distinct foci of prostate adenocarcinoma in a prostatectomy specimen (H&E stained slide of frozen tissue, original magnification 2x). (B) The tumor on the left side (blue box) has Gleason score 3+4=7, is ERG-negative by immunohistochemistry (D) without *ERG* rearrangement by FISH (inset), and demonstrates F133V *SPOP*-mutation (F). (C) The tumor on the right side (orange box) has Gleason score 3+3=6, is ERG-positive by immunohistochemistry (E) with *ERG*-rearrangement by FISH (inset), and demonstrates *SPOP* wild-type sequence (F). H&E and immunohistochemistry for ERG in slides of frozen tissue, original magnification 20x; *ERG* break-apart fluorescent *in situ* hybridization assay, and *SPOP* DNA sequence by Sanger sequencing.



Supplementary Figure 14. Detection of SPOP mutation in high-grade prostatic intraepithelial neoplasia (HGPIN). (A) Low power view of prostate adenocarcinoma (blue box) and HGPIN (orange box) in a prostatectomy specimen. Cancer area before (B) and after (C) Laser Capture Microdissection (LCM). Images of HGPIN before (E) and after (F) LCM. DNA sequence demonstrates F133V *SPOP*-mutation in both adenocarcinoma (D) and HGPIN (G). H&E stained slide of frozen tissue, original magnification 10x; *SPOP* DNA sequence by Sanger sequencing.

