HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 October 02.

Published in final edited form as:

Nat Genet. 2018 May ; 50(5): 645–651. doi:10.1038/s41588-018-0078-z.

The long tail of oncogenic drivers in prostate cancer

Joshua Armenia[#1,2], **Stephanie A.M. Wankowicz**[#3,4], **David Liu**[#3,4], **Jianjiong Gao**[1,2], **Ritika Kundra**[1,2], **Ed Reznik**[1,2], **Walid K. Chatila**[1,2], **Debyani Chakravarty**[1,2], **G. Celine Han**[3,4], **Ilsa Coleman**[5], **Bruce Montgomery**[6], **Colin Pritchard**[7], **Colm Morrissey**[8], **Christopher E. Barbieri**[9], **Himisha Beltran**[10,11,12], **Andrea Sboner**[9], **Zafeiris Zafeiriou**[13], **Susana Miranda**[13], **Craig M. Bielski**[1,2], **Alexander V. Penson**[1,2], **Charlotte Tolonen**[4], **Franklin W. Huang**[3,4], **Dan Robinson**[14], **Yi Mi Wu**[14], **Robert Lonigro**[14], **Levi A. Garraway**[3,4], **Francesca Demichelis**[15], **Philip W. Kantoff**[16], **Mary-Ellen Taplin**[3], **Wassim Abida**[16], **Barry S. Taylor**[1,2,17], **Howard I. Scher**[16], **Peter S. Nelson**[5,6], **Johann S. de Bono**[13], **Mark A. Rubin**[9,11,12], **Charles L. Sawyers**[1], **Arul M. Chinnaiyan**[14], **PCF/SU2C International Prostate Cancer Dream Team**, **Nikolaus Schultz**[#1,2,17], and **Eliezer M. Van Allen**[#3,4]

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY.

[2]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY.

[3]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA.

[4]Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA.

[5]Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA.

[6]Department of Medicine, University of Washington, Seattle, WA.

[7]Department of Laboratory Medicine, University of Washington, Seattle, WA.

[8]Department of Urology, University of Washington, Seattle, WA.

[9]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY.

[10]Department of Medicine, Division of Medical Oncology, Weill Cornell Medicine, New York, NY.

CORRESPONDING AUTHORS: Eliezer M. Van Allen, MD, Dana-Farber Cancer Institute 450 Brookline Avenue, Smith 1036B Boston, MA 02215 eliezerm_vanallen@dfci.harvard.edu; Nikolaus Schultz, PhD Marie-Josée and Henry R. Kravis Center for Molecular Oncology Department of Epidemiology & Biostatistics Memorial Sloan Kettering Cancer Center 1275 York Ave New York, NY 10065 schultzn@mskcc.org.

AUTHOR CONTRIBUTIONS

J.A., S.A.M.W., N.S., E.M.V.A., D.L., J.G., R.K., E.R., W.K.C., D.C., G.C.H., C.E.B., A.S., C.M.B., A.V.P., C.T., F.D., M.A.R., B.S.T. contributed with algorithm development and analysis of genomic data. I.C., B.M., C.P., C.M., H.B., Z.Z., S.M., F.W.H., D.R., Y.M.W., P.W.K., M.-E.T., W.A., H.I.S., P.S.N., J.S.d.B., M.A.R., C.L.S., A.M.C. developed the patient cohort, obtained tumor biopsies, performed molecular testing for metastatic cases, and data interpretation of the overall cohort. J.A., S.A.M.W., D.L., N.S., and E.M.V.A. performed final aggregate cohort assembly, mutation review, interpretation, and manuscript preparation.

COMPETING FINANCIAL INTEREST

Dr. Van Allen is a consultant for Tango Therapeutics and Genome Medical.

[11]Englander Institute for Precision Medicine, Weill Cornell Medical College-New York Presbyterian Hospital. New York, NY.

[12]Sandra and Edward Meyer Cancer Center at Weill Cornell Medical College

[13]Biomarkers Team, Division of Clinical Studies, The Institute of Cancer Research and Royal Marsden Hospital, London, UK.

[14]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI.

[15]Centre for Integrated Biology, University of Trento, Trento, Italy.

[16]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY.

[17]Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY.

[#] These authors contributed equally to this work.

Abstract

Comprehensive genomic characterization of prostate cancer has identified recurrent alterations in androgen signaling, DNA repair, and PI3K among others. However, larger and uniform genomic analysis may reveal additional recurrently mutated genes at lower frequencies. Here we aggregate and uniformly analyze exome sequencing data from 1013 prostate cancers. We identify and validate a new class of E26 transformation-specific (ETS) fusion negative tumors defined by mutations in epigenetic regulators, as well as alterations in pathways not previously implicated in prostate cancer, such as the spliceosome pathway. We find that the incidence of significantly mutated genes (SMGs) follows a long-tail distribution, with many genes mutated in less than 3% of cases. We identify a total of 97 SMGs, including 70 not previously implicated in prostate cancer, such as the ubiquitin ligase *CUL3* and the transcription factor *SPEN*. Finally, comparing primary and metastatic prostate cancer reveals a set of genomic markers that may inform risk stratification.

Keywords

Prostate cancer; whole exome sequencing; cancer genomics; mutational significance

The genomic landscape of primary and metastatic prostate cancer has been robustly assessed through whole exome sequencing (WES) of tumors and matched germline samples. These studies have identified multiple recurrently altered genes and pathways, including androgen signaling, DNA repair, and phosphoinositide 3-kinase (*PI3K*)/*AKT* signaling[1,2]. Additionally, they revealed genomically distinct classes of prostate cancer, defined by *ETS* transcription family fusions[3] or mutations in *SPOP*[4], *FOXA1*[4], or *IDH1*[2]. Nevertheless, prostate cancer harbors significant interpatient genomic heterogeneity, and power analyses have suggested that larger WES studies may reveal additional statistically significant mutated genes occurring at lower frequencies, indicating that the spectrum of novel prostate cancer genes is incompletely defined[5]. As the aggregation and uniform meta-analysis of WES data has been transformative to research and clinical interpretation of germline genetics[6], we hypothesized that mutational significance analysis using statistical and

biological frameworks in a large and uniformly analyzed WES cohort may similarly identify novel genes and pathways to refine the genomic landscape of prostate cancer.

