nature genetics

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

Christopher E Barbieri[1,2,16], Sylvan C Baca[3–5,16], Michael S Lawrence[3,16], Francesca Demichelis[6,7], Mirjam Blattner[1], Jean-Philippe Theurillat[3], Thomas A White[8], Petar Stojanov[3], Eliezer Van Allen[3,5], Nicolas Stransky[3], Elizabeth Nickerson[3], Sung-Suk Chae[1], Gunther Boysen[1], Daniel Auclair[3], Robert C Onofrio[3], Kyung Park[1], Naoki Kitabayashi[1], Theresa Y MacDonald[1], Karen Sheikh[1], Terry Vuong[1], Candace Guiducci[3], Kristian Cibulskis[3], Andrey Sivachenko[3], Scott L Carter[3], Gordon Saksena[3], Douglas Voet[3], Wasay M Hussain[1,6], Alex H Ramos[3,4], Wendy Winckler[3], Michelle C Redman[3], Kristin Ardlie[3], Ashutosh K Tewari[2], Juan Miguel Mosquera[1], Niels Rupp[9], Peter J Wild[9], Holger Moch[9], Colm Morrissey[10,11], Peter S Nelson[8,10,11], Philip W Kantoff[4,5], Stacey B Gabriel[3], Todd R Golub[3,12–14], Matthew Meyerson[3–5,14], Eric S Lander[3,4,15], Gad Getz[3,17], Mark A Rubin[1,2,17] & Levi A Garraway[3–5,14,17]

Prostate cancer is the second most common cancer in men worldwide and causes over 250,000 deaths each year[1]. Overtreatment of indolent disease also results in significant morbidity[2]. Common genetic alterations in prostate cancer include losses of *NKX3.1* (8p21)[3,4] and *PTEN* (10q23)[5,6], gains of *AR* (the androgen receptor gene)[7,8] and fusion of ETS family transcription factor genes with androgen-responsive promoters[9–11]. Recurrent somatic base-pair substitutions are believed to be less contributory in prostate tumorigenesis[12,13] but have not been systematically analyzed in large cohorts. Here, we sequenced the exomes of 112 prostate tumor and normal tissue pairs. New recurrent mutations were identified in multiple genes, including *MED12* and *FOXA1*. *SPOP* was the most frequently mutated gene, with mutations involving the SPOP substrate-binding cleft in 6–15% of tumors across multiple independent cohorts. Prostate cancers with mutant *SPOP* lacked ETS family gene rearrangements and showed a distinct pattern of genomic alterations. Thus, *SPOP* mutations may define a new molecular subtype of prostate cancer.

We performed exome capture followed by paired-end massively parallel sequencing on genomic DNA from 112 prostate adenocarcinomas and matched normal samples. We focused on treatment-naïve radical prostatectomy specimens from US and Australian subjects that spanned a range of grades, stages and risks of recurrence (Online Methods and **Supplementary Table 1**). The baits for exon capture targeted 98.2% of genes in the Consensus CDS (CCDS) database. A mean coverage depth of 118× per sample was achieved, with 89.2% of targets covered at a depth of ≥20× (**Supplementary Fig. 1** and **Supplementary Table 2**). DNA from tumors and normal samples was also analyzed by Affymetrix SNP 6.0 arrays to detect somatic copy-number alterations. In addition, transcriptome sequencing (RNA-seq) was performed on 22 exome-sequenced tumors and 41 independent samples (**Supplementary Fig. 2**).

We identified 5,764 somatic mutations that were present in DNA from tumors but absent in DNA from peripheral blood or non-cancerous prostate tissue (**Supplementary Table 3**). Of these variants, 997 occurred in a single tumor that harbored a frameshift mutation of the mismatch repair gene *MSH6* (**Supplementary Fig. 3**). Excluding this highly mutated sample, the remaining tumors harbored a median of 10 silent and 30 non-silent mutations (range 10 to 105 total mutations) or ~1.4 mutations per megabase of DNA covered (**Supplementary Fig. 3**). Analysis of 229 non-silent mutations by mass spectrometry genotyping validated 95.6% of variants with allelic fraction of ≥0.2 (confidence interval (CI) = 92–98%) (**Supplementary Table 3**). The mutation rate of this cohort exceeded that of seven published prostate tumor genomes (0.9 mutations per megabase)[14], perhaps because increased exome sequence coverage improved the detection of variants present at lower allelic fractions.

[1]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York, USA. [2]Department of Urology, Weill Cornell Medical College, New York, New York, USA. [3]The Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, Massachusetts, USA. [4]Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [6]Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York, USA. [7]Centre for Integrative Biology, University of Trento, Trento, Italy. [8]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [9]Institute of Surgical Pathology, University Hospital Zurich, Zurich, Switzerland. [10]Department of Medicine, University of Washington, Seattle, Washington, USA. [11]Department of Urology, University of Washington, Seattle, Washington, USA. [12]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA. [13]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [14]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [15]Department of Biology, MIT, Cambridge, Massachusetts, USA. [16]These authors contributed equally to this work. [17]These authors jointly directed this work. Correspondence should be addressed to M.A.R. (rubinma@med.cornell.edu) or L.A.G. (levi_garraway@dfci.harvard.edu).