Nature Genetics: doi:10.1038/ng.2279

A



SPOP mutation	PTEN (FISH) del	PTEN (FISH) wt
+	0	11
-	18	47

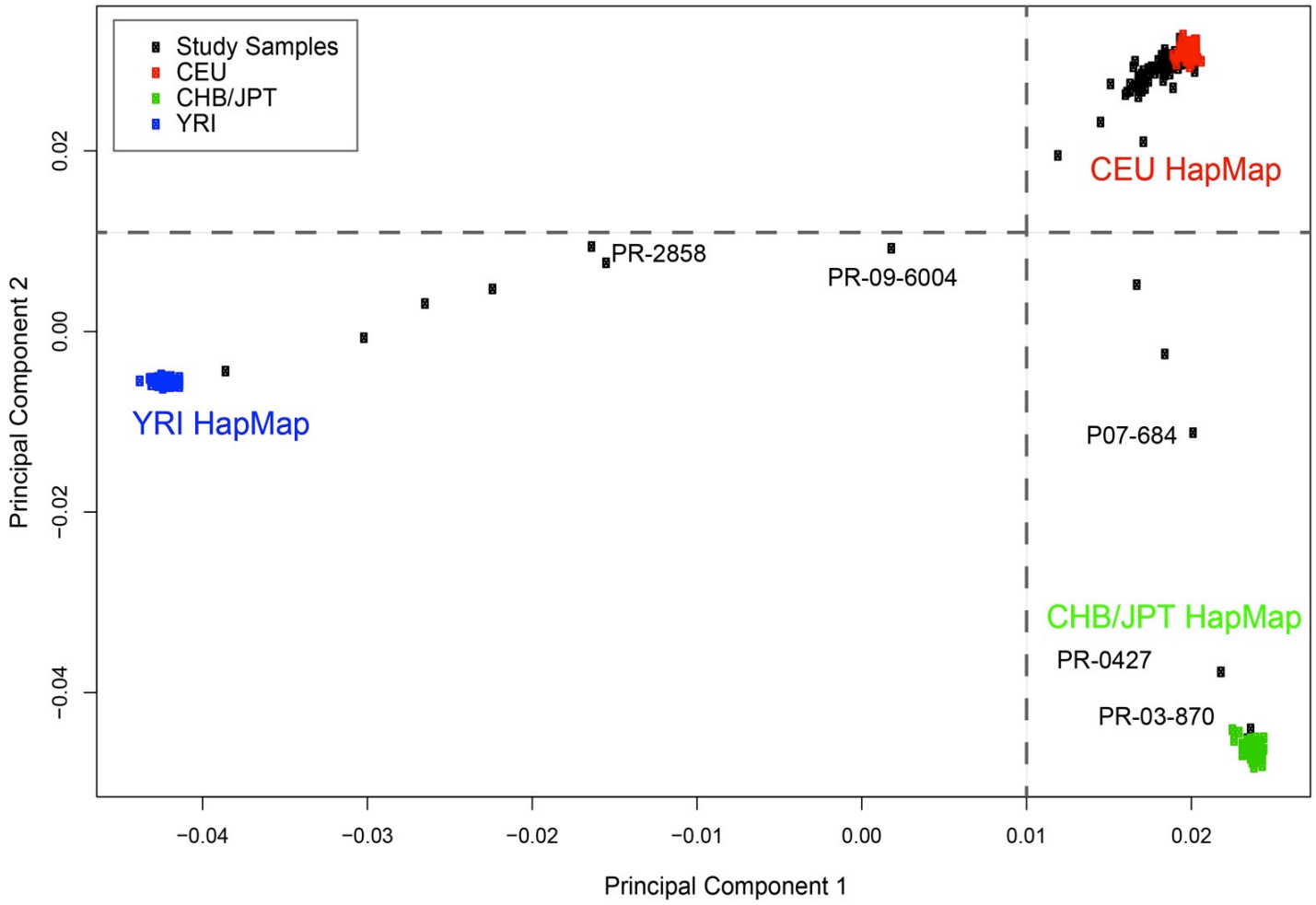
Primary
n=76
P=0.0572

B

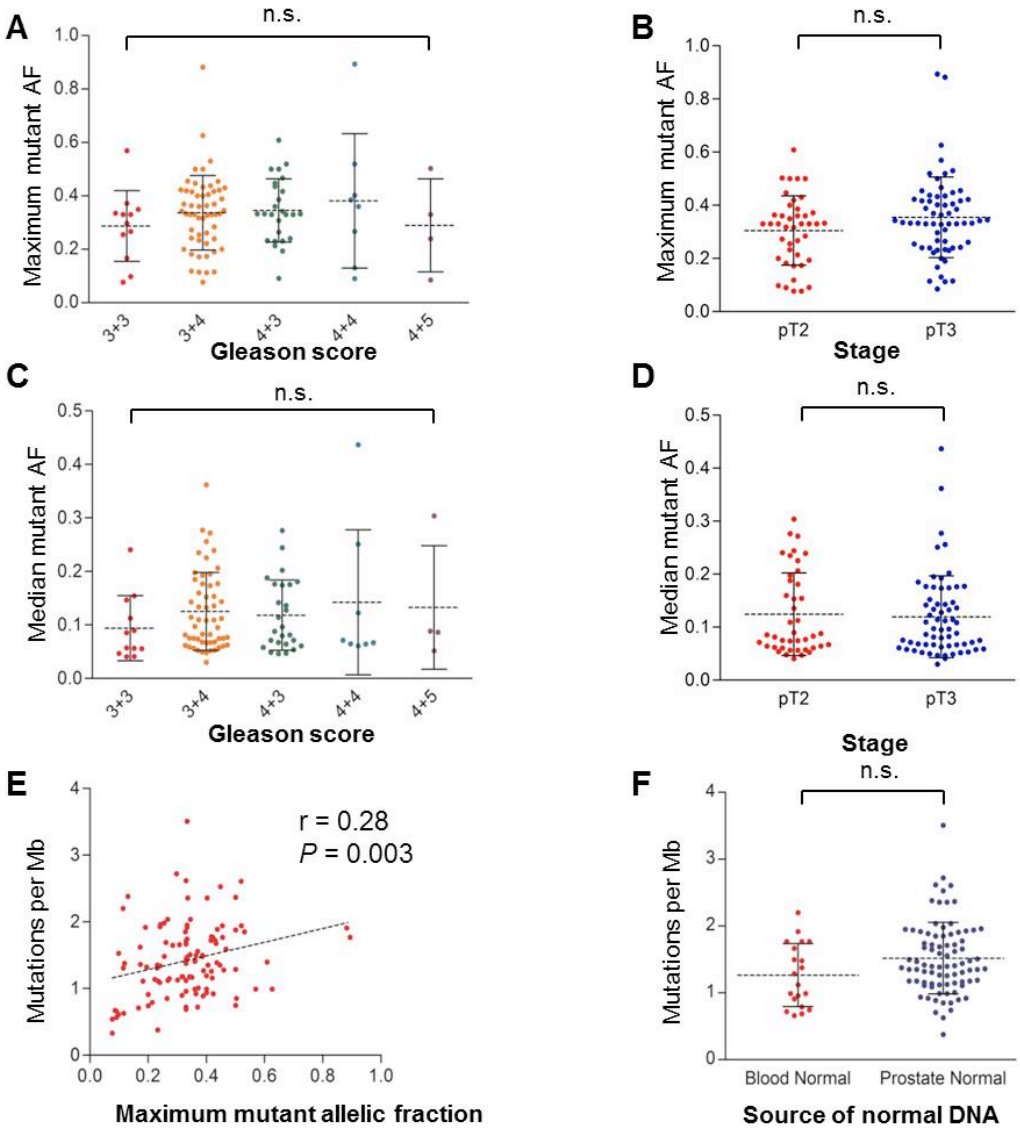
SPOP mutation	PTEN (CGH) del	PTEN (CGH) wt
+	4	2
-	25	10

Metastases
n=41
P=1.0

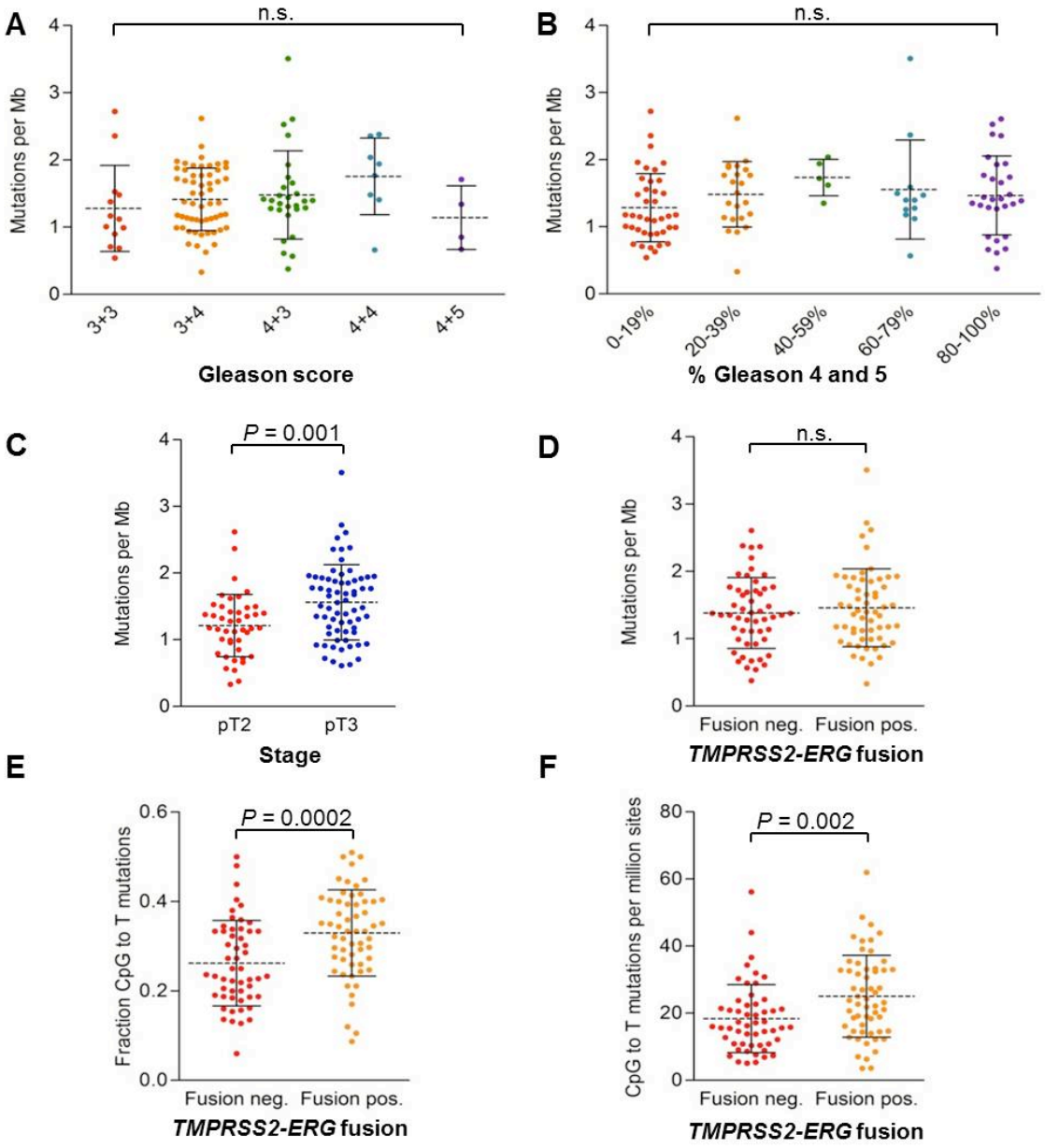
Supplementary Figure 15. Tumors with SPOP mutation lack PTEN deletion in primary but not metastatic prostate cancer. (A) Relationship of *SPOP* mutation and *PTEN* deletion determined by FISH in primary prostate cancers from the WCMC cohort. (B) Relationship of *SPOP* mutation and *PTEN* deletion determined by CGH in prostate cancer metastases from the UW cohort.



Supplementary Figure 16. Ethnicity analysis of exome-sequenced DNA. Principal component analysis was performed to assess the origin of the study individuals using SNP array data. HapMap Phase II samples representing three distinct populations, European (CEU) (red), Yoruban (YRI) (blue) and Chinese/Japanese (CHB/JPT) (green), were included in the analysis. The study identifiers of the exome-sequenced individuals whose genetic profiles deviate from the CEU pattern are shown.



Supplementary Figure 17. Relative ability to detect mutations in subgroups of tumors. (A-D) The allelic fraction (AF) values of mutations were used to characterize the relative purity of cancer DNA in each tumor. (A) and (B), maximum mutant AF observed in each tumor, grouped by Gleason score and stage. The top fifth percentile of AF values was removed in each tumor to exclude values that were elevated due to copy number variation at mutated sites. (C) and (D), as in (A) and (B), but showing median AF values across all mutations for each tumor. The relative purity of cancer DNA as assessed by AF did not vary by pathological stage or Gleason score. (E) Maximum mutant AF correlated only moderately with the number of mutations detected, implying that detection of mutations in most tumors was not limited by admixture of normal DNA. (F) Mutations per Mb sequenced in tumors grouped by source of paired normal DNA: peripheral blood (n=22) or non-cancerous prostate (n=89). No difference in mutation rates was observed between the two groups. Two-tailed p values were calculated using the Mann Whitney (two groups) or Kruskal-Wallis (multiple groups) tests and the Spearman test for correlation. Error bars indicate standard deviation. n.s., not significant. Statistical analysis was performed using GraphPad Prism.