We assembled and uniformly analyzed whole exome sequencing data from 1013 tumor and matched germline prostate cancers (680 primary and 333 metastatic tumors) (Supplementary Table 1)[1,2,4,7–9] that passed joint quality control parameters (Fig. 1a, Supplementary Fig. 1, Supplementary Table 2, Supplementary Note, Methods). Patient characteristics, including age at diagnosis, Gleason score, and metastatic site, are shown in Table 1 and Supplementary Table 3. The mean non- synonymous mutational load for primary and metastatic prostate cancers was 1.36 mutations/Mb and 2.93 mutations/Mb, respectively (Supplementary Fig. 2). As previously reported[2], mutational load was significantly higher in metastatic tumors ($p<0.001$, estimated 1.43 mut/MB higher mutational load adjusted for differences in tumor sequencing depth and tumor purity) (Supplementary Fig. 2, Methods). Additionally, copy-number burden was significantly higher in metastatic tumor ($p<0.001$). In primary tumors, increased age and higher Gleason score was associated with higher mutation burden ($p = 0.02$, $p < 0.001$) and copy-number burden ($p < 0.001$, $p < 0.001$) (all adjusted for tumor purity and tumor sequencing depth) (Supplementary Fig. 2)[10].

Mutational significance analysis of point mutations and short insertion/deletions using MutSig2CV[11] and additional biological significance filters (Methods) identified 97 significantly recurrently mutated genes (SMGs) (Fig. 1a,b; Supplementary Fig. 3, 4a-c; Supplementary Table 4). As predicted by prior power analyses[5], the majority of these new SMGs occurred in less than 5% of the overall cohort and could only be discovered in cohorts with over 900 samples (Supplementary Table 5). SMGs include well known prostate cancer genes[1,2,4,7–9], such as *AR*, *SPOP*, *FOXA1*, *TP53*, and *PTEN* (Fig. 1b and Supplementary Table 4). We identified 70 SMGs previously implicated in cancer, but not previously reported as significantly altered in prostate cancer[1,2,4,7–9] and an additional 9 SMGs not previously identified as recurrently altered in any cancer type (Fig. 1b and Supplementary Table 4)[5,12].

We then integrated focal copy-number events and available ETS fusion data to stratify these findings by pathway and function and developed a categorized set of significantly mutated genes in prostate cancer. Through this approach, we identified 20% of prostate cancer samples with mutations, frequently truncating, in epigenetic modifiers or chromatin remodeling genes (Fig. 1c,d and Supplementary Table 6). Within this class of tumors, 5% had mutations in genes that encode SWI/SNF nucleosome remodeling complex members (Fig. 1d), including *ARID1A* (1.6%), *ARID4A* (1%), *ARID2* (1.3%), *SMARCA1* (1.1%) similar to observations made in other tumor types[12,13]. In primary tumors, mutations in epigenetic regulators and chromatin modifiers are significantly associated with higher Gleason score (10% Gleason 3+4, 22% Gleason 8–10, $p=0.001$ Fisher's exact test). Furthermore, upon examination of the subset of our cohort for which ETS fusion status was available (n=765), we found that alterations in epigenetic regulators and chromatin remodelers were significantly more common in tumors that lack an ETS fusion ($p=1e-04$, Fisher's exact test), and in tumors without previously known drivers (ETS fusion, *IDH1/SPOP/CUL3*, or *FOXA1*) ($p = 0.007$, Fisher's exact test) (Fig. 1c,d and Supplementary Table 6).

Our analysis also identified recurrently mutated genes in the ubiquitin protease (USP) and ligase gene family, of which *SPOP* is a member, with mutations found in *USP28* (1.4%), *USP7* (1.2%), *CUL3* (1.3%) (Fig. 2a). *CUL3* encodes part of a cullin-RING-based (*BTB-CUL3-RBX1*) E3 ubiquitin ligase complex with *SPOP*[14,15], and mutations may affect degradation of prostate cancer tumorigenesis regulators including *AR, SRC-3, and TRIM24*[16,17]. *CUL3* mutations were primarily in a hotspot, p.Met299Arg, and were mutually exclusive with *SPOP* mutations (Fig. 2a), although this cohort size was not sufficiently powered to establish statistical significance. *CUL3* mutant tumors also exhibited copy-number profiles similar to those of *SPOP*-mutant tumors, with losses at chromosomes 5q, 6q and 13 (Fig. 2b and Supplementary Fig. 5)[18]. To confirm this finding in an orthogonal cohort, we identified nine additional somatic *CUL3* mutations in an independent cohort of advanced prostate cancers (1.3% in the MSK-IMPACT data[1920]), including three p.Met299Arg mutations (Supplementary Fig. 6a).

In addition, the splicing pathway was altered in 4% of prostate tumors (Fig. 2c), most notably through hotspot mutations in *SF3B1* (1.1%) and *U2AF1* (0.5%). Mutations in *SF3B1* mostly clustered around the highly conserved HEAT repeats in the C-terminus (Fig. 2c), similar to other cancer types[21,22]. This alteration is thought to disrupt the recognition and binding of 3’ splice sites[23].

We also identified SMGs in previously known prostate cancer pathways, including *AR* signaling, WNT/beta-catenin, PI3K, and RAS/MAPK. Within the *AR*/hormone signaling pathway, our analysis identified *SPEN*, which encodes a hormone inducible transcription repressor, mutated in 2.4% of this cohort, mostly through truncating mutations (Fig. 3a,b). The *SPEN* protein known to repress the estrogen receptor via *NCOR2*, by recruiting histone deacetylases and SRA, an RNA co-activator, interaction[24,25]. *SPEN* is activated via estrogen, and potentially other hormones[25], and its overexpression is associated with response to tamoxifen in breast cancer[25,26]. *SPEN* mutations were significantly enriched in metastatic samples (q=0.008, Fisher’s exact test) and clonal (Fig 3a), suggestive of *SPEN* being a driver in advanced disease.

The PI3-Kinase pathway was altered in 25% of our samples, primarily due to homozygous loss and truncating mutations in *PTEN* (16%). Our analysis identified a novel prostate cancer gene in the PI3-kinase pathway, *PIK3R2* (1%), which, like *PIK3R1*, encodes a PI3K regulatory subunit[27]. One of the *PIK3R2* mutations, p.Asp557Tyr, is paralogous to the known oncogenic p.Asp560Tyr mutation in *PIK3R1* (Supplementary Fig. 6b), and was also found in our validation cohort.

