



Published in final edited form as:

Cell. 2015 November 5; 163(4): 1011–1025. doi:10.1016/j.cell.2015.10.025.

The molecular taxonomy of primary prostate cancer

The Cancer Genome Atlas Research Network

The Cancer Genome Atlas Program Office, National Cancer Institute at NIH, 31 Center Drive, Bldg. 31, Suite 3A20, Bethesda, MD 20892, USA

Summary

There is substantial heterogeneity among primary prostate cancers, evident in the spectrum of molecular abnormalities and its variable clinical course. As part of The Cancer Genome Atlas (TCGA), we present a comprehensive molecular analysis of 333 primary prostate carcinomas. Our results revealed a molecular taxonomy in which 74% of these tumors fell into one of seven subtypes defined by specific gene fusions (*ERG*, *ETV1/4*, *FLII*) or mutations (*SPOP*, *FOXAI*, *IDHI*). Epigenetic profiles showed substantial heterogeneity, including an *IDHI*-mutant subset with a methylator phenotype. Androgen receptor (AR) activity varied widely and in a subtype-specific manner with *SPOP* and *FOXAI* mutant tumors having the highest levels of AR-induced transcripts. 25% of the prostate cancers had a presumed actionable lesion in the PI3K or MAPK signaling pathways, and DNA repair genes were inactivated in 19%. Our analysis reveals molecular heterogeneity among primary prostate cancers, as well as potentially actionable molecular defects.

Introduction

Prostate cancer is the second most common cancer in men and the fourth most common tumor type worldwide (Ferlay et al., 2013). It is estimated that in 2015, prostate cancer will be diagnosed in 220,800 men in the United States alone and that 27,540 will die of their disease (Siegel et al., 2015). Multiple genetic and demographic factors, including age,

Corresponding authors: Nikolaus Schultz (schultz@cbio.mskcc.org), Massimo Loda (massimo_loda@dfci.harvard.edu), Chris Sander (chris@sanderlab.org).

Author Contributions

Project leaders: M. Loda and C. Sander; **analysis leaders:** N.S. and B.S.T.; **project coordinators:** I.F. and M. Sheth; **data coordinators:** J. Armenia and N.S.; **supplement coordinator:** J. Armenia; **pathology review:** M.I., M. Loda, V.E.R., B.R., M.A.R., P. Troncoso, L.D.T., and H.Y.; **clinical data:** J. Bowen, N.M.C., L.I., K.M.L., and T.M.L.; **tumor cellularity analysis:** J. Ahn, P.C.B., A.D.C., F.D., S.H., D.P., M.A.R., J. Shih, S.T., and W.W.; **exome sequencing analysis:** M. Gupta, C. Sougnez, and E.V.A.; **whole-genome sequencing analysis:** J. Armenia, Y.F., M.B.G., M. Gupta, E.K., L. Lochovsky, A. Pantazi, C. Sougnez, and E.V.A.; **copy-number analysis:** A.D.C., and J. Shih; **mRNA expression and fusion analysis:** A. Sboner, N.S., M.D.W., and C.-C.W.; **androgen receptor analysis:** J. Armenia, R.K.B., H.D., R.A.M., A.J.M., P.S.N., and N.S.; **DNA methylation analysis:** P.W.L., S.K.R., and H.S.; **miRNA analysis:** A.G.R.; **RPPA analysis:** R.A., W.L., Y.L., and G.B.M.; **integrative analysis:** A. Arora, A.D.C., D.I.H., L.I., N.S., R. Shen, B.S.T., and M.D.W.; **batch effects analysis:** R.A., A.K., K.-V.L., S. Ling, A.R., and J.N.W.; **mRNA degradation analysis:** A.K., K.-V.L., G.R., M.A.R., N.S., and M.D.W.; **manuscript writing:** C.E.B., C.C.B., P.R.C., Y.C., A.D.C., F.D., S.M.L., M. Loda, J. Simko, P.S.N., S.K.R., A.G.R., M.A.R., C. Sander, A. Sboner, N.S., B.S.T., S.A.T., E.V.A., and J.N.W.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

family history, genetic susceptibility, and race, contribute to the high incidence of prostate cancer (Al Olama et al., 2014).

In the current era of PSA screening, nearly 90% of prostate cancers are clinically localized at the time of their diagnosis (Penney et al., 2013). The clinical behavior of localized prostate cancer is highly variable – while some men will have aggressive cancer leading to metastasis and death from the disease, many others will have indolent cancers that are cured with initial therapy or may be safely observed. Multiple risk stratification systems have been developed, combining the best currently available clinical and pathological parameters (such as Gleason score, PSA levels, clinical and pathological staging); however, these tools still do not adequately predict outcome (Cooperberg et al., 2009, D'Amico et al., 1998, Kattan et al., 1998). Further risk stratification using molecular features could potentially help distinguish indolent from aggressive prostate cancer.

Molecular and genetic profiles are increasingly being used to subtype cancers of all types and to guide selection of more precisely targeted therapeutic interventions. Several recent studies have explored the molecular basis of primary prostate cancer and have identified multiple recurrent genomic alterations that include mutations, DNA copy-number changes, rearrangements, and gene fusions (Baca et al., 2013, Barbieri et al., 2012, Berger et al., 2011, Lapointe et al., 2007, Pflueger et al., 2011, Taylor et al., 2010, Tomlins et al., 2007, Wang et al., 2011). The most common alterations in prostate cancer genomes are fusions of androgen-regulated promoters with *ERG* and other members of the ETS family of transcription factors. In particular, the *TMPRSS2-ERG* fusion is the most common molecular alteration in prostate cancer (Tomlins et al., 2005), being found in between 40 and 50% of prostate tumor foci, translating to more than 100,000 cases annually in the United States (Tomlins et al., 2009). Nevertheless, among treated prostate cancers, and despite extensive study, patients with fusion-bearing tumors do not appear to have a significantly different prognosis following prostatectomy than those without (Gopalan et al., 2009, Pettersson et al., 2012). Prostate cancers also have varying degrees of DNA copy-number alteration; indolent and low-Gleason tumors have few alterations, whereas more aggressive primary and metastatic tumors have extensive burdens of copy-number alteration genome-wide (Taylor et al., 2010, Hieronymus et al., 2014, Lalonde et al., 2014). In contrast, somatic point mutations are less common in prostate cancer than in most other solid tumors. The most frequently mutated genes in primary prostate cancers are *SPOP*, *TP53*, *FOXA1*, and *PTEN* (Barbieri et al., 2012). Only recently has the spectrum of epigenetic changes in prostate cancer genomes been explored (Börno et al., 2012, Friedlander et al., 2012, Kim et al., 2011, Kobayashi et al., 2012, Mahapatra et al., 2012).