Supplementary Figure 18. Mutational landscape across a spectrum of primary prostate cancers. (**A-D**) Mutations per Mb of covered DNA sequence for 111 primary prostatic adenocarcinomas grouped by clinical parameters. The horizontal axes denote: (A) Gleason score (major pattern + minor pattern); (B) Percent of cancer with Gleason pattern 4 or 5 histology; (C) Pathological tumor stage, where T3 indicates extra-prostatic extension; (D) Presence or absence of the *TMPRSS2-ERG* fusion based on fluorescence in-situ hybridization (FISH). Mutation rates are higher in pT3 tumors but do not vary by Gleason pattern or *TMPRSS2-ERG* fusion status. (E) Fraction of mutations in each tumor that are C to T transitions at CpG sites. (F) Number of CpG to T transitions per million CpG sites. Both the number and proportion of CpG to T mutations is increased in *TMPRSS2-ERG* fusion positive tumors. Two-tailed p values from the Mann Whitney test (two groups) or Kruskal-Wallis test (multiple groups) are indicated for each comparison. Statistical analysis was performed using GraphPad Prism. Error bars indicate standard deviation. n.s., not significant.

Supplementary References:

55 Rubin, M. A. *et al.* Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. *Clin. Cancer Res.* **6**, 1038-1045, (2000).

56 Trotman, L. C. *et al.* Pten dose dictates cancer progression in the prostate. *PLoS Biol.* **1**, E59, (2003).

57 McMenamin, M. E. *et al.* Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage. *Cancer Res.* **59**, 4291-4296, (1999).

58 Kim, J. H. *et al.* Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res.* **21**, 1028-1041, (2011).

59 Carver, B. S. *et al.* Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat. Genet.* **41**, 619-624, (2009).

60 King, J. C. *et al.* Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nat. Genet.* **41**, 524-526, (2009).

61 Gao, L. & Alumkal, J. Epigenetic regulation of androgen receptor signaling in prostate cancer. *Epigenetics* **5**, 100-104, (2010).

62 Cooperberg, M. R., Hilton, J. F. & Carroll, P. R. The CAPRA-S score: A straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer* **117**, 5039-5046, (2011).

63 Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744, (2009).

64 Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812, (2008).

65 Carpten, J. D. *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448**, 439-444, (2007).

66 Carter, B. S., Epstein, J. I. & Isaacs, W. B. Ras gene mutations in human prostate cancer. *Cancer Res.* **50**, 6830-6832 (1990).

67 Buhrman, G., Wink, G. & Mattos, C. Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. *Structure* **15**, 1618-1629, (2007).

68 Canman, C. E. Activation of the ATM Kinase by Ionizing Radiation and Phosphorylation of p53. *Science* **281**, 1677-1679, (1998).

69 Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158, (2007).

70 Sasaki, T. *et al.* ATM mutations in patients with ataxia telangiectasia screened by a hierarchical strategy. *Human Mutation* **12**, 186-195, (1998).

71 Miyaki, M. *et al.* Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. *Cancer Res.* **54**, 3011-3020 (1994).

72 Beroud, C. & Soussi, T. APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res.* **24**, 121-124 (1996).

73 Urick, M. E. *et al.* PIK3R1 (p85alpha) is somatically mutated at high frequency in primary endometrial cancer. *Cancer Res.* **71**, 4061-4067, (2011).

74 Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* **43**, 830-837, (2011).
75 Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435-439, (2011).
76 van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat. Genet.* **41**, 521-523, (2009).
77 Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43**, 875-878, (2011).
78 Wiegand, K. C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363**, 1532-1543, (2010).
79 Oricchio, E. *et al.* The eph-receptor a7 is a soluble tumor suppressor for follicular lymphoma. *Cell* **147**, 554-564, (2011).
80 Guan, M., Xu, C., Zhang, F. & Ye, C. Aberrant methylation of EphA7 in human prostate cancer and its relation to clinicopathologic features. *Int. J. Cancer* **124**, 88-94, (2009).
81 Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-231, (2010).
82 Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-363, (2010).
83 Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469**, 539-542, (2011).
84 Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806, (2008).
85 Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274, (2006).
86 Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005-1010, (2009).
87 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, (2008).
88 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058-1066, (2009).
89 Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196, (2010).
90 Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190, (2010).

Supplementary Table 1A. Clinical Characteristics of Exome-Sequenced Primary Prostate Cancers

Summary of exome-sequenced patient cohort

Characteristic	Whole exome-sequenced tumors	
Age, years		
Median (range)	63	(34 – 77)
Pre-operative Serum PSA (ng/μL)		
Median (range)	7.8	(2.7 – 31.5)
Pathologic Stage, N %		
Stage pT2 Total	44	39%
Stage pT2a	4	4%
Stage pT2b	1	1%
Stage pT2c	39	35%
Stage pT3 Total	68	61%
Stage pT3a	49	44%
Stage pT3b	19	17%
Gleason Pattern (major + minor), N %*		
Gleason 3+3	13	12%
Gleason 3+4	58	52%
Gleason 4+3	29	26%
Gleason 4+4	8	7%
Gleason 4+5	4	4%
Percentage of Gleason Pattern 4 and 5, N %*		
0-19%	40	36%
20-39%	23	12%
40-59%	5	5%
60-79%	12	11%
80-100%	29	26%
***TMPRSS2-ERG* Fusion Status, N % †**		
Fusion-negative	53	48%
Fusion with interstitial deletion	34	31%
Fusion without interstitial deletion	24	22%

* Gleason scores based on review of hematoxylin and eosin slides from site of tumor chosen for DNA extraction and sequencing

†*TMPRSS2-ERG* fusion status assessed by fluorescence *in situ* hybridization (FISH)

Supplementary Table 1B. Clinical Characteristics of Exome-Sequenced Primary Prostate Cancers