Genomic alterations in the *WNT/CTNNB1* pathway were found in 10% of the cohort (Fig. 3c and Supplementary Table 6). For *CTNNB1*, while the majority of mutations clustered in the N-terminal domain (Fig. 3d), three residues, including a novel p.Lys335Ile hotspot cluster around the *CTNNB1* interacting domain of *AXIN* (Fig. 3e, *CTNNB1* binding domain of *AXIN* highlighted in light gray). The *RAS/RAF/MAPK* pathway was altered in 5% of samples (Supplementary Table 6), including SMGs in *KRAS* and *BRAF*, mostly due to established hotspot mutations not previously enriched for significance in prostate cancer.

As previously reported, we observed a significant number of inactivating alterations in DNA repair genes (16% of samples, Supplementary Table 6). Novel prostate cancer specific SMGs in this pathway included *MRE11A* and *PALB2*. *CDK12* was mutated primarily by truncating mutations ($p<0.001$, binomial test), as previously observed in ovarian cancer. Of note, *CDK12* missense variants significantly clustered in the kinase domain ($p<0.001$, binomial test) (Supplementary Fig. 6c), suggesting a putative functional relevance. Furthermore, 15 of 31 *CDK12* mutant tumors (as well as 27 of 56 samples in the validation cohort) harbored two mutations in the gene, suggestive of frequent biallelic inactivation of the gene. Broadly, these results expand on SMGs in known cancer pathways not previously implicated in prostate cancer, and further delineate the genomic heterogeneity of mutations in the long tail of this disease.

Finally, we conducted a systematic comparison of primary and metastatic tumors to identify which events are associated with advanced disease (Fig. 4a, Methods). Genes with enrichment in metastatic samples include *TP53*, *AR*, *PTEN*, *RB1*, *FOXA1*, *APC*, and *BRCA2* (Fig. 4a). Alterations in epigenetic regulators, including *KMT2C* and *KMT2D* are also significantly enriched in metastatic tumors, and in aggregate define a genomic signature of high risk disease. Conversely, mutations in *SPOP* were significantly enriched in primary tumors (Fig. 4a). After correction for differences in mutational load, *IDH1 and ZMYM3* mutations were also enriched in primary tumors ($p=0.01$, mutation rate-adjusted permutation test). At the pathway level, PI3K, DNA repair, Cell cycle, *WNT/CTNNB1*, and epigenetic regulators were significantly more frequently altered in metastatic compared to primary tumors ($p<0.0001$ Fisher's exact test, Fig. 4b and Supplementary Table 7).

Within a given cancer type, the ability to redefine mutational significance with rapidly expanding sample sizes may identify new biologically and clinically relevant genes and pathways not previously appreciated. This study has leveraged this strategy to identify novel driver genes and pathways potentially implicated in the pathogenesis of prostate cancer. While many of the significantly altered genes and pathways are mutated at low frequencies, given the incidence of prostate cancer these alterations still impact large patient populations. In addition, whereas expanded analysis of primary indolent prostate cancer suggests near saturation for gene discovery[28], this analysis, which includes more advanced cases, has revealed new biologically and clinically relevant events and creates an opportunity to prospectively assess a metastasis-associated genomic marker for clinical stratification in localized prostate cancer.

Combined statistical and biological significance analysis enabled a focused assessment of the SMGs identified herein, and efforts to functionally characterize this long tail of SMGs in prostate cancer may inform their relative phenotypic effects on oncogenicity, metastatic potential, and response characteristics to known or emerging prostate cancer therapeutics. Indeed, many of the genes identified through statistical analysis alone are of unknown function and suggest that even larger sample sizes paired with functional analysis will be necessary to discriminate which are relevant to prostate cancer oncogenesis. Subsequent studies that harmonize even larger prostate cancer molecular cohorts through uniform genomic analysis may also orthogonally validate these findings and further mitigate technical differences, such as stochastic effects of sequencing on variant detection, when

analyzed in aggregate. Overall, our analysis demonstrates the utility of uniform genomic analysis in a single cancer type at a larger scale than previously reported, thereby redefining the molecular landscape of prostate cancer and providing rationale to revisit mutational significance in other cancer types as data generation scales by orders of magnitude.

ONLINE METHODS

Cohort collection and quality control

Samples were included in this study if tumor and matched germline WES raw sequencing data (BAM or Fastq files) were accessible and met downstream quality control characteristics (see Quality Control, below). These cohorts were identified through review of the literature and expert review (Supplementary Table 1). All cohorts had institutional review board approval for access from the original studies, listed in the citation. We obtained the WES BAM files from all samples. All samples underwent uniform alignment through the same version of the PICARD pipeline. Details of versions and parameters for all tasks within the PICARD pipeline are provided in the Supplementary Note. All tumor samples were required to have at least 50x mean target coverage and all paired normals were required to have at least 30x mean target coverage. Mean target coverage across the cohorts for tumors was 104.7x and for normals was 103.8x. ContEst was used to estimate the level of contamination with foreign DNA[29]. All samples had Contest scores lower than 5%, and the mean Contest value was 0.6%.

Clinical Data

All clinicopathological annotations were obtained from the original papers[1,2,4,7–9]. All primary tumors were treatment-naive; all metastatic tumors were castration resistant.

Variant Calling

To restrict the analysis to consider sites in the common pool of bases covered the bait sets used in the respective source projects, an intersected BED file was created using the bedtools intersect tool (Supplementary Table 8; Supplementary Fig. 7)

(http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html). Single nucleotide variants (SNVs) were called with MuTect (version 1.1.6)[30], using the intersected BED file. Unfiltered MuTect mutation call are located in Supplementary Table 9.

Artifacts introduced by DNA oxidation during sequencing or formalin fixation process were removed when appropriate[31]. Specifically regarding artifacts from formalin fixation, formalin fixation introduces multiple types of DNA damage including deamination, which converts cytosine to uracil and leads to downstream mispairing in PCR: C>T / G>A. Because deamination occurs prior to ligation of palindromic Illumina adapters, likely deamination artifacts will have a read orientation bias. We then use this read orientation to identify artifacts and calculate a Phred scaled Q-score for FFPE artifacts[32].