Received 22 December 2011; accepted 19 April 2012; published online 20 May 2012; doi:10.1038/ng.2279

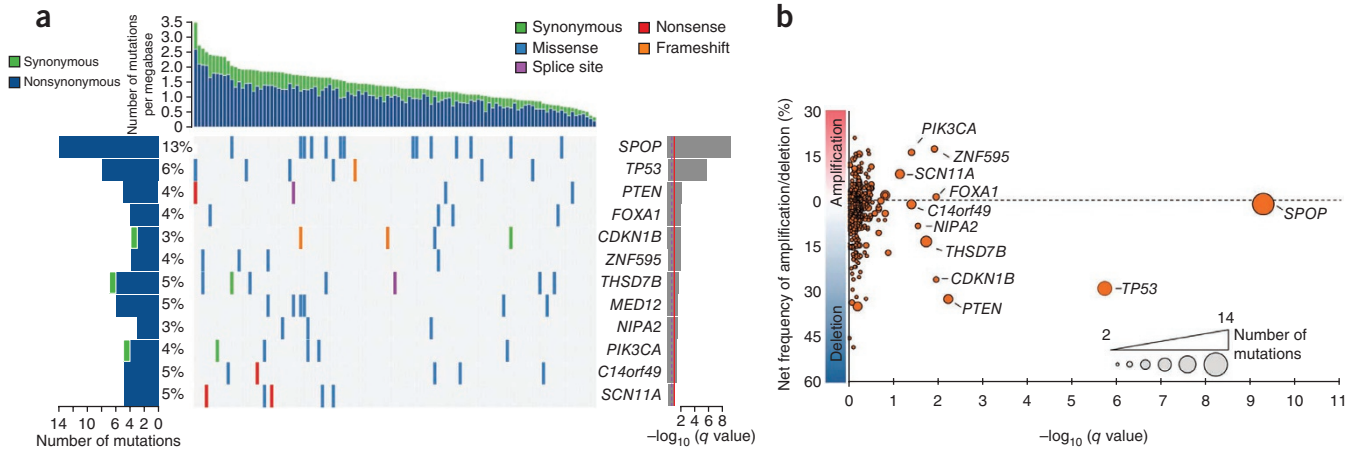

Figure 1 Significantly mutated genes in aggressive primary prostate cancer. (**a**) Top, a cohort of 111 primary prostate tumors is ordered by number of mutations per megabase sequenced. Center, mutations in significantly mutated genes, colored by the coding consequence of the mutation. Each column represents a tumor, and each row represents a gene. Left, number and percentage of tumors with mutations in a given gene. Right, the negative log of the *q* values for the significance level of mutated genes is shown (for all genes with $q < 0.1$). (**b**) Net frequency of gene deletion and/or amplification across 169 copy number–profiled tumors. Significantly mutated genes are indicated. Only autosomal genes with two or more mutations are shown.

We searched for genes that harbored more nonsynonymous mutations than expected by chance given gene size, sequence context and the frequency of mutations for each tumor (**Fig. 1a** and **Supplementary Table 4**). Twelve genes were enriched for mutations at *q* value < 0.1, the majority of which are highly expressed at the transcript level in prostate tumors (**Supplementary Fig. 4**). The identification of *PIK3CA*, *TP53* and *PTEN* confirmed that our approach detected alterations known to promote tumorigenesis in prostate cancer and other malignancies. We also found evidence of enrichment for mutations in the *PTEN* pathway, cell cycle regulatory machinery and other gene sets (**Supplementary Table 5**)[12].

The most frequently mutated gene was *SPOP* (in 13% of tumors; **Fig. 1**), which encodes the substrate-binding subunit of a Cullin-based E3 ubiquitin ligase[15,16]. Although isolated *SPOP* mutations have been reported in prostate cancer[14,17], this gene has not previously been found to be mutated at a significant frequency in any malignancy. Several genes not previously known to undergo somatic alteration in prostate cancer were enriched for mutations, including *FOXA1*, *MED12*, *THSD7B*, *SCN11A* and *ZNF595*. The *CDKN1B* gene encoding $p27^{Kip1}$ was somatically mutated in 3 samples and deleted in 16 others (**Fig. 1b**). $p27^{Kip1}$ constrains prostate tumor growth in mice[18] and *CDKN1B* harbors a germline prostate cancer risk allele[19], but somatic substitutions had not previously been observed in this cell cycle regulator. Infrequent mutations were also detected in multiple proto-oncogenes, tumor suppressors and chromatin-modifying enzymes (**Supplementary Table 4** and **Supplementary Note**).

The forkhead transcription factor gene *FOXA1* harbored non-silent mutations in 4 of 111 exomes and 4 of 41 independent RNA-seq samples. The FOXA1 protein is required for epithelial cell differentiation in the mouse prostate[20] and promotes cell cycle progression in castration-resistant prostate cancer[21]. Notably, FOXA1 modulates androgen receptor–driven transcription[22] and activates expression of *CDKN1B*[23]. Mutations strictly affected residues in the forkhead domain (**Supplementary Table 4**) that resides near the DNA-binding surface (**Fig. 2a**)[24]. The clustered nature of these mutations suggests that they may disrupt binding of FOXA1 to DNA targets.

Mutations affecting *MED12* were observed in 6 of 111 exomes, with a recurrent L1224F variant (encoded by c.3670C>T or c.3669_3670TC>GT) being identified in five samples (**Fig. 2b** and **Supplementary Table 4**). *MED12* encodes a subunit of the mediator complex and the cyclin-dependent kinase 8 (CDK8) subcomplex that regulates basal and stimulus-specific transcriptional programs[25–27]. Recently, *MED12* mutations were reported in 70% of uterine leiomyomas[28], benign stromal tumors of the smooth muscle lineage. Mutations in prostate cancer affected distinct codons from those in leiomyoma and occurred in epithelial cells rather than stroma, as determined by laser-capture microdissection (LCM) (**Supplementary Fig. 5**). *MED12* mutations may conceivably perturb CDK8-dependent modulation of transcriptional programs linked to p53 and androgen signaling[26,27].