Clinical information for individual patients

Patient	Age	Gleason Score	GS % 4 and 5	Stage	Serum PSA (ng/mL)	TMPRSS2-ERG Fusion Status	mod CAPRA-S #
P00-000450	59	4+3	80	pT3a	8.3	Positive with interstitial deletion	4
P01-28	61	4+4	100	pT3a	14	Negative	6
P02-1562	67	4+3	95	pT3b	5.4	Negative	5
P02-2035	63	3+4	20	pT2c	5.8	Positive	1
P03-1334	69	3+4	2	pT3a	5.6	Negative	2
P03-1426	69	4+3	60	pT2c	13.9	Positive	4
P03-1906	75	3+4	30	pT3b	12.2	Positive	6
P03-2345	56	3+4	10	pT2c	7	Positive with interstitial deletion	2
P03-2620	68	3+4	20	pT3a	12.5	Negative	4
P03-3391	62	3+3	0	pT3b	18.6	Positive	5
P03-595	58	4+5	100	pT3a	24	Positive with interstitial deletion	7
P03-871	59	4+3;5	60	pT2c	6.5	Positive	3
P04-1084	65	3+4	40	pT3a	22	Negative	5
P04-1243	63	3+4	30	pT3a	7.3	Positive with interstitial deletion	3
P04-1421	54	3+4	20	pT2a	8.4	Positive	2
P04-1790	56	4+3;5	75	pT3a	9	Positive with interstitial deletion	4
P04-2599	68	4+3	60	pT2c	6.9	Negative	3
P04-2641	56	3+4	20	pT3a	6	Negative	3
P04-2666	58	4+5;3	95	pT3a	5	Negative	4
P04-2740	71	4+3	70	pT3a	11	Positive	5
P04-47	55	4+3;5	82	pT2b	14.6	Negative	4
P04-594	66	3+4	5	pT3a	8.5	Negative	3
P05-2212	66	3+4	30	pT3a	9.3	Positive with interstitial deletion	3
P05-2594	60	3+4	20	pT3b	13.5	Negative	6
P05-3436	62	4+3	80	pT2c	5.3	Positive	2
P05-3829	59	3+3	0	pT2c	10.8	Negative	2
P05-3852	62	4+3;5	60	pT3a	12.5	Negative	5
P05-3859	68	4+3	80	pT3a	8.2	Negative	4
P05-620	59	4+3	90	pT3a	5.8	Negative	3
P06-1125	66	3+4	30	pT3a	5.6	Positive with interstitial deletion	2
P06-1696	64	4+3;5	80	pT3a	4.6	Negative	3
P06-2325	62	4+3	70	pT2c	6.8	Negative	3
P06-3676	59	3+4	15	pT3a	4.8	Positive	2
P06-3939	71	4+5	100	pT3bN1	8.3	Negative	8
P06-4428	47	3+4	20	pT3a	31.5	Positive with interstitial deletion	5
P07-144	65	3+4	30	pT3a	2.7	Negative	2
P07-360	57	4+4	100	pT2c	16.7	Negative	5
P07-5036	52	3+4;5	60	pT3a	5.2	Positive	2
P07-684	58	4+3	90	pT3b	9.4	Negative	6
P07-718	57	3+4	5	pT2c	6.4	Negative	2
P07-837	69	4+3	80	pT3a	7.8	Negative	4
P08-2516	75	4+4	100	pT3b	9	Negative	7
P08-590	65	3+4	5	pT2c	9.2	Positive with interstitial deletion	2
P09-120	61	3+4	5	pT2c	6	Positive	1
P09-1372	73	3+4	5	pT3a	6.2	Positive with interstitial deletion	3
P09-1580	62	3+4;5	50	pT3a	5.2	Positive	2
P09-2497	62	3+3	0	pT3a	8	Positive with interstitial deletion	2
P09-649	71	4+3	80	pT2c	8.1	Negative	3
PR-00-1165	54	3+4	5	pT3b	6.7	Negative	5
PR-00-160	60	3+4	40	pT2c	9.4	Negative	2
PR-00-1823	60	3+4;5	12	pT3a	6.9	Positive with interstitial deletion	3
PR-0099	70	3+4	10	pT3a	10.3	Positive	4
PR-01-1934	59	3+4	20	pT3b	6.7	Negative	4
PR-01-2382	50	3+4	10	pT3a	6.2	Positive	3
PR-01-2492	67	3+4	10	pT3a	Not known	Negative	NA
PR-01-2554	69	4+3	90	pT3b	8.9	Positive	5
PR-02-1082	61	3+4	30	pT3b	11.5	Positive with interstitial deletion	6
PR-02-169	65	3+4	20	pT2c	10.6	Positive with interstitial deletion	3
PR-02-1736	58	3+4	5	pT3b	13	Negative	5
PR-02-1899	66	3+3	0	pT3a	13.4	Positive with interstitial deletion	3
PR-02-2072	54	4+3;5	62	pT3a	9	Positive with interstitial deletion	4
PR-02-2480	69	3+4	20	pT3b	10.4	Positive with interstitial deletion	5
PR-02-254	74	4+3	70	pT3b	17.6	Positive	7
PR-03-022	70	3+4	10	pT3a	9.7	Positive	3
PR-03-1026	56	4+3;5	85	pT3a	12.8	Positive	5
PR-03-870	52	4+3	90	pT3a	5.3	Negative	3
PR-04-1367	53	3+3	0	pT2c	9.9	Positive with interstitial deletion	1
PR-0415	71	4+4	100	pT3a	7.6	Positive with interstitial deletion	5
PR-04-194	65	3+4	30	pT2c	7.1	Positive with interstitial deletion	2
PR-0427	69	4+3		pT2a	5.1	Negative	1
PR-04-3113	66	3+3	0	pT3b	12	Positive	4
PR-04-3222	66	3+4	5	pT2c	5.2	Negative	1
PR-04-3347	59	3+3	0	pT2c	7.5	Negative	1
PR-04-639	61	3+4	5	pT3a	5.1	Negative	2
PR-04-903	68	3+4	5	pT3a	7.8	Positive with interstitial deletion	3
PR-05-3440	63	4+3	80	pT3a	8.5	Positive with interstitial deletion	4
PR-05-3595	64	3+4	50	pT2c	7.2	Negative	2
PR-05-839	69	3+4	30	pT2a	9	Negative	2
PR-06-1651	60	4+5;3	95	pT2c	5.8	Positive with interstitial deletion	3
PR-06-1749	60	4+3	60	pT2c	6.7	Negative	3
PR-06-1999	56	3+4	40	pT2a	30	Positive	4
PR-09-2517	64	3+4	5	pT3a	14	Positive with interstitial deletion	4
PR-09-2744	48	3+3	0	pT3a	4.4	Positive with interstitial deletion	1
PR-09-2767	58	4+4	100	pT3a	4.6	Positive with interstitial deletion	4
PR-09-3421	65	3+4	10	pT3a	4.8	Positive with interstitial deletion	2
PR-09-3566	65	4+3	80	pT3a	5	Positive with interstitial deletion	3
PR-09-3687	73	3+4	5	pT2c	7.8	Positive	2
PR-09-5094	77	4+4	100	pT3b	17	Negative	7
PR-09-5245	72	4+4	100	pT3a	6.1	Negative	5
PR-09-5446	67	3+3	0	pT2c	13	Positive with interstitial deletion	2
PR-09-5630	64	4+3	60	pT3b	14	Positive with interstitial deletion	7
PR-09-5700	60	3+4	20	pT3a	14	Positive	4
PR-09-5702	68	3+4	20	pT2c	8.2	Positive with interstitial deletion	2
PR-1024	73	3+4;5	22	pT3b	7.2	Negative	4
PR-1043	59	3+3	0	pT2c	4.9	Positive	0
PR-2661	59	3+3	0	pT2c	4.5	Positive with interstitial deletion	0
PR-2682	71	4+3	80	pT2c	7.6	Negative	3
PR-2740	58	3+3	0	pT2c	2.8	Negative	0
PR-2761	56	3+3	0	pT2c	4.3	Negative	0
PR-2762	68	3+4	5	pT2c	5.5	Negative	1
PR-2858	52	3+4	N/A	pT3a	5.3	Positive	2
PR-2872	62	3+4	5	pT3a	6	Positive with interstitial deletion	2
PR-2915	66	3+4	30	pT2c	3.6	Negative	1
PR-2916	65	3+4	2	pT3a	9.1	Negative	3
PR-3023	67	4+3	95	pT2c	5	Negative	2
PR-3026	56	3+4	20	pT3b	24.2	Negative	7
PR-3034	70	4+4	100	pT2c	14.6	Negative	5
PR-3035	34	3+4	5	pT2c	4.2	Positive with interstitial deletion	1
PR-3036	47	3+4	5	pT2c	5.1	Negative	1
PR-3048	64	3+4	10	pT2c	3.5	Negative	1
PR-3051	60	3+4	20	pT2c	20	Negative	3
PR-3127	55	3+4	2	pT2c	4.2	Negative	1