Importantly, no studies have comprehensively integrated diverse omics data types to assess the robustness of previously defined subtypes and potentially prognostic alterations. Here, to gain further insight into the molecular-genetic heterogeneity of primary prostate cancer and to establish a molecular taxonomy of the disease for future diagnostic, prognostic, and therapeutic stratification, the TCGA Network has comprehensively characterized 333 primary prostate cancers using seven genomic platforms. This analysis reveals novel molecular features that provide a better understanding of this disease and suggest potential therapeutic strategies.

Results

Cohort and platforms—The cohort of primary prostate cancers analyzed resulted from extensive pathologic, analytical, and quality control review, yielding 333 tumors from 425 available cases. Images of frozen tissue were evaluated by multiple expert genitourinary pathologists and cases were excluded if no tumor cells were identifiable in the sample, or if there was evidence of significant RNA degradation (Figure S1, Supplementary Methods). For the subset of cases reviewed by two pathologists, tumor cellularity estimates were within 20% of each other in 71% of cases. In total, 78% of Gleason scores were concordant within one grade of the secondary pattern (Supplementary Methods). Moreover, due to the challenge of acquiring primary prostate cancer specimens of high tumor cellularity, we also performed a multi-platform analysis of tumor content, estimating tumor purity with analytical approaches utilizing both DNA (Carter et al., 2012, Prandi et al., 2014) and RNA (Quon et al., 2013, Ahn et al., 2013) sequencing data. The molecular and pathologic estimates are presented in Table S1A and Figure S1. The clinical and pathological characteristics of the final cohort are presented in Table 1. The average follow-up time following radical prostatectomy was just under two years, which precluded outcomes analysis due to the long natural history of primary prostate cancer.

We characterized isolated biomolecules from these 333 tumor samples using four platforms: whole-exome sequencing for somatic mutations, array-based methods for profiling both somatic copy-number changes and DNA methylation, and messenger RNA (mRNA) sequencing. We also performed microRNA (miRNA) sequencing on 330 of these samples, reverse-phase protein array (RPPA) on 152 samples, and low-pass and high-pass whole-genome sequencing (WGS) on 100 and 19 tumor/normal pairs, respectively (Supplementary Methods). For 19 samples, non-malignant adjacent prostate samples were also examined for DNA methylation and RNA/miRNA expression analyses.

The molecular taxonomy of primary prostate cancer

Previous studies indicate that many genomically distinct subsets of prostate cancer exist. These are driven in some cases by frequent events, such as androgen-regulated fusions of *ERG* and other ETS-family members, or recurrent *SPOP* mutations, and in other cases by less common genomic aberrations. Given the comprehensive nature of our data, we sought to unify these disparate findings to establish a molecular taxonomy of primary disease that integrates results from somatic mutations, gene fusions, somatic copy-number alterations (SCNA), gene expression, and DNA methylation. We first performed unsupervised clustering of data from each molecular platform as well as integrative clustering using iCluster (Shen et al., 2009). These analyses uncovered both known and novel associations with 74% of all tumors being assignable to one of seven molecular classes based on distinct oncogenic drivers: fusions involving 1) *ERG*, 2) *ETV1*, 3) *ETV4*, or 4) *FLII* (46, 8, 4, and 1% respectively), or mutations in 5) *SPOP*, 6) *FOXA1*, or 7) *IDH1* mutations (11, 3, and 1% respectively) (Figures 1 and S2, Table S1A).

In total, 53% of tumors were found to have ETS-family gene fusions (*ERG*, *ETV1*, *ETV4*, and *FLII*) after analysis with two complementary algorithms (Sboner et al., 2010, Wang et

al., 2010) (see **Methods**). While *TMPRSS2* was the most frequent fusion partner in all ETS fusions, we identified fusions with other previously described androgen-regulated 5' partner genes, including *SLC45A3* and *NDRG1* (Table S5). We also identified several tumors that overexpressed full-length ETS transcripts that were mutually exclusive with ETS fusions (12 *ETV1* high tumors, 6 *ETV4*, and 2 *FLII*) (Table S5). ETS overexpression in these cases could possibly be mediated via epigenetic mechanisms or cryptic translocations of the entire gene locus to a transcriptionally active neighborhood. In the one case with elevated *ETV1* full-length expression studied by whole-genome sequencing, we identified a cryptic genomic rearrangement 3' of the *ETV1* locus with a region on chromosome 14 near the *MIPOL1* gene adjacent to *FOXA1*. This event is similar to previously described *ETV1* translocations in LNCaP and MDA-PCa2b cell lines and in patient samples (Tomlins et al., 2007, Gasi et al., 2011). Overall, while fusions in the four genes were mostly mutually exclusive, three tumors showed evidence for fusions involving more than one of these genes (Table S5). Given that histologically defined single tumor foci have been shown to be rarely composed of different ETS fusion positive clones (Cooper et al., 2015, Kunju et al., 2014, Pflueger et al., 2011), it is likely these cases reflect convergent phenotypic evolution in clonally heterogeneous tumors. Tumors defined by *SPOP* mutations were mutually exclusive with all ETS fusion-positive cases, though four of the *SPOP*-mutant tumors also possessed *FOXA1* mutations. In all four of these tumors, both the *SPOP* and *FOXA1* mutations were clonal, indicating they are present in the same tumor cells.

Beyond the class-defining lesions, there were multiple patterns of both known and novel concurrent alterations in key prostate cancer genes. The former included the preponderance of *PTEN* deletions in *ERG* fusion-positive cases (Taylor et al., 2010). Similarly, *SPOP* mutations have previously been found to occur in approximately 10% of clinically localized prostate cancers, were mutually exclusive of tumors defined by ETS rearrangements, and may designate a distinct molecular class of disease based primarily on distinctive SCNA profiles (including deletion of *CHDI*, 6q, and 2q) (Barbieri et al., 2012, Blattner et al., 2014). Beyond reaffirming these known patterns, our taxonomy revealed new relationships and subtypes. Specifically, the *SPOP*-mutant/*CHDI*-deleted subset of prostate cancers had notable molecular features, including elevated levels of DNA methylation, homogeneous gene expression patterns, as well as frequent overexpression of *SPINK1* mRNA, supporting *SPOP* mutation as a key feature in the molecular taxonomy of prostate cancer. Interestingly, mRNA, copy-number, and methylation profiles were similar in tumors with *FOXA1* mutations and those with *SPOP* mutations. Furthermore, we identified a new genomically distinct subtype of prostate cancer defined by hotspot mutations in *IDH1*, described in greater depth below.