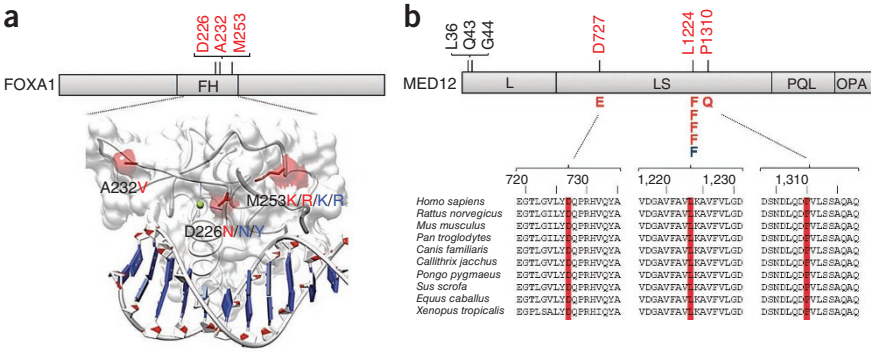

Figure 2 Recurrent somatic alterations in FOXA1 and MED12. Residues affected by mutations detected by exome sequencing are depicted (red), as are variants identified in non-overlapping transcriptome sequencing data (blue). (**a**) Structural analysis of alterations in FOXA1. Mutated residues are mapped to the structure of the HNF3γ forkhead domain (FH) (Protein Data Bank (PDB) coordinate file 1VTN)[24] and are highlighted in red. (**b**) Recurrent MED12 alterations in prostate cancer (red) are distinct from those reported in uterine leiomyoma (black)[28]. Domains of MED12 are denoted as previously[25]; L, leucine-rich domain; LS, leucine- and serine-rich domain; PQL, proline-, glutamine-, leucine-rich domain; OPA, opposite paired domain. Multispecies conservation of the mutated sites is shown below each alteration in prostate cancer.

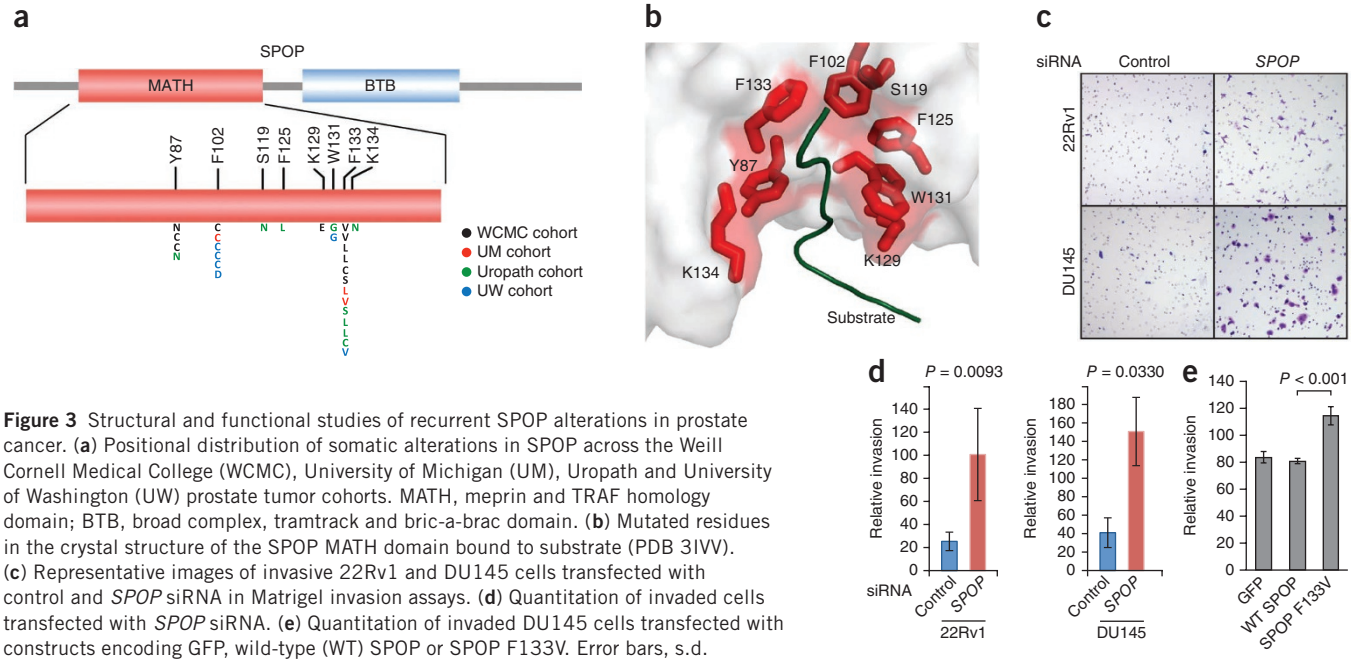

Figure 3 Structural and functional studies of recurrent SPOP alterations in prostate cancer. (**a**) Positional distribution of somatic alterations in SPOP across the Weill Cornell Medical College (WCMC), University of Michigan (UM), Uropath and University of Washington (UW) prostate tumor cohorts. MATH, meprin and TRAF homology domain; BTB, broad complex, tramtrack and bric-a-brac domain. (**b**) Mutated residues in the crystal structure of the SPOP MATH domain bound to substrate (PDB 3IVV). (**c**) Representative images of invasive 22Rv1 and DU145 cells transfected with control and *SPOP* siRNA in Matrigel invasion assays. (**d**) Quantitation of invaded cells transfected with *SPOP* siRNA. (**e**) Quantitation of invaded DU145 cells transfected with constructs encoding GFP, wild-type (WT) SPOP or SPOP F133V. Error bars, s.d.

Although *SPOP* mutations were originally reported in genomic studies of prostate cancer (**Supplementary Table 6**)[14,17], their prevalence and functional relevance remained unknown. We therefore sequenced this gene in multiple additional cohorts comprising over 300 primary tumors and metastases from individuals in the United States and Europe. Using RNA-seq and Sanger sequencing of tumor and matched germline DNA, recurrent heterozygous *SPOP* substitutions were identified in 6–13% of primary prostate adenocarcinomas (**Fig. 3a**, **Supplementary Fig. 6** and **Supplementary Table 7**). No mutations were identified in 36 benign prostate tissue samples, 5 samples of prostate stroma or 6 common prostate cell lines. *SPOP* mutations were also found in 6 of 41 individuals with metastatic disease (14.5%) (**Fig. 3a** and **Supplementary Table 7**). Thus, *SPOP* mutations occurred at a frequency of 6–15% across localized and advanced prostate tumors.