Modified CAPRA-S score, derived from CAPRA-S score (62) without including surgical margins

Nature Genetics: doi:10.1038/ng.2279

Supplementary Table 4. Mutations in Significantly-Mutated Genes and other Selected Genes

Selected mutations detected by exome sequencing are annotated based on the indicated UCSC Genome Browser identifier. Additional mutations detected by Sanger sequencing or transcriptome sequencing of independent tumors are also listed. Mutations are grouped by three catogies: (1) Mutations in significantly mutated genes. (2) Mutations previously reported in cancer. (3) Selected genes with damaging mutations (splice site-alteration, truncation or frame-shift) and are inactivated by similar mutations in other cancer types. References are listed that support the role of either the particular mutation (for missense mutations) or the type of mutation (for nonsense, frame-shift or splice site mutations) in prostate cancer or other neoplasms. "RNA" denotes the median $\log_{10}$(RPKM+1) value for each transcript across 63 tumors that underwent transcriptome sequencing. fs, frame-shift mutation; del, deletion; CV, coefficient of variation.

Category	Gene symbol (Annotation transcript)	Patient	Genomic coordinates (hg19)	Variant type	Codon change	Protein change	Median RNA	Mean RNA	RNA CV	RNA Percentile	References
1. Significantly mutated genes (q-value < 0.1)	***SPOP*** (uc010dbk.2)	P04-2599	g.chr17:47696425A>G	Missense	c.(397-399)TTC>TCC	p.F133S	3.88	3.78	0.13	88%	
		P07-684	g.chr17:47696448A>T	Missense	c.(373-375)TTT>TTA	p.F125L					
		P07-837	g.chr17:47696424G>C	Missense	c.(397-399)TTC>TTG	p.F133L					
		P09-649	g.chr17:47696467C>T	Missense	c.(355-357)AGT>AAT	p.S119N					
		PR-00-160	g.chr17:47696421C>G	Missense	c.(400-402)AAG>AAC	p.K134N					
		PR-01-2492	g.chr17:47696425A>C	Missense	c.(397-399)TTC>TGC	p.F133C					
		PR-04-3222	g.chr17:47696424G>T	Missense	c.(397-399)TTC>TTA	p.F133L					
		PR-05-839	g.chr17:47696689A>T	Missense	c.(259-261)TAC>AAC	p.Y87N					
		PR-09-5245	g.chr17:47696432A>C	Missense	c.(391-393)TGG>GGG	p.W131G					
		PR-2661	g.chr17:47696426A>C	Missense	c.(397-399)TTC>GTC	p.F133V					
		PR-2740	g.chr17:47696424G>T	Missense	c.(397-399)TTC>TTA	p.F133L					
		PR-2761	g.chr17:47696425A>G	Missense	c.(397-399)TTC>TCC	p.F133S					
		PR-2915	g.chr17:47696426A>C	Missense	c.(397-399)TTC>GTC	p.F133V					
		PR-3051	g.chr17:47696438T>C	Missense	c.(385-387)AAA>GAA	p.K129E					
		STID0000000420_D (RNA-seq)	g.chr17:47696424G>C	Missense	c.(397-399)TTC>TTG	p.F133L					
		STID0000000508_A (RNA-seq)	g.chr17:47696426A>C	Missense	c.(397-399)TTC>GTC	p.F133V					
		STID0000002525_A (RNA-seq)	g.chr17:47696689A>T	Missense	c.(259-261)TAC>AAC	p.Y87N					
		STID0000002844_D (RNA-seq)	g.chr17:47696424G>T	Missense	c.(397-399)TTC>TTA	p.F133L					
		STID0000002943_C (Sanger seq)	g.chr17:47696688T>C	Missense	c.(259-261)TAC>TGC	p.Y87C					
		STID0000003043_B56 (Sanger seq)	g.chr17:47696643A>C	Missense	c.(304-306)TTC>TGC	p.F102C					
		STID0000003131_B51 (Sanger seq)	g.chr17:47696425A>C	Missense	c.(397-399)TTC>TGC	p.F133C					
		STID0000003640_B53 (RNA-seq)	g.chr17:47696688T>C	Missense	c.(259-261)TAC>TGC	p.Y87C					
	TP53 (uc002gim.2)	P03-871	g.chr17:7579417-7579426(del)	Frame shift deletion	c.(259-270)fs	p.P87fs	3.99	3.95	0.16	89%	
		P04-1243	g.chr17:7577121G>A	Missense	c.(817-819)CGT>TGT	p.R273C					
		P06-4428	g.chr17:7577547C>T	Missense	c.(733-735)GGC>GAC	p.G245D					
		P06-4428	g.chr17:7578457C>T	Missense	c.(472-474)CGC>CAC	p.R158H					
		P07-5036	g.chr17:7577538C>T	Missense	c.(742-744)CGG>CAG	p.R248Q					
		PR-02-254	g.chr17:7578542G>A	Missense	c.(388-390)CTC>TTC	p.L130F					
		PR-0415	g.chr17:7577529A>T	Missense	c.(751-753)ATC>AAC	p.I251N					
		PR-06-1999	g.chr17:7579359G>A	Missense	c.(328-330)CGT>TGT	p.R110C					
	PTEN (uc001kfb.2)	P04-1790	g.chr10:89692900G>C	Missense	c.(382-384)AAG>AAC	p.K128N	2.78	2.55	0.29	75%	
		P06-3676	g.chr10:89692905G>A	Missense	c.(388-390)CGA>CAA	p.R130Q					
		PR-02-254	g.chr10:89720857C>G	Nonsense	c.(1006-1008)TAC>TAG	p.Y336*					
		PR-02-254	g.chr10:89692901G>C	Missense	c.(385-387)GGA>CGA	p.G129R					
		PR-3026	g.chr10:89711874G>A	Splice site mutation	N/A	p.G165_splice					
	FOXA1 (uc001wuf.2)	P05-620	g.chr14:38061231A>T	Missense	c.(757-759)ATG>AAG	p.M253K	6.92	6.94	0.09	99%	
		PR-04-1367	g.chr14:38061294G>A	Missense	c.(694-696)GCA>GTA	p.A232V					
		PR-0427	g.chr14:38061231A>C	Missense	c.(757-759)ATG>AGG	p.M253R					
		PR-09-3687	g.chr14:38061313C>T	Missense	c.(676-678)GAC>AAC	p.D226N					
		STID0000003071_B51 (RNA-seq)	g.chr14:38061231A>C	Missense	c.(757-759)ATG>AGG	p.M253R					
		STID0000000423_Ct (RNA-seq)	g.chr14:38061231A>T	Missense	c.(757-759)ATG>AAG	p.M253K					
		STID0000016350_B53T (RNA-seq)	g.chr14:38061313C>T	Missense	c.(676-678)GAC>AAC	p.D226N					
		STID0000002832_C (RNA-seq)	g.chr14:38061313C>A	Missense	c.(676-678)GAC>TAC	p.D226Y					
	CDKN1B (uc001rat.2)	P00-000450	g.chr12:12870910A>G	Missense	c.(136-138)GAG>GGG	p.E46G	3.90	3.72	0.15	88%	
		P03-1426	g.chr12:12871788-12871788(del)	Frame shift deletion	c.(505-507)fs	p.R169fs					
		PR-00-160	g.chr12:12871228-12871228(del)	Frame shift deletion	c.(454-456)fs	p.R152fs					