Despite this detailed molecular taxonomy of primary prostate cancers, 26% of all tumors studied appeared to be driven by still-occult molecular abnormalities or by one or more frequent alterations that co-occur with the genomically defined classes. Some of these tumors showed a high burden of copy-number alterations or DNA hypermethylation. Enrichment analysis indicated that this subset of tumors was enriched for mutations in *TP53*, *KDM6A*, and *KMT2D*; deletions of chromosomes 6 and 16; and amplifications of chromosomes 8 (spanning *MYC*) and 11 (*CCND1*) (Table S2). To further characterize this

group further, we performed whole genome sequencing of 19 tumor specimens and their matched normal tissues, a subset of which had high tumor cellularity but still lacked DNA copy-number alterations or any known or presumed driver lesions. Interestingly, no occult driver abnormalities or highly recurrent regulatory mutations were identified, such as the *TERT* promoter mutation common to many other tumor types (Khurana et al., 2013). Therefore, a significant (up to 26%) subset of primary prostate cancers of both good and poor clinical prognosis (including those with Gleason scores of >8) are driven by as yet unexplained molecular alterations.

mRNA clusters were tightly correlated with ETS fusion status, where mRNA cluster 1 consisted primarily of ETS-negative tumors and mRNA clusters 2 and 3 were split among ETS fusion-positive tumors (Figure 1, Figure S4). miRNA clustering showed a similar pattern, revealing a general difference in miRNA expression between ETS-positive and negative tumors (Figure 1, Figure S6). Clustering of RPPA data identified three distinct subgroups, with cluster 3 exhibiting elevated PI3K/AKT, MAP-Kinase and receptor tyrosine kinase activity (Figure S7A). The cluster was not enriched, however, in genomic alterations in these pathways, and in general there was little correlation of increased pathway activity (as measured by phospho-AKT and other downstream phospho-proteins) with the frequent genomic alterations in the pathways (see the example of PTEN deletions in Figure S7B).

Recurrently altered genes and their patterns across subtypes

The overall mutational burden of the cohort, inferred from whole-exome sequencing, was 0.94 mutations per megabase (median, range 0.04 – 28 per megabase), which corresponds to 19 non-synonymous mutations per tumor genome (median; 13–25, 25th and 75th percentiles respectively). This is consistent with prior exome and genome-scale sequencing results for localized prostate cancers (Barbieri et al., 2012, Baca et al., 2013), and is lower than the mutational burden of metastatic prostate cancers (Gundem et al., 2015, Grasso et al., 2012, Robinson et al., 2015). These results reaffirm that prostate cancer possesses a lower mutational burden than many other epithelial tumor types that are not associated with a strong exogenous mutagen (Alexandrov et al., 2013, Lawrence et al., 2013). Prior exome sequencing of 112 prostate cancers identified 12 recurrently mutated genes through focused assessment of point mutations and short insertions and deletions (Barbieri et al., 2012). By comparison, mutational significance analysis of these 333 tumor-normal pairs by MutSigCV (Lawrence et al., 2013, 2014) identified 13 significantly mutated genes (q -value < 0.1), seven of which had not been previously identified (Figure 2 and Tables S1B–S1C). Among the significantly mutated genes, *SPOP*, *TP53*, *FOXA1*, *PTEN*, *MED12*, and *CDKN1B* were previously identified as recurrently mutated. Additional clinically relevant genes were identified with lower mutation frequencies; these included genes within canonical kinase signaling pathways (*BRAF*, *HRAS*, *AKT1*), the beta-catenin pathway (*CTNNB1*), and the DNA repair pathway (*ATM*). The rate of *BRAF* mutations (2.4%) seen in this study is higher than previously reported; these include several known activating mutations, but curiously not the canonical V600E hotspot. We identified no *BRAF* fusions, which had previously been reported in a subset of clinically advanced prostate cancer (Palanisamy et al., 2010). *NKX3-1*, previously implicated in familial prostate cancer syndromes and often found to be deleted, was also somatically mutated in this cohort (1% of tumors). While its functional

significance is unknown, *ZMYM3*, an epigenetic regulatory protein not previously implicated in prostate cancer but infrequently mutated in Ewing sarcomas (Tirode et al., 2014) and various pediatric cancers (Huether et al, 2014), was also recurrently mutated (2% of tumors). Genes with known biological relevance that were mutated at frequencies just below the threshold of significance (q-value < 0.01) included *KMT2C* (*MLL3*), *KMT2D* (*MLL4*), *APC*, *IDH1*, and *PIK3CA* (Figure 2 and Tables S1B–S1C). Mutations in the tumor suppressor genes *KMT2C*, *KMT2D*, and *APC* were mostly truncating; the *IDH1* and *PIK3CA* mutations occurred in previously characterized hotspots and thus may have therapeutic relevance for those occasional tumors with these mutations.

Notwithstanding these key somatic mutations, the most frequent molecular abnormalities involved chromosomal arm-level copy-number alterations (Taylor et al., 2010). These alterations included recurrent genomic gains of chromosome 7 and 8q and heterozygous losses of 8p, 13q, 16q and 18 (Figure S3A). Significance analysis of recurrent focal DNA copy-number alterations revealed 20 amplifications and 35 deletions (q-value < 0.25, GISTIC 2.0; Figure S3A and Table S1D). Recurrent focal amplifications included those spanning known oncogenes such as *CCND1* (11q13.2, 2%), *MYC* (8q24.21, 8%), as well as *FGFR1* and *WHSC1L1* (8p11.23, 8%). Recurrent focal deletions were much more common: Homozygous deletions spanning the *PTEN* locus occurred at one of the highest rates of any tumor type studied thus far (15%). Focal deletions of the region between the *TMPRSS2* and *ERG* genes on 21q22.3, which result in *TMPRSS2-ERG* fusions, were unique to prostate cancers as expected. Other focal deletions include those spanning tumor suppressors *TP53* (17p13.1), *CDKN1B* (12p13.1), *MAP3K1* (5q11.2), *FANCD2* (3p26), as well as *SPOPL* (2q22.1) and the complex locus spanning *FOXPI/RYPB/SHQ1* (3p13). *MAP3K7* (6q.12-22) was also frequently deleted, along with deletion of *CHD1* (5q15-q21); co-deletion of these loci has been associated with aggressive ETS-negative prostate cancer (Kluth et al., 2013, Rodrigues et al., 2015).