All *SPOP* mutations affected conserved residues in the structurally defined substrate-binding cleft (**Fig. 3b** and **Supplementary Fig. 7**)[16]. Several recurrently altered residues have key roles in substrate interaction; moreover, alteration of Tyr87, Trp131 and Phe133 disrupted substrate binding *in vitro*[16]. These results strongly suggest that *SPOP* mutations in prostate cancer are biologically consequential. To test this hypothesis, we examined the consequences of mutant SPOP protein expression or *SPOP* knockdown on tumorigenic phenotypes *in vitro*. Prostate cancer cells transfected with the most common SPOP mutant (the F133V variant) or with *SPOP* siRNA showed increased invasion compared to controls (**Fig. 3c–e** and **Supplementary Fig. 8**), but cell growth and viability were largely unaltered (**Supplementary Fig. 9**). The SPOP-CUL3 complex affects a variety of substrates that participate in multiple pathways, including hedgehog, c-Jun N-terminal kinase (JNK) and steroid receptor signaling cascades[29–31]. *SPOP* undergoes amplification in other malignancies and may be overexpressed in renal cell carcinoma[29]; however, multiple prostate cancer cohorts showed no evidence of *SPOP* amplification or upregulation (**Fig. 1b** and **Supplementary Figs. 10** and **11**). Prostate cancer–associated *SPOP* mutations may conceivably result in *de novo* gain-of-function alterations (for example, interaction with a distinct substrate profile), dominant-negative effects or more subtle alterations in substrate specificity. Further studies are necessary to determine the specific ubiquitin ligase functions and cellular pathways deregulated by *SPOP* mutation in prostate cancer.

Notably, all exomes with *SPOP* mutations lacked the *TMPRSS2-ERG* fusion or other ETS family gene rearrangements present in up to 50% of prostate cancers (**Fig. 4** and **Supplementary Fig. 12**)[11,32]. This mutually exclusive relationship between *SPOP* mutation and *ERG* rearrangement ($P < 0.001$, Fisher's exact test) was confirmed in evaluable samples across all five cohorts tested (**Supplementary Fig. 12**), even within an individual prostate tumor (**Supplementary Fig. 13**). Thus, *SPOP* mutation and ETS family gene fusions may represent early and divergent driver events in prostate carcinogenesis. *SPOP* mutations were identified in LCM-analyzed high-grade intraepithelial neoplasia (HG-PIN) adjacent to invasive adenocarcinoma, providing further support for the idea that *SPOP* mutation comprises an early event in prostate tumorigenesis (**Supplementary Fig. 14**).

In light of previous studies suggesting that prostate cancer may be classified by co-occurring genomic alterations[12,33,34], we investigated whether tumors with *SPOP* mutations were enriched for other genomic lesions (**Fig. 4**). Recurrent somatic deletions at 5q21 and 6q21 were found to be enriched in these tumors ($P = 1.4 \times 10^{-11}$ and 3.4×10^{-7}, respectively, Fisher's exact test), both in the cohort used for whole-exome sequencing and in an independent prostate cancer collection (**Fig. 4** and **Supplementary Table 8**). Thus, loss of tumor suppressor genes in these regions may cooperate with *SPOP* mutation to promote tumorigenesis. The relevant 5q21 locus contains *CHD1*, which encodes a chromatin-modifying enzyme that also undergoes disruptive rearrangements in prostate cancer[14]. The 6q21 region encompasses several genes, including *FOXO3*, a *FOXA1* homolog that has previously been implicated in prostate carcinogenesis and progression[35], and *PRDM1*, a tumor suppressor in lymphoma[36]. In contrast, *TP53* lesions were generally absent in tumors with *SPOP* mutations ($P = 0.015$, Fisher's exact test), despite the fact that this tumor suppressor was recurrently mutated and deleted (**Fig. 1**). *SPOP* mutations also trended inversely with point mutations and/or copy-number loss involving the *PTEN* locus in primary tumors ($P = 0.044$, Fisher's exact test) (**Fig. 4**); this pattern was supported by FISH analysis for *PTEN* deletion

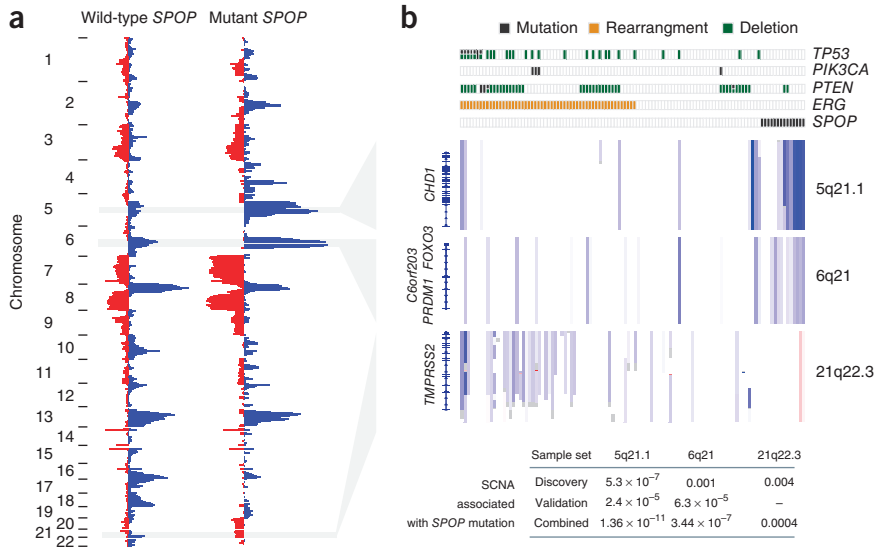

Figure 4 *SPOP* mutation defines a distinct genetic subclass of prostate cancer. (**a**) Frequency of genomic copy-number alterations in tumors with mutant and wild-type *SPOP*. The height of the peaks reflects the frequency of copy-number loss (blue) or gain (red). (**b**) Heatmap showing selected recurrent somatic copy-number aberrations (SCNAs). Each column represents a single prostate cancer sample. Samples are annotated for mutations in *SPOP*, *PTEN*, *PIK3CA* and *TP53*, deletions of *PTEN* and *ERG* rearrangements. Deletions positively correlated (5q21.1, 6q21) or inversely correlated (21q22.3) with *SPOP* mutation are shown. *P* values of peak association with *SPOP* mutation in both discovery and validation cohorts are given below (Fisher's exact test). Regions are not to scale; full coordinates are available in **Supplementary Table 8**.