	ZNF595 (uc003fzv.1)	P03-1334	g.chr4:59330T>C	Missense	c.(10-12)GTA>GCA	p.V4A	1.23	1.25	0.36	53%	
		P04-594	g.chr4:59357T>A	Missense	c.(37-39)TTC>TAC	p.F13Y					
		PR-09-3687	g.chr4:85888A>T	Missense	c.(493-495)ATA>TTA	p.I165L					
		PR-09-5702	g.chr4:86116C>A	Missense	c.(721-723)ACA>AAA	p.T241K					
	THSD7B (uc002tva.1)	PR-01-1934	g.chr2:138163182G>C	Splice site mutation	N/A	p.A803_splice	0.07	0.13	1.41	27%	
		PR-03-1026	g.chr2:137988713C>T	Missense	c.(1729-1731)ACG>ATG	p.T577M					
		PR-05-839	g.chr2:138000058C>T	Missense	c.(2089-2091)CGC>TGC	p.R697C					
		PR-09-2767	g.chr2:138373832G>C	Missense	c.(3424-3426)GTG>CTG	p.V1142L					
		PR-09-5702	g.chr2:138033591C>A	Missense	c.(2401-2403)ACA>AAA	p.T801K					
		PR-2872	g.chr2:137928505C>T	Missense	c.(1627-1629)CGC>TGC	p.R543C					
	MED12 (uc004dyy.2)	P04-594	g.chrX:70349258C>T	Missense	c.(3670-3672)CTC>TTC	p.L1224F	2.92	2.93	0.12	77%	
		P07-684	g.chrX:70349946C>A	Missense	c.(3928-3930)CCA>CAA	p.P1310Q					
		PR-00-160	g.chrX:70344951C>A	Missense	c.(2179-2181)GAC>GAA	p.D727E					
		PR-3026	g.chrX:70349258C>T	Missense	c.(3670-3672)CTC>TTC	p.L1224F					
		PR-3035	g.chrX:70349258C>T	Missense	c.(3670-3672)CTC>TTC	p.L1224F					
		PR-3036	g.chrX:70349257TC>GT	Missense	c.(3667-3672)GTTCTC>GTGTTC	p.L1224F					
		PR-3084 (RNA-seq)	g.chrX:70349257TC>GT	Missense	c.(3667-3672)GTTCTC>GTGTTC	p.L1224F					
	NIPA2 (uc001yux.2)	P03-1906	g.chr15:23012398A>C	Missense	c.(328-330)CTT>CGT	p.L110R	3.58	3.61	0.17	85%	
		P07-837	g.chr15:23006670C>A	Missense	c.(634-636)GTG>TTG	p.V212L					
		PR-09-3421	g.chr15:23014493C>A	Missense	c.(232-234)GCG>TCG	p.A78S					
	PIK3CA (uc003fjk.2)	PR-04-639	g.chr3:178952085A>G	Missense	c.(3139-3141)CAT>CGT	p.H1047R	1.39	1.31	0.28	55%	
		PR-09-2517	g.chr3:178917478G>A	Missense	c.(352-354)GGT>GAT	p.G118D					
		PR-09-2744	g.chr3:178936095A>C	Missense	c.(1636-1638)CAG>CCG	p.Q546P					
		PR-09-3421	g.chr3:178952006T>C	Missense	c.(3061-3063)TAC>CAC	p.Y1021H					
	C14orf49 (uc001yei.3)	P00-000450	g.chr14:95884213C>T	Missense	c.(2878-2880)GCC>ACC	p.A960T	0.38	0.42	0.57	38%	
		P07-718	g.chr14:95932380C>T	Missense	c.(514-516)CGG>CAG	p.R172Q					
		PR-09-5700	g.chr14:95899740C>A	Nonsense	c.(2545-2547)GAG>TAG	p.E849*					
		PR-2916	g.chr14:95921763G>A	Missense	c.(1087-1089)GCG>GTG	p.A363V					
		PR-3023	g.chr14:95906394G>A	Missense	c.(1930-1932)CGG>TGG	p.R644W					
	SCN11A (uc011ays.1)	P07-144	g.chr3:38926826A>T	Nonsense	c.(3016-3018)TTA>TAA	p.L1006*	0.02	0.03	1.15	24%	
		PR-00-1823	g.chr3:38938566G>A	Missense	c.(2173-2175)CGT>TGT	p.R725C					
		PR-01-2554	g.chr3:38936068C>A	Nonsense	c.(2791-2793)GAA>TAA	p.E931*					
		PR-06-1999	g.chr3:38938451C>T	Missense	c.(2287-2289)CGC>CAC	p.R763H					
		PR-09-2517	g.chr3:38938451C>T	Missense	c.(2287-2289)CGC>CAC	p.R763H					
2. Mutations previously reported in cancer	***IDH1*** (uc002vcs.2)	P06-2325	g.chr2:209113112C>T	Missense	c.(394-396)CGT>CAT	p.R132H	6.00	5.94	0.14	98%	(63),(64)
		PR-05-3595	g.chr2:209113112C>T	Missense	c.(394-396)CGT>CAT	p.R132H					
	AKT1 (uc001ypk.2)	P06-1125	g.chr14:105246551C>T	Missense	c.(49-51)GAG>AAG	p.E17K	5.51	5.51	0.06	97%	(65)
	HRAS (uc001lpv.2)	PR-09-5702	g.chr11:533874T>C	Missense	c.(181-183)CAG>CGG	p.Q61R	4.98	4.94	0.08	95%	(66),(67)
	ATM *(uc001pkb.1)*	P00-000450	g.chr11:108218029G>T	Missense	c.(8608-8610)GAT>TAT	p.D2870Y	1.33	1.30	0.30	54%	(68-70)
		PR-09-2767	g.chr11:108216537C>T	Missense	c.(8485-8487)CCA>CTA	p.P2829L					
		PR-02682	g.chr11:108201148-108201148(del)	Frame shift deletion	c.(7513-7515)fs	p.K2505fs					
3. Genes with a damaging mutation that undergo damaging mutations in other cancer types	***APC*** (uc010jby.2)	P05-620	g.chr5:112173635A>T	Nonsense	c.(2344-2346)AAG>TAG	p.K782*	1.11	1.18	0.34	51%	(71),(72)
		PR-00-160	g.chr5:112157644-112157648(del)	Frame shift deletion	c.(1363-1368)fs	p.K455fs					
		PR-0415	g.chr5:112175979-112175979(del)	Frame shift deletion	c.(4687-4689)fs	p.L1563fs					
	PIK3R1 (uc003jva.2)	PR-09-3566	g.chr5:67591246A>G	Splice site mutation	N/A	p.M582_splice	2.29	2.22	0.23	69%	(73)
	MLL2 (uc001rta.3)	PR-2661	g.chr12:49438067G>A	Nonsense	c.(5104-5106)CGA>TGA	p.R1702*	3.26	3.13	0.17	81%	(39),(74),(75)
	MLL3 (uc003wla.2)	PR-3048	g.chr7:151859711C>A	Nonsense	c.(10951-10953)GAG>TAG	p.E3651*	2.25	2.22	0.21	68%	(39),(75)
	KDM6A (uc011mkz.1)	P07-360	g.chrX:44941865-44941872(del)	Frame shift deletion	c.(3343-3354)fs	p.S1115fs	2.00	2.01	0.18	64%	(76)
		PR-2916	g.chrX:44945155G>T	Missense	c.(3634-3636)GGC>GTC	p.G1212V					
		PR-3036	g.chrX:44945146T>A	Missense	c.(3625-3627)ATA>AAA	p.I1209K					
	ARID1A (uc001bmv.1)	P04-2599	g.chr1:27101666C>T	Nonsense	c.(4948-4950)CAG>TAG	p.Q1650*	4.03	3.98	0.09	89%	(77),(78)
		PR-09-3421	g.chr1:27099857A>T	Missense	c.(3736-3738)ATG>TTG	p.M1246L					
	EPHA7 (uc003poe.2)	P04-1084	g.chr6:93967815A>C	Splice site mutation	N/A	p.G704_splice	0.92	0.95	0.47	48%	(79),(80)
	NCOR1 (uc002gpo.2)	P05-620	g.chr17:16012098A>C	Splice site mutation	N/A	p.V728_splice	3.07	3.10	0.13	79%	(77)
		P02-2480	g.chr17:15935762G>A	Missense	c.(7171-7173)CGG>TGG	p.R2391W					
	MLL (uc001ptb.2)	PR-2682	g.chr11:118355577G>T	Nonsense	c.(4219-4221)GAG>TAG	p.E1407*	2.35	2.31	0.16	69%	(77)