To further reduce low confidence mutations with potential strand bias, we performed a Fisher's exact test on each called mutation site in aggregate to identify variants occurring significantly more frequently in one read direction than in the other. A false discovery rate

threshold, measured by Benjamini-Hochberg, of <0.0001 was used. In addition, all SNVs were required to have an allelic fraction of >= 0.01 to be called.

Insertions and deletions (indels) were called with Strelka (version 1.0.11)[33]. SNVs and indels were also filtered through a large panel of normals to extract additional poor calls. Any mutations in hotspot genes, defined by cancerhotspots.org[34], initially called by MuTect but subsequently filtered out were rescued for the final variant list. When possible, we used ERG fusions calls defined as per the original source data[2,4]. For the 126 additional TCGA samples that were not part of the TCGA manuscript, we derived ERG fusion status via mRNA expression levels, inferring from samples with outlier expression of ERG likely contain an ERG fusion[2].

Exome-wide copy-number ratios were inferred from coverage information using ReCapSeg (http://gatkforums.broadinstitute.org/gatk/categories/recapseg). For 303 prostate cancer samples that were analyzed by TCGA, we compared the segmented copy-number profiles generated by ReCapSeg to those from SNP6 data[2] (Supplementary Fig. 8). We generated a scatter plot to compare the segment means of matched segments >200KB from the SNP6 and the ReCapSeg data, resulting in a pearson correlation of 0.92. Significant focal copy-number alterations were identified from segmented data using GISTIC 2.0[35]. In addition, we called the allelic copy-number of well known prostate cancer genes, accounting for purity and ploidy, obtained from FACETS (version 0.5.10)[36] (genes examined: *TP53, APC, PTEN, RB1, BRCA2, CDKN1B, FANCA, ATM, AR*). We performed manual review of copy number calls for selected oncogenes and tumor suppressors. All data is available for visualization and analysis in the cBioPortal for Cancer Genomics at http://www.cbioportal.org/[37].

Mutation and Copy-Number Burden

Mutational burden was calculated as the number of mutations over the number of bases covered per sample and is reported as mutations per megabase. Copy-number burden was calculated as fraction of genome altered using copy-number segments with >|0.2|, as previously defined[2]. A multivariate linear regression adjusting for purity and coverage was used to evaluate the difference in mutational and copy-number burden in metastatic and primary tumors. Additional information is provided in the Supplementary Note.

Mutational Significance Analysis

All mutations that passed QC were analyzed using Mutsig2CV[5] to identify significantly mutated genes (SMGs). Mutsig2CV integrates three separate significance algorithms: MutsigCV, MutsigFN, which looks at the functionality of a mutation in a gene, and MutsigCL, which looks at the clustering of mutations within the gene, specifically looking for hotspot mutations. Both MutsigFN and MutsigCL measure significance based on permutations. Significantly mutated genes (SMGs) fell within two different categories: 1) *q* values less than 0.1 and altered in at least 10 samples, 2) q values between 0.1 and 0.25, altered in 10 samples, and in known cancer genes[5,38]. Additionally, genes with low median allelic fraction (<0.1) were removed from the SMG list. Genes whose length was >1500aa (except for cancer genes[5,38] or if the gene had a fraction of truncating variants larger than

50% of total mutations, indicating a putative tumor suppressor) were also removed from the SMG list. Genes with low expression in prostate cancer (median expression below bottom tertile TCGA RNAseq[2]) were also removed from the SMG list. Finally, genes with at least five oncogenic variants (according to OncoKB, http://oncokb.org), but were not previously included in the SMG list, were added to the SMG list[39].

Comparison of genomic alterations between primary and metastatic tumors

Enrichment analysis of mutations and copy-number alterations observed in metastatic tumors compared to primary tumors was performed by tabulating the frequency of mutations or copy-number events observed in either metastatic or primary prostate cancer and performing a two-sided Fisher's exact test on a set of biologically relevant cancer genes (n=650 genes)[5,38]. Multiple hypothesis test correction was performed using Benjamini-Hochberg method. To adjust for differences due to increased mutation load in metastatic tumors, we also performed a modified Fisher's exact test; a permutation test where the probability of mutation in each sample is weighted by the mutation rate in that sample, and a simulation of 10,000 permutations performed with a two-sided p-value calculated as the proportion of those permutations with the observed or more extreme outcome. This directly corrects for differential observed mutation rates between primary and metastatic tumors, and represents the null hypothesis that mutations are equally likely to be found in primary vs. metastatic tumors, adjusting for differences in mutation rate. We were able to perform this mutational-rate based adjustment in genes where the only events were mutations. In cases where functional events included both gene mutations and copy-number changes (e.g. *PTEN*), we performed only a Fisher's Exact test.

Clonality analysis

Clonality of mutations was estimated as cancer cell fraction (CCF)[40], and implemented in the FACETS algorithm[36]. Additional information is provided in the Supplementary Note.

Statistical Analysis

Two-tailed Fisher's exact test was used to assess enrichment of alterations in epigenetic regulators and chromatin remodelers in ETS-negative tumors. Association of mutation burden and fraction genome altered with metastasis status, age at diagnosis and Gleason score were evaluated using Mann Whitney Wilcoxon test and permutation test. All statistical analysis were performed using R version 3.3.1 (https://www.r-project.org).

Validation Datasets

To validate mutations detected in this study cohort, we queried cancer panel data from two sources: 1) Foundation Medicine, 204 patients with prostate cancer, as published[41]. Mutation calling for this cohort was obtained as previously described and data is available in phs001179. 2) Clinical sequencing data from 706 samples from Memorial Sloan Kettering patients (MSK-IMPACT)[19,20]. Mutation calling for this cohort was obtained as previously described and data is available from the paper or at cBioPortal.org.

Data Availability

BAM files are accessible as described for the original cohorts (Supplementary Table 1). In addition, all mutation calls and clinical annotation were deposited into cBioPortal for analysis and visualization: http://www.cbioportal.org/study?id=prad_p1000.

Code Availability

Bioinformatics tools used in the analysis of this data set are publicly available. Any that are not are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the patients for participating in this study. We also thank the Broad Cancer Genome Analysis and Data Sciences groups for analysis methodology and computational support.