(**Supplementary Fig. 15**). Tumors with *SPOP* mutations also lacked *PIK3CA* mutations (**Fig. 4**). Although the inverse relationship between *SPOP* mutations and *PTEN* and *PIK3CA* alterations was observed in primary tumors ($P = 0.041$, Fisher's exact test), these events co-occurred more frequently in metastatic tumors (**Supplementary Fig. 15**). Further studies are needed to determine whether these genetic relationships also occur in other populations of affected individuals and to elucidate the biological interactions that may underlie this phenomenon. Taken together, these results suggest that *SPOP* mutations may anchor a distinct genetic subtype of ETS fusion–negative cancers.

In summary, whole-exome sequencing has identified genes that are recurrently mutated in prostate cancer. These efforts have also revealed a distinct ETS fusion–negative subclass of prostate cancer characterized by recurrent *SPOP* mutations and enriched for both 5q21 and 6q21 deletions. In the future, this expanded genetic framework may articulate new mechanisms of carcinogenesis that inform both disease modeling and patient stratification for clinical trials of experimental agents. Together with additional comprehensive analyses of the prostate cancer genome, epigenome and transcriptome, these systematic approaches should illuminate the landscape of alterations that underlie disease biology and therapeutic vulnerability in this common and clinically heterogeneous malignancy.

URLs. CCDS, http://www.ncbi.nlm.nih.gov/CCDS/; Protein Data Bank, http://www.pdb.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Binary sequence alignment/map (BAM) files and SNP array data were deposited in the database of Genotypes and Phenotypes (dbGaP; phs000447.v1.p1).

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful for the assistance of members of the Broad Institute Biological Samples Platform, Genetic Analysis Platform and Genome Sequencing Platform. We thank S. Banerjee for computational assistance, R. Kim and R. Leung for their critical contributions to the Weill Cornell Prostate Cancer Tumor Bank, P. Schraml, S. Dettwiler and M. Storz for assistance with the University Hospital Zurich cohort and biobank and the members of the University of Washington Rapid Autopsy Program. We thank the University of Michigan Prostate Cancer Special Program of Research Excellence (SPORE) (K. Pienta) for sample contribution to this study. We are also grateful to the individuals with prostate cancer and the families who contributed to these studies. This work was supported by the US National Human Genome Research Institute (NHGRI) Large Scale Sequencing Program (U54 HG003067 to the Broad Institute, E.S.L.), the Kohlberg Foundation (L.A.G.), the Starr Cancer Consortium (M.A.R., F.D. and L.A.G), the Prostate Cancer Foundation (M.A.R.), US Department of Defense Synergy Awards (PC101020 to F.D., L.A.G. and M.A.R. and PC093372 to T.W. and P.S.N.), the Dana-Farber/Harvard Cancer Center Prostate Cancer SPORE (US National Institutes of Health (NIH) P50 CA090381), a New Investigator Award (PC094516 to F.D.), the US National Cancer Institute, Early Detection Research Network (U01CA111275 and NCI EDRN to F.D. and M.A.R.), the US National Cancer Institute (R01 CA125612 to F.D. and M.A.R.), the Pacific Northwest Prostate Cancer SPORE (P50CA097186 to C.M. and P.S.N.), the Swiss Science Foundation (PASMP3_134379/1 to J.-P.T.) and a US NIH Director's New Innovator Award (DP2OD002750 to L.A.G.). S.C.B. is supported by a Medical Scientist Training Program (MSTP) grant from the US NIH. C.E.B. is supported by a Prostate Cancer Foundation Young Investigator Award.

AUTHOR CONTRIBUTIONS

S.C.B., M.S.L., P.S., W.M.H., E.V.A., N.S., K.C., A.S., S.L.C., G.S., D.V. and A.H.R. performed computational analyses. M.B., J.-P.T., T.A.W., S.-S.C., K.S., G.B., T.Y.M., K.P. and T.V. designed and performed experiments. E.N., D.A., R.C.O., C.G., W.W., M.C.R., K.A. and S.B.G. processed samples and supervised exome sequencing. J.M.M., K.P., N.K., A.K.T., N.R., P.J.W., H.M., C.M. and P.S.N. coordinated sample acquisition, processing, pathologic review and analysis. C.E.B., S.C.B., F.D., P.W.K., T.R.G., M.M., E.S.L., G.G., M.A.R. and L.A.G. designed the study. C.E.B., S.C.B., F.D., G.G., M.A.R. and L.A.G. analyzed the data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2279.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61**, 69–90 (2011).
2. Daskivich, T.J. *et al.* Overtreatment of men with low-risk prostate cancer and significant comorbidity. *Cancer* **117**, 2058–2066 (2011).
3. He, W.W. *et al.* A novel human prostate-specific, androgen-regulated homeobox gene (*NKX3.1*) that maps to 8p21, a region frequently deleted in prostate cancer. *Genomics* **43**, 69–77 (1997).
4. Bhatia-Gaur, R. *et al.* Roles for *Nkx3.1* in prostate development and cancer. *Genes Dev.* **13**, 966–977 (1999).
5. Li, J. *et al. PTEN*, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**, 1943–1947 (1997).
6. Cairns, P. *et al.* Frequent inactivation of *PTEN/MMAC1* in primary prostate cancer. *Cancer Res.* **57**, 4997–5000 (1997).
7. Linja, M.J. & Visakorpi, T. Alterations of androgen receptor in prostate cancer. *J. Steroid Biochem. Mol. Biol.* **92**, 255–264 (2004).