Supplementary Table 5. Significantly Mutated Gene Sets

Curated gene sets representing canonical pathways were analyzed for enrichment of mutations in their constituent genes (Supplementary Methods). Sets that are significantly mutated above a q-value of 0.05 (Benjamini-Hochberg adjustment) (*52*) are listed.

Gene Set	No. mutations	No. tumors with a mutation	B-H q-value	Genes in set mutated (no. tumors affected)
SA_G1_AND_S_PHASES	11	10	0.00033	CDKN1B(3), TP53(8)
RBPATHWAY	14	13	0.00033	ATM(3), CDC25A(1), MYT1(1), TP53(8), WEE1(1)
P53HYPOXIAPATHWAY	14	13	0.0048	ABCB1(2), AKT1(1), ATM(3), TP53(8)
TERTPATHWAY	9	8	0.0048	SP1(1), TP53(8)
PLK3PATHWAY	11	10	0.0062	ATM(3), TP53(8)
IGF1MTORPATHWAY	14	13	0.0072	AKT1(1), GSK3B(1), IGF1R(1), INPPL1(1), PIK3CA(4), PIK3R1(1), PTEN(5)
ARFPATHWAY	13	12	0.008	PIK3CA(4), PIK3R1(1), TP53(8)
G1PATHWAY	16	14	0.008	ATM(3), CDC25A(1), CDKN1B(3), GSK3B(1), TP53(8)
G2PATHWAY	17	16	0.011	ATM(3), CDC25A(1), EP300(1), MYT1(1), PRKDC(1), RPS6KA1(1), TP53(8), WEE1(1)
CHEMICALPATHWAY	16	14	0.011	AKT1(1), ATM(3), CASP3(1), CASP9(1), TLN1(2), TP53(8)
PTENPATHWAY	17	16	0.011	AKT1(1), BCAR1(1), CDKN1B(3), ILK(1), PIK3CA(4), PIK3R1(1), PTEN(5), SHC1(1)
RNAPATHWAY	9	8	0.011	DNAJC3(1), TP53(8)
P53PATHWAY	11	10	0.017	ATM(3), TP53(8)
COMPLEMENT_ACTIVATION_CLASSICAL	10	10	0.017	C1QB(1), C1S(1), C3(3), C6(1), C8A(1), C9(2), MASP1(1)
CLASSICPATHWAY	9	9	0.018	C1QB(1), C1S(1), C3(3), C6(1), C8A(1), C9(2)
COMPPATHWAY	10	10	0.018	C1QB(1), C1S(1), C3(3), C6(1), C8A(1), C9(2), MASP1(1)
HCMVPATHWAY	10	10	0.018	AKT1(1), CREB1(1), MAP2K6(1), MAPK14(1), PIK3CA(4), PIK3R1(1), SP1(1)
TELPATHWAY	14	13	0.025	AKT1(1), EGFR(1), IGF1R(1), POLR2A(1), TEP1(1), TP53(8), XRCC5(1)
ALTERNATIVEPATHWAY	7	7	0.027	C3(3), C6(1), C8A(1), C9(2)
CDC42RACPATHWAY	8	7	0.027	ACTR2(1), PDGFRA(2), PIK3CA(4), PIK3R1(1)
SA_PTEN_PATHWAY	12	11	0.04	AKT1(1), ILK(1), PIK3CA(4), PTEN(5), SHC1(1)
RACCYCDPATHWAY	11	11	0.04	AKT1(1), CDKN1B(3), HRAS(1), PIK3CA(4), PIK3R1(1), RAF1(1)
CELL_CYCLE_KEGG	29	28	0.042	ATM(3), BUB3(1), CDC20(1), CDC25A(1), CDC6(1), CDH1(1), EP300(1), ESPL1(1), GSK3B(1), HDAC2(2), HDAC3(1), HDAC5(2), MAD1L1(1), MCM4(1), PRKDC(1), SMAD4(1), TP53(8), WEE1(1)
IGF1RPATHWAY	11	11	0.042	AKT1(1), HRAS(1), IGF1R(1), IRS1(1), PIK3CA(4), PIK3R1(1), RAF1(1), SHC1(1)

Supplementary Table 6. Systematic Sequencing Studies Including *SPOP*

Cancer type	Samples (N)	*SPOP* mutations	Reference	Approach
Prostate	7	2: F102C, F133V	(14)	Paired-end whole-genome sequencing
Prostate	58	2: F125V, F133V	(17)	Mismatch repair detection (MRD)
Lung	134	2: N169S, L190F	(17)	Mismatch repair detection (MRD)
Ovarian	316	1: E249*	(42)	Whole exome sequencing
Ovarian	8	1: E47K	(81)	Whole exome sequencing
Ovarian	58	None	(17)	Mismatch repair detection (MRD)
HNSCC	76	None	(38)	Whole exome and paired-end whole-genome sequencing
RCC	101	None	(82)	PCR-based exon resequencing
RCC	7	None	(83)	Whole exome sequencing
Pancreatic	8	None	(17)	Mismatch repair detection (MRD)
Pancreatic	24	None	(84)	Whole exome sequencing
Breast	183	None	(17)	Mismatch repair detection (MRD)
Breast	11	None	(85)	Whole exome sequencing
Breast	24	None	(86)	Paired-end whole-genome sequencing
Medulloblastoma	22	None	(75)	Whole exome sequencing
GBM	22	None	(64)	Whole exome sequencing
Colorectal	11	None	(85)	Whole exome sequencing
AML	1	None	(87)	Single-end whole-genome sequencing
AML	1	None	(88)	Paired-end whole-genome sequencing
Melanoma	1	None	(89)	Paired-end whole-genome sequencing
SCLC	1	None	(90)	Paired-end whole-genome sequencing