As the pattern and extent of SCNAs in prostate cancer genomes have been associated with probability of disease recurrence and metastasis in primary prostate cancers (Taylor et al., 2010, Hieronymus et al., 2014, van Dekken et al., 2004, Paris et al., 2004), we sought to identify similar structure in the burden of SCNAs by performing hierarchical clustering of arm-level alterations. We identified three major groups of prostate cancers, one with mostly unaltered genomes (hereafter referred to as *quiet*), a second group encompassing 50% of all tumors with an intermediate level of SCNAs, and a third group with a high burden of arm level genomic gains and losses (Figure S3B–S3C). While a formal outcome analysis was not possible due to the limited clinical follow-up available for this cohort, the subset of tumors with the greatest burden of SCNAs had significantly higher Gleason scores and PSA levels than the other two groups (Figure S3B–S3D). The tumors in this group also had significantly higher tumor cellularity (Figure S3C).

Epigenetic changes define molecularly distinct subtypes of prostate cancer

Integrative analysis of genetic and epigenetic changes revealed a diversity of DNA methylation changes that defined molecularly distinct subsets of primary prostate cancer (Figure 3). Unsupervised hierarchical clustering of the most variably hypermethylated CpGs

identified four epigenetically distinct groups of prostate cancers (Figure S5A–S5B). When integrated with the molecular taxonomy defined above, we found a number of striking associations. Among these was a notable pattern within ERG fusion-positive tumors. Specifically, while nearly two-thirds of all ERG fusion-positive tumors belonged to an unsupervised cluster with only moderately elevated DNA methylation (DNA methylation cluster 3), the remaining ERG fusion-positive tumors comprised a distinct hypermethylated cluster (cluster 1) that was almost exclusively associated with ERG fusions. On average, this cluster contained twice the number of hypermethylated loci as DNA methylation cluster 3 (Figure S5A), and the epigenetic patterns were largely distinct from those of *ETV1* and *ETV4* fusion-positive tumors, which showed more heterogeneous methylation. What drives these epigenetically distinct groups of ETS fusion-positive tumors is unknown, but there is considerable diversity in their DNA methylation profiles that may reflect altered epigenetic silencing (Figure S5A–S5B). Together, these results support further ETS fusion-based subtyping of disease, but also reveal a greater molecular and likely biological diversity among ERG fusion-positive tumors than previously appreciated. Likewise, these results are consistent with *in vivo* mouse modeling and expression profiling studies that suggest important molecular and clinicopathological differences between *ERG* and non-*ERG* ETS fusion positive tumors (Baena et al., 2013, Tomlins et al., 2015).

SPOP and *FOXA1*-mutant tumors exhibited homogeneous epigenetic profiles. These tumors belonged almost exclusively to DNA methylation cluster 2, a group that also contained a majority of the *ETV1* and *ETV4* but not *ERG*-positive tumors. Lastly, the *IDH1*-mutant tumors were notable given their strongly elevated levels of genome-wide DNA hypermethylation (Figure S5B). While of low incidence, these *IDH1* R132-mutant tumors defined a distinct subgroup of what appears to be early onset prostate cancer (Figure 3B) that possesses fewer DNA copy-number alterations (see Figure 1) or other canonical genomic lesions that are common to most other prostate cancers. *IDH1* and *IDH2* mutations have been associated with a DNA methylation phenotype in other tumor types, most notably in gliomas (Noushmehr et al., 2010) and acute myeloid leukemias (AML, Figueroa et al., 2010). Curiously, *IDH1*-mutant prostate cancers possessed even greater levels of genome-wide hypermethylation than do either glioma or AML *IDH1*-mutant tumors (Figure 3B). After further investigating DNA methylation differences between IDH-mutant and wild type tumors among prostate cancers, gliomas, and AMLs, we found that hypermethylated loci were specific to the cancer type rather than IDH mutants (Figure S5F).

Integrating these epigenetic data with mRNA expression levels, we identified 164 genes that were epigenetically silenced in subsets of the cohort (Figure S5C, Table S1F). These silenced genes were significantly enriched for genes previously found to be differentially expressed in prostate cancer; specifically, genes that are down-regulated in metastatic prostate cancer (Chandran et al., 2007) and genes involved in prostate organ development (Schaeffer et al., 2008) (q -value $< 2.0 \cdot 10^{-5}$). These 164 silenced genes displayed heterogeneous frequencies of epigenetic silencing across the cohort. For example, *SHF*, *FAXDC2*, *GSTP1*, *ZNF154*, and *KLF8* were epigenetically silenced in almost all tumors (>85%) whereas *STAT6* was silenced predominantly in ETS fusion-positive tumors and not in *SPOP* and *IDH1*-mutant tumors. Conversely, *HEXA* was silenced preferentially in *SPOP*-

mutant tumors compared to ERG fusion-positive tumors (86.5 versus 14.5% respectively, $p < 5.4 \cdot 10^{-15}$). Consistent with their increased DNA hypermethylation, the *IDH1*-mutant prostate tumors also possessed the greatest number of epigenetically silenced genes among all prostate tumors (Table S1F).

AR activity is variable in primary prostate cancers

The androgen receptor (AR) regulates normal prostate development as well as critical growth and survival programs in prostate carcinoma. Primary prostate cancer is androgen dependent, and androgen activity is a central axis in prostate cancer pathogenesis, driving the creation and overexpression of most ETS fusion genes (Lin et al., 2011, Mani et al., 2009, Tomlins et al., 2005). However, the extent to which individual primary prostate cancers differ in androgen sensitivity or dependence is unknown, and the issue has translational implications because AR targeting is therapeutically important. To address these questions, we sought to infer the AR output of tumors by calculating an AR activity score from the expression pattern of 20 genes that are experimentally validated AR transcriptional targets (Hieronymus et al., 2006). This score suggested that a broad spectrum of AR activity exists across all prostate tumors as well as between genomic subtypes (Figure 4A). Although ETS fusion genes are under AR control, the ETS fusion-positive groups had variable AR transcriptional activity. In contrast, we found that tumors with *SPOP* or *FOXAI* mutations had the highest AR transcriptional activity of all genotypically distinct subsets of prostate cancer ($p = 1.1 \cdot 10^{-6}$ and 0.04, respectively, t-test). Consistent with this, *SPOP* mutations have been previously implicated in androgen signaling in model systems, since both AR and AR coactivators are substrates deregulated by *SPOP* mutation (Geng et al. 2013, An et al., 2014, Geng et al. 2014), providing a possible explanation for the associated increase in AR activity seen in this subtype of prostate cancers.