8. Visakorpi, T. *et al. In vivo* amplification of the androgen receptor gene and progression of human prostate cancer. *Nat. Genet.* **9**, 401–406 (1995).
9. Perner, S. *et al. TMPRSS2:ERG* fusion–associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res.* **66**, 8337–8341 (2006).
10. Tomlins, S.A. *et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595–599 (2007).
11. Tomlins, S.A. *et al.* Recurrent fusion of *TMPRSS2* and *ETS* transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
12. Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
13. Kumar, A. *et al.* Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17087–17092 (2011).
14. Berger, M.F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
15. Nagai, Y. *et al.* Identification of a novel nuclear speckle-type protein, SPOP. *FEBS Lett.* **418**, 23–26 (1997).
16. Zhuang, M. *et al.* Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases. *Mol. Cell* **36**, 39–50 (2009).
17. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
18. Majumder, P.K. *et al.* A prostatic intraepithelial neoplasia–dependent $p27^{Kip1}$ checkpoint induces senescence and inhibits cell proliferation and cancer progression. *Cancer Cell* **14**, 146–155 (2008).
19. Kibel, A.S. *et al. CDKN1A* and *CDKN1B* polymorphisms and risk of advanced prostate carcinoma. *Cancer Res.* **63**, 2033–2036 (2003).
20. Gao, N. *et al.* Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431–3443 (2005).
21. Zhang, C. *et al.* Definition of a FoxA1 cistrome that is crucial for G1 to S-phase cell-cycle transit in castration-resistant prostate cancer. *Cancer Res.* **71**, 6738–6748 (2011).
22. Gao, N. *et al.* The role of hepatocyte nuclear factor–3α (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol. Endocrinol.* **17**, 1484–1507 (2003).
23. Williamson, E.A. *et al.* BRCA1 and FOXA1 proteins coregulate the expression of the cell cycle–dependent kinase inhibitor $p27^{Kip1}$. *Oncogene* **25**, 1391–1399 (2006).
24. Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
25. Zhou, R. *et al.* SOX9 interacts with a component of the human thyroid hormone receptor–associated protein complex. *Nucleic Acids Res.* **30**, 3245–3252 (2002).
26. Wang, Q., Sharma, D., Ren, Y. & Fondell, J.D. A coregulatory role for the TRAP-mediator complex in androgen receptor–mediated gene expression. *J. Biol. Chem.* **277**, 42852–42858 (2002).
27. Donner, A.J., Szostek, S., Hoover, J.M. & Espinosa, J.M. CDK8 is a stimulus-specific positive coregulator of p53 target genes. *Mol. Cell* **27**, 121–133 (2007).
28. Mäkinen, N. *et al. MED12*, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science* **334**, 252–255 (2011).
29. Liu, J. *et al.* Analysis of *Drosophila* segmentation network identifies a JNK pathway factor overexpressed in kidney cancer. *Science* **323**, 1218–1222 (2009).
30. Wang, C., Pan, Y. & Wang, B. Suppressor of fused and Spop regulate the stability, processing and function of Gli2 and Gli3 full-length activators but not their repressors. *Development* **137**, 2001–2009 (2010).
31. Li, C. *et al.* Tumor-suppressor role for the SPOP ubiquitin ligase in signal-dependent proteolysis of the oncogenic co-activator SRC-3/AIB1. *Oncogene* **30**, 4350–4364 (2011).
32. Mosquera, J.M. *et al.* Prevalence of *TMPRSS2-ERG* fusion prostate cancer among men undergoing prostate biopsy in the United States. *Clin. Cancer Res.* **15**, 4706–4711 (2009).
33. Demichelis, F. *et al.* Distinct genomic aberrations associated with *ERG* rearranged prostate cancer. *Genes Chromosom. Cancer* **48**, 366–380 (2009).
34. Lapointe, J. *et al.* Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res.* **67**, 8504–8510 (2007).
35. Yang, J.Y. & Hung, M.C. A new fork for clinical application: targeting forkhead transcription factors in cancer. *Clin. Cancer Res.* **15**, 752–757 (2009).
36. Mandelbaum, J. *et al. BLIMP1* is a tumor suppressor gene frequently disrupted in activated B cell–like diffuse large B cell lymphoma. *Cancer Cell* **18**, 568–579 (2010).

ONLINE METHODS

DNA extraction and exome sequencing. Slides for hematoylin and eosin staining were cut from all frozen tissue blocks and examined by a board-certified pathologist to select for high-density cancer foci with <10% stroma or other non-cancerous material to ensure high purity of cancer DNA. Biopsy cores were then taken from the corresponding frozen tissue block for DNA extraction. From each sample, 25–30 mg of tissue was homogenized. DNA was extracted from homogenate and quantified using Picogreen dsDNA Quantitation Reagent (Invitrogen). Samples were separated on agarose gels (E-Gel, Invitrogen) to assess structural integrity. All DNA samples were stored at −20 °C. Ancestry analysis of sequenced samples is shown in **Supplementary Figure 16**.

Whole-exome capture libraries were constructed from 100 ng of DNA from tumors and normal tissue after sample shearing, end repair, phosphorylation and ligation to barcoded sequencing adaptors[37]. Ligated DNA was size selected for lengths between 200–350 bp and subjected to exonic hybrid capture using SureSelect v2 Exome bait (Agilent). Samples were multiplexed and sequenced on multiple Illumina HiSeq flow cells for average target exome coverage of 118×.