Supplementary Table 7. SPOP Mutations in Multiple Cohorts

Cohort	*SPOP* mutation prevalence	Technology	Mutated residues
WCMC	13.3 % (11/83)	WES, RNA-seq, Sanger	Y87N, Y87C, F102C, K129E, F133V, F133S, F133L, F133C
Uropath	10.1% (9/89)	WES	Y87N, S119N, F125L, W131G, F133S, F133L, F133C, K134N
UM	6.1% (3/49)	RNA-seq	F102C, F133L, F133V
UHZ	8.3% (16/193)	Sanger	Y87N, F102C, F102S, W131C, F133V, F133L
UW	14.5% (6/39)	Sanger	F102C, F102D, W131G, F133V

Supplementary Table 8. Somatic Copy Number Alterations Associated with *SPOP* Mutation

					Association with *SPOP* mutation:					
					Discovery set		Validation set		Combined	
Type	Cytoband	Area Coordinates (hg18)	Peak Coordinates (hg18)	q.values	P-value	Odds ratio	P-value	Odds ratio	P-value	Odds ratio
Del	6q15	chr6:84993071-90693659	chr6:89844069-89854517	7.60E-14	0	49.0936	0.0494	4.6152	0	14.9182
Del	5q14.3	chr5:86736848-92949317	chr5:90698354-90718446	4.76E-06	0	28.9982	0.0023	10.5654	0	18.5816
Del	6q21	chr6:102623781-118338352	chr6:107452805-107484929	2.72E-12	0	30.1055	0.0003	14.5979	0	23.0189
Del	5q21.1	chr5:96540739-99903527	chr5:98211239-98298151	6.24E-17	0	26.1837	0	23.6731	0	26.5255
Del	5q21.1	chr5:98288994-100177733	chr5:99892827-99956161	1.12E-11	0.0001	13.0406	0.0006	11.2936	0	12.8012
Del	6q14.1	chr6:79992719-89377471	chr6:82929067-83023893	3.50E-10	0.0001	10.9692	0.0427	4.4101	0	7.6868
Del	5q21.3	chr5:100266177-130524037	chr5:108694667-108775892	2.46E-06	0.0012	7.412	0.0001	20.5115	0	10.8653
Del	6q13	chr6:71718724-79638241	chr6:74188319-74228252	6.62E-05	0.0017	6.8441	0.2702	2.4264	0.0015	4.4487
Del	21q22.3	chr21:41569692-42033506	chr21:41754509-41805040	3.20E-13	0.0043	0	0.1836	0	0.0004	0
Del	5q11.2	chr5:54500232-57785740	chr5:55428291-55450265	3.84E-08	0.0044	5.4502	0.0107	7.6381	0.0004	4.9509
Amp	8q24.3	chr8:142130394-143948850	chr8:143649163-143686346	0.005918	0.0104	4.7801	0.4292	0.3188	0.1905	1.9013
Del	17p13.1	chr17:7099211-7923319	chr17:7723851-7759444	2.24E-07	0.0157	0.1101	0.6699	0.409	0.0089	0.1662
Del	2q21.1	chr2:114430192-140707944	chr2:131673017-131764875	0.0027275	0.0179	5.1272	0.0191	6.0238	0.0007	5.6709
Del	3p13	chr3:70098695-73514455	chr3:71884403-71891255	2.46E-06	0.0188	0	0.6731	0.4708	0.0179	0.1198
Amp	21q21.1	chr21:18980067-18997817	chr21:18980170-18994744	0.024377	0.0279	7.9899	1	0	0.144	2.9523
Amp	3q22.3	chr3:138503860-138525897	chr3:138503876-138514781	0.0046072	0.0284	3.8842	1	1.0223	0.1342	2.3893
Amp	7q11.23	chr7:71946015-71962929	chr7:71948417-71962425	0.076891	0.0333	4.1148	0.6154	1.7416	0.0554	2.9006
Amp	7q21.2	chr7:91212265-91215184	chr7:91212282-91215163	0.059607	0.0333	4.1148	0.6154	1.7416	0.0554	2.9006
Del	2q23.3	chr2:151939541-152050232	chr2:151970139-152045679	0.074786	0.0349	4.8599	0.65	1.4242	0.0829	2.9745
Del	17q21.31	chr17:39091591-40194083	chr17:39820952-39945190	1.17E-05	0.0366	0	0.1836	0	0.0026	0
Amp	7p15.2	chr7:26355665-26394413	chr7:26356137-26393733	0.059607	0.0435	3.7332	0.5879	2.2144	0.0554	2.9006

Supplementary Table 10. Mutation of *PIK3CA* and *PTEN* is Enriched in Locally Advanced Tumors
(A) Tumors are grouped by stage and mutational status of *PIK3CA* or *PTEN*. Only stage pT3 tumors displayed mutations in either gene (two-sided p = 0.011, Fisher's exact test).

(B) Mutations in *PTEN* or *PIK3CA* detected by exome sequencing. The hyper-mutated tumor, PR-00-1165, was included in this analysis and contained two canonical mutations in *PTEN*. Amino acids are numbered based on RefSeq protein ID NP_000305 for PTEN and NP_006209 for PIK3CA. All substitutions have been documented previously in prostate cancer or other cancer types in the Cosmic database (http://www.sanger.ac.uk/genetics/CGP/cosmic/). *, nonsense mutation.

A

Genotype:	Stage pT2	Stage pT3	Total
PIK3CA/PTEN mutant:	0	9	9
PIK3CA/PTEN wild-type:	44	59	103
Total:	44	68	112

B

PTEN mutations:	Reported in Cosmic?
K128N	Yes
R130Q	Yes
Y336*	Yes
G129R	Yes
R173H, R233*	Yes, Yes

PIK3CA mutations:	Reported in Cosmic?
p.H1047R	Yes
p.G118D	Yes
p.Q546P	Yes
p.Y1021H	Yes

Supplementary Table 11. Primer Sequences

Amplification	**sense (5' --> 3')**	**antisense (5' --> 3')**
SPOP		
Exons 6 and 7	TTCTATGGGGCCTGCATTT	CTCCACTTGGGGCTTTTTCT

Sequencing	**sense (5' --> 3')**	**antisense (5' --> 3')**
SPOP		
Exon 6	TTTTCTATCTGTTTTGGACAGG	CAAAGCCACAACTTGTCAGTG
Exon 7	TTTGCGAGTAAACCCCAAAG	CTCATCAGATCTGGGAACTGC

qPCR	**sense (5' --> 3')**	**antisense (5' --> 3')**
SPOP	CTTCTGCGAGGTGAGTGTTG	TCCCACAGTCCTCCTAACTCA
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG

Supplementary Table 12. BAC Probes Used for FISH

Assay	BAC
***ERG* break-apart**	
5' probe	RP11-372O17
3' probe	RP11-24A11
***ETV1* break-apart**	
5' probe	RP11-661L15
3' probe	RP11-79G16
***ETV4* break-apart**	
5' probe	RP11-147C10
3' probe	CTD-3215I16
***ETV5* break-apart**	
5' probe	RP11-822O23
3' probe	RP11-480B15
***PTEN* deletion**	
PTEN probe	CTD-2047N14
Reference probe	RP11-431P18