FUNDING: This work was supported by the SU2C-PCF Prostate Cancer International Dream Team, Prostate Cancer Foundation Young Investigator Awards (B.S.T., C.P., N.S., E.M.V.A.), Prostate Cancer Foundation-V Foundation Challenge Award (E.M.V.A., P.S.N., J. S. d.B.), NIH K08CA188615 (E.M.V.A.), NCI P50-CA097186 and NCI P50-CA92629 SPOREs in Prostate Cancer, the Marie-Josée and Henry R. Kravis Center for Molecular Oncology, a National Cancer Institute Cancer Center Core Grant (P30-CA008748), and the Robertson Foundation (B.S.T, N.S.).

REFERENCES

1. Robinson D et al. Integrative clinical genomics of advanced prostate cancer. Cell 161, 1215–1228 (2015). [PubMed: 26000489]
2. The Molecular Taxonomy of Primary Prostate Cancer. Cell 163, 1011–1025 (2015). [PubMed: 26544944]
3. Tomlins SA et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310, 644–648 (2005). [PubMed: 16254181]
4. Barbieri CE et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat. Genet 44, 685–689 (2012). [PubMed: 22610119]
5. Lawrence MS et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495–501 (2014). [PubMed: 24390350]
6. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016). [PubMed: 27535533]
7. Beltran H et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat. Med 22, 298–305 (2016). [PubMed: 26855148]
8. Kumar A et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat. Med 22, 369–378 (2016). [PubMed: 26928463]
9. Baca SC et al. Punctuated evolution of prostate cancer genomes. Cell 153, 666–677 (2013). [PubMed: 23622249]
10. Hieronymus H et al. Copy number alteration burden predicts prostate cancer relapse. Proc. Natl. Acad. Sci. U. S. A 111, 11139–11144 (2014). [PubMed: 25024180]
11. Lawrence MS et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218 (2013). [PubMed: 23770567]
12. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499, 43–49 (2013). [PubMed: 23792563]

13. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 507, 315–322 (2014). [PubMed: 24476821]
14. Theurillat J-PP et al. Ubiquitylome analysis identifies dysregulation of effector substrates in SPOP-mutant prostate cancer. Science 346, 85–89 (2014). [PubMed: 25278611]
15. Geng C et al. Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover. Proc. Natl. Acad. Sci. U. S. A 110, 6997–7002 (2013). [PubMed: 23559371]
16. Yuan W-C et al. A Cullin3-KLHL20 Ubiquitin ligase-dependent pathway targets PML to potentiate HIF-1 signaling and prostate cancer progression. Cancer Cell 20, 214–228 (2011). [PubMed: 21840486]
17. Groner AC et al. TRIM24 Is an Oncogenic Transcriptional Activator in Prostate Cancer. Cancer Cell 29, 846–858 (2016). [PubMed: 27238081]
18. Boysen G et al. SPOP mutation leads to genomic instability in prostate cancer. Elife 4, (2015).
19. Abida W et al. Prospective Genomic Profiling of Prostate Cancer Across Disease States Reveals Germline and Somatic Alterations That May Affect Clinical Decision Making. JCO Precis Oncol 2017, (2017).
20. Zehir A et al. Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. Nat. Med (2017). doi:10.1038/nm0817-1004c
21. Dolatshad H et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia 29, 1092–1103 (2015). [PubMed: 25428262]
22. Ciriello G et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell 163, 506–519 (2015). [PubMed: 26451490]
23. Papaemmanuil E et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N. Engl. J. Med 365, 1384–1395 (2011). [PubMed: 21995386]
24. McHugh CA et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521, 232–236 (2015). [PubMed: 25915022]
25. Shi Y Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. Genes Dev 15, 1140–1151 (2001). [PubMed: 11331609]
26. Légaré S et al. The Estrogen Receptor Cofactor SPEN Functions as a Tumor Suppressor and Candidate Biomarker of Drug Responsiveness in Hormone-Dependent Breast Cancers. Cancer Res 75, 4351–4363 (2015). [PubMed: 26297734]
27. Kuchay S et al. FBXL2- and PTPL1-mediated degradation of p110-free p85β regulatory subunit controls the PI(3)K signalling cascade. Nat. Cell Biol 15, 472–480 (2013). [PubMed: 23604317]
28. Fraser M et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature (2017). doi: 10.1038/nature20788
29. Cibulskis K et al. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics 27, 2601–2602 (2011). [PubMed: 21803805]
30. Cibulskis K et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol 31, 213–219 (2013). [PubMed: 23396013]
31. Costello M et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res 41, e67 (2013). [PubMed: 23303777]
32. Van Allen EM et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat. Med 20, 682–688 (2014). [PubMed: 24836576]
33. Saunders CT et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817 (2012). [PubMed: 22581179]
34. Chang MT et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat. Biotechnol 34, 155–163 (2016). [PubMed: 26619011]
35. Mermel CH et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 12, R41 (2011). [PubMed: 21527027]

36. Shen R & Seshan VE FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 44, e131 (2016). [PubMed: 27270079]
37. Cerami E et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2, 401–404 (2012). [PubMed: 22588877]
38. Cheng DT et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J. Mol. Diagn 17, 251–264 (2015). [PubMed: 25801821]
39. Chakravarty D. OncoKB: A Precision Oncology Knowledge Base. Journal of Clinical Oncology Precision Oncology (2017).
40. McGranahan N et al. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci. Transl. Med 7, 283ra54 (2015).
41. Hartmaier RJ et al. High-Throughput Genomic Profiling of Adult Solid Tumors Reveals Novel Insights into Cancer Pathogenesis. Cancer Res (2017). doi:10.1158/0008-5472.CAN-16-2479