Copy-number analysis. DNA from tumors and normal tissue was analyzed by Affymetrix SNP 6.0 arrays to detect regions of somatic copy-number alteration (**Supplementary Table 9**). Quality control, segmentation and copy-number analysis were performed as previously described[33], with one additional step aimed at diminishing the number of recurrent lesions possibly caused by germline signal: we applied the same detection pipeline to the DNA samples from normal tissue alone. All peaks detected in both analyses were excluded from the list of recurrent somatic copy-number aberrations. The remaining lesions with Genomic Identification of Significant Targets in Cancer (GISTIC) q value of <0.1 were included in analysis of copy-number alterations associated with mutated genes. Two-tailed Fisher's exact tests were applied for all association tests.

Sequence data processing and quality control. Exome sequence data processing and analysis were performed using pipelines at the Broad Institute[14,38,39]. A BAM file aligned to the hg19 human genome build was generated from Illumina sequencing reads for each tumor and normal sample by the Picard pipeline. The Firehose pipeline was used to manage input and output files and submit analyses for execution by GenePattern[40].

Quality control modules in Firehose were used to compare genotypes derived from Affymetrix arrays and sequencing data to ensure concordance. Genotypes from SNP arrays were also used to monitor for low levels of cross-contamination between samples from different individuals in sequencing data using the ContEst algorithm[41].

Mutation calling and identification of significantly mutated genes. The MuTect algorithm from the Broad Institute Genome Analysis Toolkit was used to identify somatic single-nucleotide variants (SSNVs) in targeted exons[39,42]. MuTect identifies candidate SSNVs by Bayesian statistical analysis of bases and their qualities in the tumor and normal BAM files at a given genomic locus. We required a minimum of 14 reads covering a site in the tumor and 8 reads in the normal sample for mutation calling. We determined the lowest allelic fraction at which SSNVs could be detected on a per-sample basis, using estimates of cross-contamination from the ContEst pipeline[41]. Small somatic insertions and deletions (indels) were detected using the Indelocator algorithm after local realignment of sequences from tumors and normal samples[42].

The MutSig algorithm from the Broad Institute was applied to identify genes or gene sets that were significantly enriched for mutations, given sequence context and genomic territory[14,39,42]. For each gene, we calculated the probability of detecting the observed constellation of mutations or a more extreme one, given the background mutation rates calculated across the data set. This P value was then adjusted for multiple hypotheses testing (Benjamini-Hochberg procedure) to obtain a q value. Mutations in significantly mutated genes and the others described were manually reviewed by examination of BAM files in the Integrative Genomics Viewer (IGV)[43].

Validation of selected mutations by mass spectrometry genotyping. We chose 240 non-silent mutations (231 SSNVs and 9 indels) across 48 pairs of tumor and normal samples to validate by mass spectrometry genotyping using the iPLEX platform (Sequenom). We targeted 74 mutations in significantly mutated genes or gene sets with q value of <0.1 and mutations reported in COSMIC. The remaining 166 non-silent mutations were chosen at random. Because the rate of validation using this technology is significantly lower when the mutant allele is present at low allelic fraction[14,38], we attempted to validate only mutations with allelic fraction of ≥0.2 (where 20% of sequence reads from the tumor contained the mutation).

Of the 240 assays attempted, 228 gave successful genotype calls, and 218 somatic mutations were confirmed (**Supplementary Table 3**). All events called in the tumor were absent from the corresponding normal sample. The overall accuracy for mutation calling was 95.6% (Clopper-Pearson 95% CI = 92–98%), in close agreement with previous studies[14,38,39].

Mutation annotation. Somatic point mutations and indels were annotated using information from publicly available databases, including the UCSC Genome Browser's UCSC Genes track[44], miRBase release 15 (ref. 45), dbSNP build 132 (ref. 46), UCSC Genome Browser's ORegAnno track[47], UniProt release 2011_03 (ref. 48) and COSMIC v51 (ref. 49).

Mutations were classified as heterozygous or homozygous using the ABSOLUTE algorithm[50]. ABSOLUTE integrates genome-wide copy-number data from SNP arrays and the allelic fraction values of somatic mutations to model gene and mutation copy number in a tumor. Only a subset of tumors contained sufficient copy-number alterations to permit analysis. Mutations in these tumors were annotated as heterozygous or homozygous and clonal or subclonal. Mutation multiplicity, equal to the average number of copies of a mutation per cancer cell, was also calculated (**Supplementary Figs. 17** and **18** and **Supplementary Table 10**).

RNA extraction and RNA-seq sample prep, sequencing and processing. RNA was extracted from the frozen cancer tissue using TRIzol (Invitrogen), according to the manufacturer's protocol. Total RNA was prepared in accordance with Illumina's sample preparation protocol for paired-end sequencing of mRNA, as previously described[51]. Paired-end reads were aligned to the human genome (hg18) using ELAND, as previously described[51]. Data were visualized using IGV[43], and candidate mutations were identified in *SPOP* coding regions.

DNA extraction and *SPOP* genotyping. DNA was extracted using phenol-chloroform and purified by an ethanol precipitation method, as previously described[14]. Direct Sanger sequencing of putative *SPOP* somatic mutations in all tumor-blood pairs was performed by standard methods, following PCR amplification using specific primers. The sequences of the primers used for amplifying and sequencing *SPOP* are given (**Supplementary Table 11**).

Laser capture microdissection. Tissue sections (5 μm thick) were cut, fixed and stained on membrane-coated slides and were dissected with the Arcturus*XT* LCM Instrument (Life Technologies). Tissue staining and LCM were performed as described[52]. A combination of inferred capture and UV laser cutting was carried out to best recover a precise subset of cells. DNA was amplified with the Whole Genome Amplification kit (WGA4), as suggested by the manufacturer for the single-cell approach (Sigma-Aldrich). Standard PCR was used for targeted enrichment of *SPOP* exons 6 and 7, and Sanger sequencing was performed.