While AR transcriptional output is a proxy for ligand-driven AR activity in many tumors, *AR* transcript variants have been described that encode truncated AR proteins that lack the ligand-binding domain and hence are capable of activating AR target genes in the absence of androgens (Dehm et al., 2008, Watson et al., 2010). Using RNA sequencing reads that spanned the splice junctions unique to each *AR* variant, we quantified the expression of these *AR* transcript variants. This analysis revealed that several *AR* splice variants, most notably AR-V7, can be detected at low levels in primary tumors and, in a few cases, in adjacent benign prostate tissue (Figure 4B), and we validated these expression levels with qPCR (Figure 4C). However, their expression was not associated with differential expression of known AR target genes or with the seven previously defined genomic subtypes. Most detected splice forms were truncated after the DNA-binding domain by the presence of a cryptic exon rather than by skipping those exons encoding the ligand-binding domain. Truncated *AR* splice variants were previously assumed to be expressed primarily in metastatic castration-resistant prostate cancers, where, at least for AR-V7, their presence was associated with resistance to hormone therapy (Antonarakis et al., 2014). Hence, our finding that they are expressed in hormone-naïve primary prostate cancers is notable.

In prostate cancers, the degree of AR pathway output is controlled not only by AR mRNA and protein expression levels, but also by expression of and mutations in AR cofactors

(Heemers et al., 2007). It is therefore notable that FOXA1 was recurrently mutated in our cohort, as it is a pioneering transcription factor that targets AR and has a demonstrated role in prostate cancer oncogenesis (Jin et al., 2013). We identified *FOXA1* mutations in 4% of the primary prostate cancers studied here, which is similar to the mutation frequency observed previously (Barbieri et al., 2012, Grasso et al., 2012) (Figure 4A). While a subset of these mutations were present in tumors that also possessed *SPOP* mutations and had elevated levels of AR output, *FOXA1* mutations were mutually exclusive with all other alterations that define the genomic subclasses described here. While there were some truncating mutations near the C-terminus and the C-terminal part of the forkhead domain, the majority of the mutations found here and in other prostate cancer cohorts were missense mutations that primarily affect the winged-helix DNA binding domain of FOXA1. Curiously, these mutations do not directly alter FOXA1 DNA-binding residues (Figure 4D–E), a pattern similar to the *FOXA1* mutations recently found in lobular breast cancers (TCGA, unpublished), which suggests that the impact of FOXA1 mutations has less to do with altering DNA binding than with disrupting or altering interactions with other chromatin-bound cofactors.

Clinically actionable DNA repair defects in primary prostate cancers

Prior data indicate that several DNA repair pathways are disrupted in a subset of prostate cancers (Karanika et al., 2014, Pritchard et al., 2014). Moreover, the PARP inhibitor olaparib is effective in some patients with prostate cancer (Mateo et al., 2014). Here, we found inactivation of several DNA repair genes that collectively affected 19% of patients (Figure 5A). While we found only one inactivating *BRCA1* germline mutation, a frameshift at V923 caused by a 4bp deletion (Clinvar RCV000083190.3), *BRCA2* inactivation affected 3% of tumors, including both germline and somatic truncating mutations. All six *BRCA2* germline mutations were K3326*, a C-terminal truncating mutation with debated functional impact but increased prevalence in several tumor types (Farrugia et al., 2008, Martin et al., 2005, Delahaye-Sourdeix et al., 2015). Two additional tumors possessed focal *BRCA2* homozygous deletions that were accompanied by very low *BRCA2* transcript expression. Four tumors (1%) possessed either loss-of-function mutations or homozygous deletion of *CDK12*, a gene that has been implicated in DNA repair by regulating expression levels of several DNA damage response genes (Blazek et al., 2011) and is recurrently mutated in metastatic prostate cancer (Grasso et al., 2012). *ATM*, an apical kinase of the DNA damage response, which is activated by the Mre11 complex and mediates downstream checkpoint signaling, was affected by a nonsense mutation in one case and by a likely kinase-dead hotspot N2875 mutation in two cases. *FANCD2* was similarly affected by diverse uncommon lesions including a truncating mutation in one tumor, homozygous deletion in two tumors, and focal heterozygous losses in 6% of the cohort (Figure 5B). *RAD51C* (3%) was affected by focal DNA losses, most of which were heterozygous. Finally, it was notable that heterozygous losses of *BRCA2* (13q13.1) almost always coincided with concurrent loss of the distant *RBI* tumor suppressor gene (13q14.2) (Figure S3D). The observation that nearly 20% of primary prostate cancers bear genomic defects involving DNA repair pathways is remarkably consistent with the recently announced TOPARP-A Phase II trial results in patients with metastatic castration-resistant prostate cancer indicating that clinical

responses to the PARP inhibitor olaparib likely occurred in the subgroup of tumors bearing defects in DNA repair genes (Mateo et al., 2014, Robinson et al., 2015).

Clinically actionable lesions in PI3K and Ras signaling

The long tail of the frequency distribution of molecular abnormalities is particularly notable among primary prostate cancers. Beyond *PTEN*, which was deleted or mutated in 17% of the cohort, various driver mutations in effectors of PI3K signaling were present at low incidence (Figure 5C). *PIK3CA*, which encodes the 110 kDa catalytic subunit of phosphatidylinositol 3-kinase, was mutated in six tumors including one case possessing coincident activating mutations (E542A and N345I), both of which appeared to be subclonal. The other four *PIK3CA* mutations were all known activating mutational hotspots (E545K, Q546K, N345I, C420R), while one had a mutation of unknown function (E474A). Focal *PIK3CA* amplification with associated mRNA overexpression occurred in ~1% of cases. Interestingly, *PIK3CB* was mutated in two tumors that also possessed coincident homozygous deletions of *PTEN*, both of which were clonal. *PIK3CB* E552K was found in one tumor at a paralogous residue to the canonical *PIK3CA* helical domain E545K mutant, and is presumably activating (Figure 5E). As *PTEN*-deleted tumors are likely *PIK3CB*-dependent due to the feedback inhibition of *PIK3CA*, co-existent loss and mutation of *PTEN* and *PIK3CB* may be elevating PI3K pathway output and perhaps indicating a set of tumors in which combined PI3K and androgen signaling inhibition may be effective (Schwartz et al., 2015). Among other lesions that drive PI3K signaling, *AKT1* was mutated in three tumors. Two tumors had the known E17K hotspot mutation, while another encoded a D323Y mutation. Whereas E17K is the most common hotspot in *AKT1* across human cancer, the D323Y variant is uncommon, having been identified previously in one lung adenocarcinoma (Cancer Genome Atlas Research Network, 2014) and one urothelial bladder cancer (Guo et al., 2013). Nevertheless, while distant linearly from the activating E17K hotspot, in three dimensions this D323Y kinase domain mutant directly abuts the PH-domain containing E17K (Figure 5D), and has been described as potentially activating (Parikh et al., 2012).