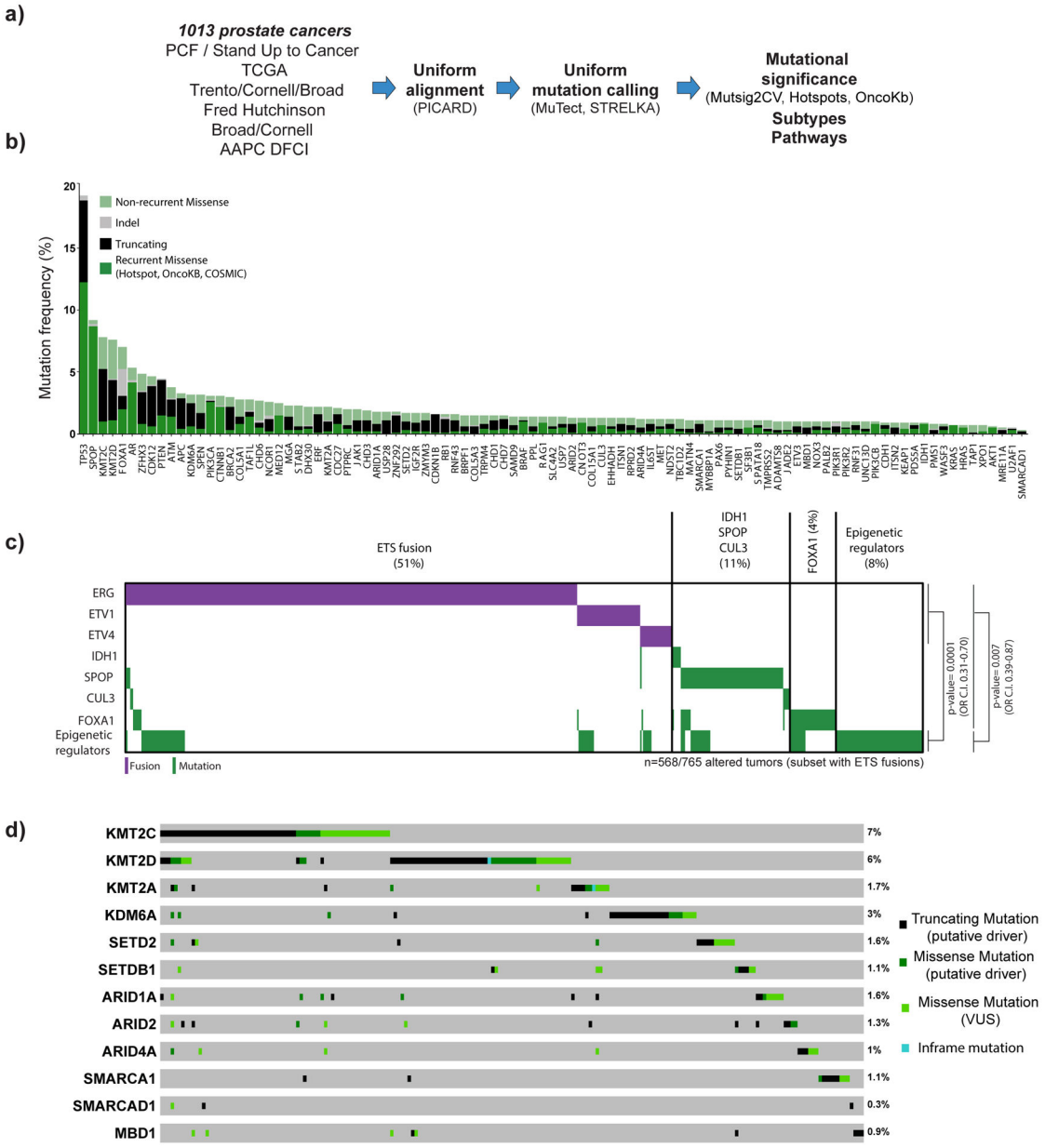

Figure 1.

Mutational significance in 1013 prostate cancers. **(a)** Uniform alignment, mutation calling, and significance analysis. **(b)** Recurrently mutated genes (n = 97). Genes are ordered by frequency, and mutations are stratified by mutation type and, for missense mutation, by recurrence. Recurrence is defined via cancerhotspots.org, OncoKB.org, and COSMIC; truncating mutations are defined as frameshift, nonsense, splice, nonstop. (**c**) Mutations in epigenetic regulators and chromatin remodelers are significantly enriched in ETS-negative tumors. p-values are calculated using a two-tailed Fisher's exact test and shown for ETS fusions compared to all epigenetic mutations (including those co-occurring with *SPOP* and *CUL3*) and for ETS fusions compared to non-overlapping mutations in epigenetic modifiers only. (**d**) Cohort-wide view of mutations in epigenetic regulators and chromatin remodelers,

which affect 20% of samples. Samples are shown from left to right (only the 202 tumors with alterations are shown, out of 1013), and gene alterations are color-coded by mutation type and, for missense mutations, by assumed driver status; mutations are assumed to be drivers if they have been previously reported and entered into COSMIC or annotated in OncoKB or variants of unknown significance (VUS).