FISH. *ETS* rearrangement status and *PTEN* deletion status were assessed on tissue slides from the same tumor nodule used for RNA and DNA extraction. FISH methods to detect *TMPRSS2-ETS* gene fusions have been previously described[9,11]. We used *ERG*, *ETV1*, *ETV4* and *ETV5* break-apart FISH assays to confirm gene rearrangement on the DNA level[53]. To assess the status of *PTEN*, we used a locus-specific probe and a reference probe, as previously described[14]. The sequences of all FISH probes are listed (**Supplementary Table 12**).

Quantitative RT-PCR. RNA was extracted using TRIzol reagent, subjected to DNase treatment (DNA-free Kit; Applied Biosystems) according to the manufacturer's instructions and used in quantitative RT-PCR. Quantitative RT-PCR was performed using the ABI 7500 Real-Time PCR System (Applied Biosystems), following the manufacturer's RNA-to-CT 1-step protocol. Each target was run in triplicate, and expression levels relative to those of the housekeeping gene *GAPDH*

doi:10.1038/ng.2279

were determined on the basis of the comparative threshold cycle method ($2^{-\Delta\Delta CT}$). The sequences for primers used in these experiments are given (**Supplementary Table 9**). All experiments were run in triplicate; results are representative of three independent experiments.

Immunohistochemistry. *ERG* rearrangement status was confirmed by immunohistochemistry, as previously described[54]. Briefly, primary rabbit monoclonal antibody to ERG was obtained from Epitomics (2805-1). Antigen recovery was conducted using heat retrieval and CC1 standard, a high pH Tris/borate/EDTA buffer (VMSI, 950-124). Slides were incubated with a 1:100 dilution of the primary antibody to ERG for 1 h at room temperature. Primary antibody was detected using the ChromoMap DAB detection kit (VMSI, 760-159) and UltraMap horseradish peroxidase (HRP)-conjugated secondary antibody to rabbit (VMSI, 760-4315). The secondary antibody was applied for 16 min at room temperature. Slides were counterstained with Hematoxylin II (VMSI, 790-2208) for 8 min and with Bluing Reagent (VMSI, 760-2037) for 4 min at 37 °C. Subjective evaluation of ERG protein expression was scored as positive or negative by pathologists for the study.

Plasmids expressing SPOP. Wild-type *SPOP* was obtained from Origene in a mammalian expression vector encoding C-terminal Myc and Flag tags. The sequence encoding SPOP F133V was created using the QuikChange II Site-Directed Mutagenesis kit (Agilent). The sequences of all plasmids were confirmed by Sanger sequencing, and protein expression was confirmed by protein blot using antibodies to Myc (Upstate, 05-724), and Flag (Sigma, F3165).

Cell culture and transfection. The human prostate cancer cell lines 22Rv1 and DU145 and the benign prostate cell line RWPE were obtained from the American Type Culture Collection (ATCC). The 22Rv1 and DU145 cells were maintained in RPMI 1640 (Invitrogen) supplemented with 10% FBS and penicillin/streptomycin. RWPE cells were maintained in Keratinocyte-SFM (Invitrogen) supplemented with human recombinant epidermal growth factor (Invitrogen) and bovine pituitary extract (Invitrogen), according to the manufacturer's instructions.

For siRNA transfection, RWPE (2.5×10^5 per well), 22Rv1 (4×10^5 per well) and DU145 (2×10^5 per well) cells were seeded on 6-well tissue culture plates. The next day, cells were transfected with 100 nM *SPOP* or non-targeting (control) siRNAs (ON-TARGETplus, Thermo Scientific) using Dharmfect 2 reagent (Invitrogen) according to the manufacturer's instructions. For plasmid transfection, DU145 (4×10^5 per well) cells were seeded on 6-well tissue culture plates. The next day, cells were transfected with 4 ug of pCMV6 encoding wild-type SPOP or pCMV encoding SPOP F133V using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

Cell viability and proliferation assays. The 22Rv1 (2×10^3 per well) and DU145 (1×10^3 per well) cells transfected with control or *SPOP* siRNA or with plasmids expressing SPOP were seeded in 96-well tissue culture plates. Cell viability and growth was determined by performing WST-1 assays (Roche), reading absorbance at 450 nm according to the manufacturer's instructions. Values from three wells were obtained for each treatment and time point. Results are representative of three independent experiments.

Invasion assays. For invasion assays, 7.5×10^4 22Rv1 and 5×10^4 DU145 cells transfected with control or *SPOP* siRNA or with plasmids expressing SPOP were resuspended in 0.5 ml of RPMI-1640 medium containing 1% FBS and placed into the top chamber of Matrigel-coated 8-μm Transwell inserts (BD Falcon). The bottom wells contained RPMI supplemented with 5–10% FBS. After 24 h (DU145) or 48 h (22Rv1), the filters were fixed and stained with 0.5% crystal violet for 30 min, and cells on the upper surface of the filters were removed with a cotton swab. Migrated cells were quantified by counting the numbers of cells that penetrated the membrane in four microscopic fields (viewed at 20× magnification) per filter. All experiments were performed in triplicate; results are representative of three independent experiments.

37. Fisher, S. *et al.* A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).
38. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
39. Chapman, M.A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
40. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
41. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27**, 2601–2602 (2011).
42. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
43. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
44. Fujita, P.A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011).
45. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
46. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
47. Griffith, O.L. *et al.* ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res.* **36**, D107–D113 (2008).
48. UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **39**, D214–D219 (2011).
49. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
50. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* published online, doi:10.1038/nbt.2203 (29 April 2012).
51. Pflueger, D. *et al.* Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. *Genome Res.* **21**, 56–67 (2011).
52. Espina, V. *et al.* Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603 (2006).
53. Svensson, M.A. *et al.* Testing mutual exclusivity of ETS rearranged prostate cancer. *Lab. Invest.* **91**, 404–412 (2011).
54. Park, K. *et al.* Antibody-based detection of *ERG* rearrangement–positive prostate cancer. *Neoplasia* **12**, 590–598 (2010).