We also identified known or presumed driver mutations in several other genes of the MAPK pathway, affecting 25% of the tumors (Figure 5A). *HRAS* was mutated in four tumors, of which three were Q61R hotspot mutations. Two mutations arose in other Ras-family small GTPases. While both *RAC1* Q61R and *RRAS2* Q72L occurred only once each, they affected residues paralogous to the RAS Q61 hotspot (Figure 5E) (Chang et al., 2015). We also identified eight *BRAF* mutations, though, curiously, none were the common V600E mutation that is prevalent in cutaneous melanomas, thyroid cancers, and many other tumor types. Five *BRAF* mutations are likely activating, including known hotspots (K601E, G469A, L597R), two of which confer sensitivity to MEK inhibitors (Dahlman et al., 2012, Bowyer et al., 2014). Another mutation was a likely activating in-frame 3 amino acid deletion at K601 (Figure 5F), while the final mutation (F468C) affected the adjacent residue to the known G469 hotspot. Together, these findings reveal a long tail of low-incidence potentially actionable predicted driver mutations present across the molecular taxonomy of prostate cancer.

Comparison with metastatic prostate cancer

To put these results in context, we compared our findings with those from a recently published cohort of 150 castration-resistant metastatic prostate cancer samples (Robinson et al., 2015). The analysis revealed some similarities and many differences between primary and treated metastatic disease. Although the overall burden of copy-number alterations and mutations was significantly higher in the metastatic samples (Figure 6A), consistent with previous findings (Taylor et al., 2010, Grasso et al., 2012), the primary and metastatic samples were remarkably similar in their subtype distribution, with the exception that the metastatic data set contained no *IDH1*-mutant tumors (Figure 6B). We compared the frequencies of all recurrently altered genes described in both studies and found that, similar to the overall burden of genomic alterations (Figure 6A), many genes and pathways have increased alteration rates in the metastatic samples (Figure 6C, Table S3). Androgen receptor signaling was more frequently altered in the metastatic samples, most often by amplification or mutation of *AR*, events that were essentially absent in primary samples. Interestingly, *SPOP* mutations were somewhat less frequent in the metastatic samples (8% vs 11% in the primary samples). DNA repair and PI3K pathway alterations were more frequent in the metastatic samples, as were mutations or deletions of *TP53*, *RBI*, *KMT2C* and *KMT2D*. Interestingly, we found no focal, clonal *MYCL* amplifications, which were recently described in primary prostate cancer (Boutros et al., 2015), in either data set nor in a separate set of 63 untreated prostate cancer samples (Hovelson et al., 2015).

Discussion

The comprehensive molecular analyses of primary prostate cancers presented here reveal highly diverse genomic, epigenomic, and transcriptomic patterns. Major subtypes could be defined by fusions of the ETS family genes *ERG*, *ETV*, *ETV4*, or *FLII* and by mutations in *SPOP*, *FOXAI*, or *IDHI*. However, even within the groups there was significant diversity in DNA copy-number alterations, gene expression, and DNA methylation. The mutational heterogeneity mirrors the heterogeneous natural history of primary prostate cancers.

Although the broad spectrum of copy-number alterations in tumors with ETS fusions has been previously characterized (Demichelis et al., 2009, Taylor et al., 2010), here we uncovered additional differences between the epigenetic profiles of those tumors. We found that *ERG* fusion-positive tumors can be subdivided into two methylation subtypes: one with lower levels of methylation, and one with a distinct spectrum of hypermethylation. Many genes were epigenetically silenced as a result of the hypermethylation in the latter tumors. While further studies will be required to determine which silencing events are linked to prostate cancer pathogenesis, the findings presented here reveal variability among what was previously considered to be genetically homogeneous prostate cancer subtypes.

We have also identified a distinct subgroup of tumors with *IDHI* R132 mutations that is associated with younger age at diagnosis. Although *IDHI* mutations have previously been identified in prostate cancer with a similar incidence (2.7%) (Ghiam et al., 2012, Kang et al., 2009), we show here that those tumors are all ETS fusion-negative and *SPOP* wildtype, have little SCNA burden, and possess elevated levels of genome-wide methylation. The levels of methylation observed in this methylator phenotype are higher than those observed

in *IDH1*-mutant GBMs and AMLs. Consistent with our observations, a recently published clinical study of 117 prostate cancers identified a single *IDH1*-mutant prostate cancer from 56-year old patient that also lacked significant copy number alterations, ETS gene fusions, or driver mutations (Hovelson et al., 2015). Future studies in cohorts with sufficient clinical follow-up will be able to ask whether the *IDH1*-mutant prostate cancers are prognostically distinct, as noted for gliomas (Noushmehr et al., 2010) and AMLs (Mardis et al., 2009), and if they are sensitive to newly developed *IDH1*-targeted therapeutics (Rohle et al., 2013).

Interestingly, 26% of the tumors in this study could not be characterized by one of the taxonomy-defining cardinal genomic alterations. The 26% were clinically and genomically heterogeneous, with some tumors exhibiting extensive DNA copy-number alterations and high Gleason scores indicative of poorer prognosis. About a third of them were genomically similar to *SPOP* and *FOXAI* mutant tumors but lacked any canonical mutation (iCluster 1, methylation cluster 2, mRNA cluster 1); others were enriched for mutations of *TP53*, *KDM6A*, and *KMT2D* or specific SCNAs spanning *MYC* and *CCND1*. Many of the tumors had low Gleason score with few if any DNA copy-number alterations and a normal-like DNA methylation pattern. As previously reported, tumors with fewer genomic alterations were also more commonly Gleason score 6 tumors (38% in the ‘quiet’ class vs 8% in the class with the greatest burden of alterations).