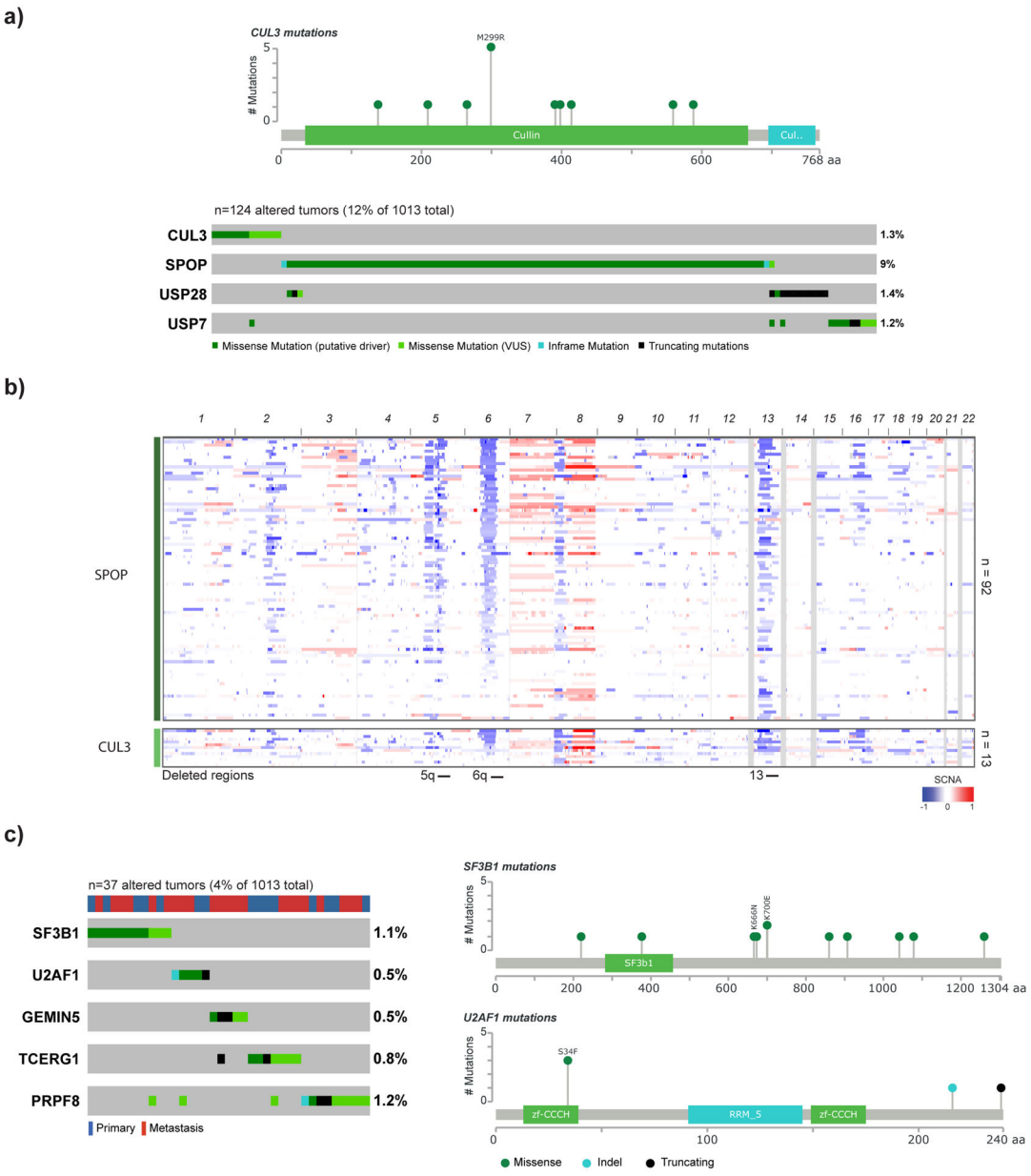

Figure 2.

Ubiquitin and splicing pathways in prostate cancer. **(a)** Mutations in *CUL3* are exclusively missense mutations, and five tumors show the recurrent p.Met299Arg mutation. *CUL3* mutations are mutually exclusive with *SPOP* mutations, and including USP28, *USP7*, 12% of tumors harbor alterations in members of the ubiquitin pathway. **(b)** *CUL3*-mutant tumors show copy-number profile similar to those of *SPOP*-mutant tumors. Chromosomes are shown from left to right, samples from top to bottom. Regions of loss are indicated by shades of blue, and gains are indicated by shades of red. **(c)** Mutations in members of the splicing pathway are found in 4% of tumors, including oncogenic mutations in *SF3B1* and *U2AF1*.

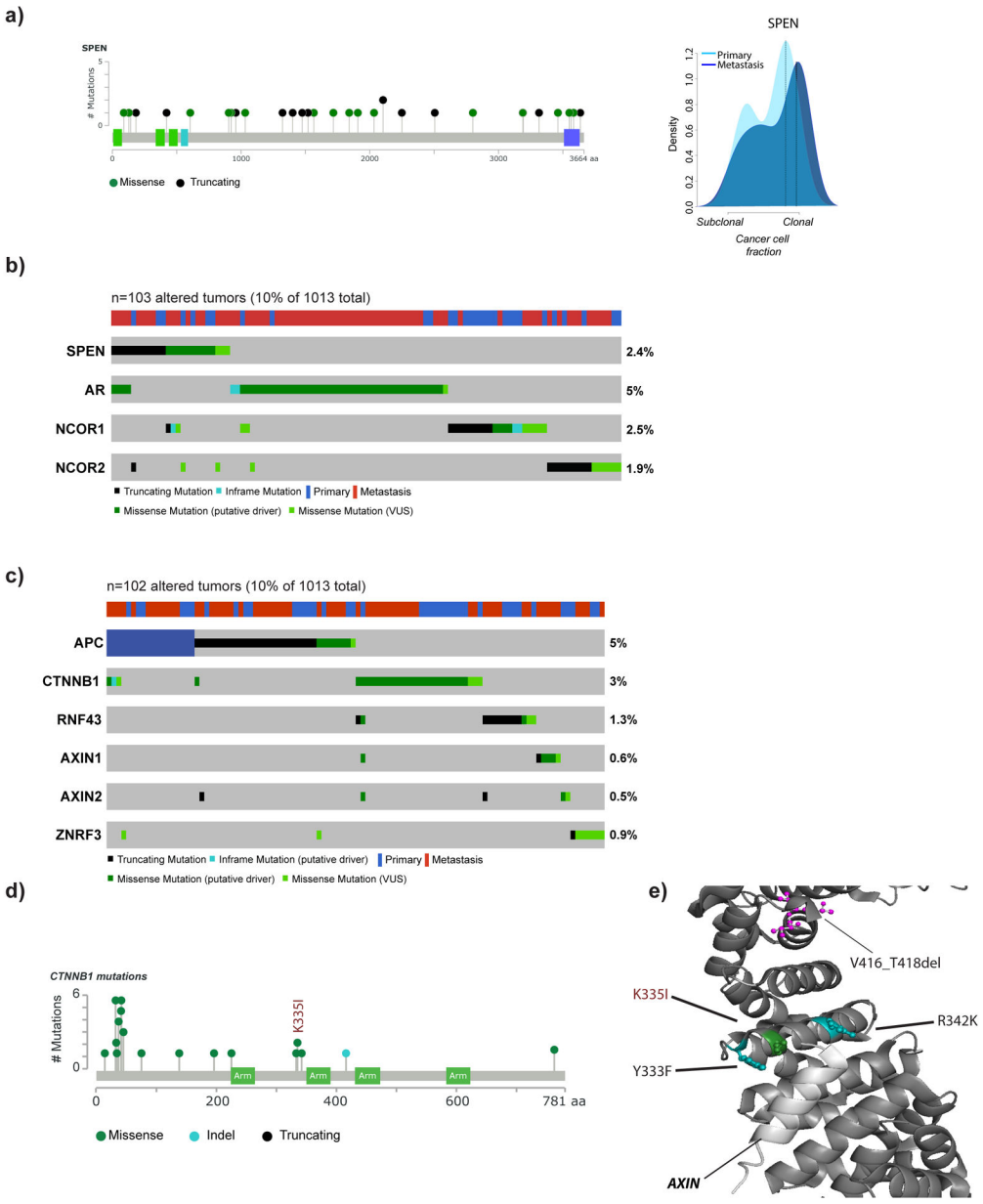

Figure 3.

SPEN mutations and WNT pathway alterations. **(a)** The majority of *SPEN* mutations are truncating and clonal in metastatic samples. **(b)** Oncoprint highlighting the distributions of *SPEN* mutations with alterations in members of the *AR* signaling. (**c**) Alterations in WNT/CTNNB1 pathway are found in 10% of tumors, primarily with loss of function mutations in *APC* and missense mutations in *CTNNB1*. **(d)** *CTNNB1* mutations cluster primarily in hotspots in the N-terminal domain. **(e)** 3D structure of *CTNNB1* showing novel mutations clustered around the CTNNB1-interacting domain of AXIN (highlighted in light gray).

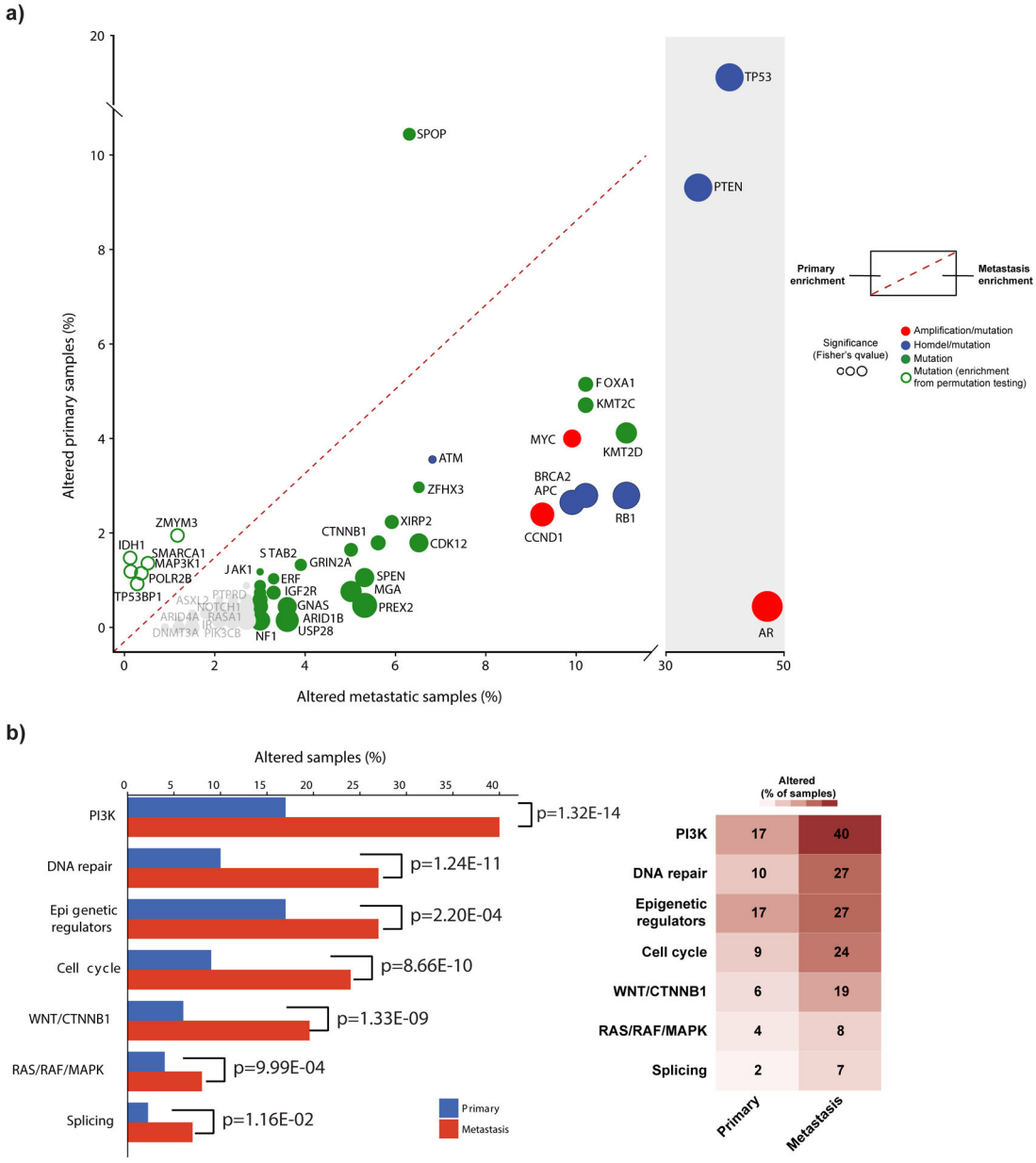

Figure 4.

Enrichment of genomic alterations in metastatic tumors. (a) Most genomic alterations are enriched in metastatic disease. Alteration percentages in metastatic samples (n=333) are shown on the x-axis, primary samples (n=680) on the y-axis. The significance of enrichment (two-sided Fisher's test q-value or weighted permutation test) is shown by the size of the dots. Genes in bold have a significant enrichment of mutations using Fisher's test and weighted permutation test correcting for mutation burden. (b) Pathway alteration frequencies in metastatic disease compared to primary disease. A sample was considered altered in a given pathway if at least a single gene in the pathway had a genomic alteration. p-values indicate the level of significance (two-sided Fisher's exact test).

Table 1

Cohort characteristics. Baseline demographic and clinical data for the aggregate cohort, including age, Gleason score, metastatic site (if applicable).

Primary Tumors (n=680)	Gleason Score	6	103
		3+4	208
		4+3	143
		8–10	196
		Unknown	30
	Age at Diagnosis	Median	62
		Unknown (n)	80
Metastatic Tumors (n=333)	Metastatic Site	Bone	80
		Lymph Node	82
		Lung	7
		Soft Tissue	2
		Other	26
		Unknown	107