Tumor cellularity, as assessed by pathology review, was lower among tumors with fewer SCNAs (one-sided Mann-Whitney test, $p = 0.0002$), indicating that the apparent lower burden of alterations in tumors with smaller volumes may be due in part to their tumor purities being lower. However, the lower cellularity of these tumors did not limit the detection of clonal molecular lesions since tumor cellularity between ETS fusion-positive and these fusion-negative tumors was not significantly different (two-sided Mann-Whitney test, $p = 0.32$). One must also keep in mind that this study was limited to a single tumor focus for each patient, even though the vast majority of primary prostate tumors are multifocal and molecular heterogeneity between different foci has been demonstrated (Cooper et al., 2015, Boutros et al., 2015, Lindberg et al., 2013). Such issues must be considered when designing new therapeutic approaches and biomarker panels for clinical use, as patients likely have more than one of these molecular subtypes present due to this commonly occurring tumor multifocality and molecular heterogeneity.

Primary prostate cancers exhibit a wide range of androgen receptor activity. This study demonstrates for the first time a direct association between mutations in *SPOP* or *FOXAI* and increased AR-driven transcription in human prostate cancers. Further studies in preclinical models as well as in clinical trial settings will be required to understand the implications of variable AR activity in the contexts of chemoprevention and prostate cancer-directed treatment strategies (Mostaghel et al., 2010). Other, more immediately actionable opportunities for targeted therapy exist for the 19% of primary prostate cancers that have defects in DNA repair, and for the nearly equal number of cancers with altered key effectors of both PI3K and MAPK pathways. While the numbers of DNA repair defects found in organ-confined prostate tumors may be lower than those found in metastatic prostate cancer (Robinson et al., 2015), an increase in the number of such defects with disease progression suggests a possible advantage to targeting DNA repair-deficient tumors at an earlier stage of

disease, perhaps at initial diagnosis. Such strategies may include preventing DNA damage as well as targeting deficient DNA repair (Ferguson et al., 2015). Alterations in the PI3K/MTOR pathway also play an important role: beyond the frequent inactivation of *PTEN*, we document rare activation of *PIK3CA*, *PIK3CB*, *AKT1*, and *MTOR*, and of several small GTPases, including *HRAS*, as well as *BRAF*. As DNA sequencing of tumor samples becomes more widely adopted earlier in the clinical care of cancer patients, such alterations may emerge as candidates for inclusion in clinical trials after front-line therapy.

In summary, our integrative assessment of 333 primary prostate cancers has confirmed previously defined molecular subtypes across multiple genomic platforms and identified novel alterations and subtype diversity. It provides a resource for continued investigation into the molecular and biological heterogeneity of the most common cancer in American men.

Experimental Procedures

Tumor and matched normal specimens were obtained from prostate cancer patients that provided informed consent and were approved for collection and distribution by local Institutional Review boards. Blocks frozen in OCT were made of all tumors and of paired benign tissue when present. A 5 micron section was cut from both the top and bottom of the OCT block of 111 tumor cases and from the top or bottom only of the OCT block of 222 tumor cases. Out of 39 normal samples included in the freeze 23 underwent pathology review, and prostate origin (i.e., no seminal vesicles) and absence of tumor and high grade prostate intraepithelial neoplasia (HGPIN) was confirmed. Tissue images were reviewed by eight genitourinary pathologists, who reported the primary and secondary Gleason patterns of cancer for each slide and estimates of tumor cellularity in 10% increments (from 0–100%). In case of discrepancies of Gleason scores between the top and bottom sections, the Gleason scores of cancer in the section with the largest area of tumor was used. A subset of 54 cases was reviewed by two pathologists. Discrepancies that occurred between the two pathologists were reconciled by blind review by a third pathologist.

DNA, RNA, and protein were purified and distributed throughout the TCGA network. Samples with evidence for RNA degradation were excluded from the study (Supplementary Methods). In total, 333 primary tumors with associated clinicopathologic data were assayed on at least four molecular profiling platforms. Platforms included exome and whole genome DNA sequencing, RNA sequencing, miRNA sequencing, SNP arrays, DNA methylation arrays, and reverse phase protein arrays. Integrated multiplatform analyses were performed.

The data and analysis results can be explored through the Broad Institute FireBrowse portal (<http://firebrowse.org/?cohort=PRAD>), the cBioPortal for Cancer Genomics (http://www.cbioportal.org/study.do?cancer_study_id=prad_tcga_pub), TCGA Batch Effects (<http://bioinformatics.mdanderson.org/tcgambatch/>), Regulome Explorer (<http://explorer.cancerregulome.org/>) and Next-Generation Clustered Heat Maps (<http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/>). See also Supplemental Information and the TCGA publication page (https://tcga-data.nci.nih.gov/docs/publications/prad_2015/).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to all the patients and families who contributed to this study. We thank Margi Sheth and Ina Felau for project management. Supported by the following grants from the United States National Institutes of Health: 5U24CA143799, 5U24CA143835, 5U24CA143840, 5U24CA143843, 5U24CA143845, 5U24CA143848, 5U24CA143858, 5U24CA143866, 5U24CA143867, 5U24CA143882, 5U24CA143883, 5U24CA144025, U54HG003067, U54HG003079, and U54HG003273, P30CA16672.

Consortia

The members of The Cancer Genome Atlas Research Network are Adam Abeshouse, Jaeil Ahn, Rehan Akbani, Adrian Ally, Samirkumar Amin, Christopher D. Andry, Matti Annala, Armen Aprikian, Joshua Armenia, Arshi Arora, J. Todd Auman, Miruna Balasundaram, Saianand Balu, Christopher E. Barbieri, Thomas Bauer, Christopher C. Benz, Alain Bergeron, Rameen Beroukham, Mario Berrios, Adrian Bivol, Tom Bodenheimer, Lori Boice, Moiz S. Bootwalla, Rodolfo Borges dos Reis, Paul C. Boutros, Jay Bowen, Reanne Bowlby, Jeffrey Boyd, Robert K. Bradley, Anne Breggia, Fadi Brimo, Christopher A. Bristow, Denise Brooks, Bradley M. Broom, Alan H. Bryce, Glenn Bubley, Eric Burks, Yaron S.N. Butterfield, Michael Button, David Canes, Carlos G. Carlotti, Rebecca Carlsen, Michel Carmel, Peter R. Carroll, Scott L. Carter, Richard Cartun, Brett S. Carver, June M. Chan, Matthew T. Chang, Yu Chen, Andrew D. Cherniack, Simone Chevalier, Lynda Chin, Juok Cho, Andy Chu, Eric Chuah, Sudha Chudamani, Kristian Cibulskis, Giovanni Ciriello, Amanda Clarke, Matthew R. Cooperberg, Niall M. Corcoran, Anthony J. Costello, Janet Cowan, Daniel Crain, Erin Curley, Kerstin David, John A. Demchok, Francesca Demichelis, Noreen Dhalla, Rajiv Dhir, Alexandre Doueik, Bettina Drake, Heidi Dvinge, Natalya Dyakova, Ina Felau, Martin L. Ferguson, Scott Frazer, Stephen Freedland, Yao Fu, Stacey B. Gabriel, Jianjiong Gao, Johanna Gardner, Julie M. Gastier-Foster, Nils Gehlenborg, Mark Gerken, Mark B. Gerstein, Gad Getz, Andrew K. Godwin, Anuradha Gopalan, Markus Graefen, Kiley Graim, Thomas Gribbin, Ranabir Guin, Manaswi Gupta, Angela Hadjipanayis, Syed Haider, Lucie Hamel, D. Neil Hayes, David I. Heiman, Julian Hess, Katherine A. Hoadley, Andrea H. Holbrook, Robert A. Holt, Antonia Holway, Christopher M. Hovens, Alan P. Hoyle, Mei Huang, Carolyn M. Hutter, Michael Ittmann, Lisa Iype, Stuart R. Jefferys, Corbin D. Jones, Steven J.M. Jones, Hartmut Juhl, Andre Kahles, Christopher J. Kane, Katayoon Kasaian, Michael Kerger, Ekta Khurana, Jaegil Kim, Robert J. Klein, Raju Kucherlapati, Louis Lacombe, Marc Ladanyi, Phillip H. Lai, Peter W. Laird, Eric S. Lander, Mathieu Latour, Michael S. Lawrence, Kevin Lau, Tucker LeBien, Darlene Lee, Semin Lee, Kjong-Van Lehmann, Kristen M. Leraas, Ignaty Leshchiner, Robert Leung, John A. Libertino, Tara M. Lichtenberg, Pei Lin, W. Marston Linehan, Shiyun Ling, Scott M. Lippman, Jia Liu, Wenbin Liu, Lucas Lochovsky, Massimo Cell 163, 1011–1025, November 5, 2015 ©2015 Elsevier Inc. 1021 Loda, Christopher Logothetis, Laxmi Lolla, Teri Longacre, Yiling Lu, Jianhua Luo, Yussanne Ma, Harshad S. Mahadeshwar, David Mallery, Armaz Mariamidze, Marco A. Marra, Michael Mayo, Shannon McCall, Ginette McKercher, Shaowu Meng, Anne-Marie Mes-Masson, Maria J. Merino, Matthew Meyerson, Piotr A. Mieczkowski, Gordon B. Mills, Kenna R. Mills Shaw, Sarah Minner, Alireza

Moinzadeh, Richard A. Moore, Scott Morris, Carl Morrison, Lisle E. Mose, Andrew J. Mungall, Bradley A. Murray, Jerome B. Myers, Rashi Naresh, Joel Nelson, Mark A. Nelson, Peter S. Nelson, Yulia Newton, Michael S. Noble, Houtan Noushmehr, Matti Nykter, Angeliki Pantazi, Michael Parfenov, Peter J. Park, Joel S. Parker, Joseph Paulauskis, Robert Penny, Charles M. Perou, Alain Piché, Todd Pihl, Peter A. Pinto, Davide Prandi, Alexei Protopopov, Nilsa C. Ramirez, Arvind Rao, W. Kimryn Rathmell, Gunnar Rättsch, Xiaojia Ren, Victor E. Reuter, Sheila M. Reynolds, Suhn K. Rhie, Kimberly Rieger-Christ, Jeffrey Roach, A. Gordon Robertson, Brian Robinson, Mark A. Rubin, Fred Saad, Sara Sadeghi, Gordon Saksena, Charles Saller, Andrew Salner, Francisco Sanchez-Vega, Chris Sander, George Sandusky, Guido Sauter, Andrea Sboner, Peter T. Scardino, Eleonora Scarlata, Jacqueline E. Schein, Thorsten Schlomm, Laura S. Schmidt, Nikolaus Schultz, Steven E. Schumacher, Jonathan Seidman, Luciano Neder, Sahil Seth, Alexis Sharp, Candace Shelton, Troy Shelton, Hui Shen, Ronglai Shen, Mark Sherman, Margi Sheth, Yan Shi, Juliann Shih, Ilya Shmulevich, Jeffrey Simko, Ronald Simon, Janae V. Simons, Payal Sipahimalani, Tara Skelly, Heidi J. Sofia, Matthew G. Soloway, Xingzhi Song, Andrea Sorcini, Carrie Sougnez, Serghei Stepa, Chip Stewart, John Stewart, Joshua M. Stuart, Travis B. Sullivan, Charlie Sun, Huandong Sun, Angela Tam, Donghui Tan, Jiabin Tang, Roy Tarnuzzer, Katherine Tarvin, Barry S. Taylor, Patrick Teebagy, Imelda Tenggara, Bernard Têtu, Ashutosh Tewari, Nina Thiessen, Timothy Thompson, Leigh B. Thorne, Daniela P. Tirapelli, Scott A. Tomlins, Felipe Amstalden Trevisan, Patricia Troncoso, Lawrence D. True, Maria Christina Tsourlakis, Svitlana Tyekucheva, Eliezer Van Allen, David J. Van Den Berg, Umadevi Veluvolu, Roel Verhaak, Cathy D. Vocke, Doug Voet, Yunhu Wan, Qingguo Wang, Wenyi Wang, Zhining Wang, Nils Weinhold, John N. Weinstein, Daniel J. Weisenberger, Matthew D. Wilkerson, Lisa Wise, John Witte, Chia-Chin Wu, Junyuan Wu, Ye Wu, Andrew W. Xu, Shalini S. Yadav, Liming Yang, Lixing Yang, Christina Yau, Huihui Ye, Peggy Yena, Thomas Zeng, Jean C. Zenklusen, Hailei Zhang, Jianhua Zhang, Jiashan Zhang, Wei Zhang, Yi Zhong, Kelsey Zhu, and Erik Zmuda.

References

- Ahn J, Yuan Y, Parmigiani G, Suraokar MB, Diao L, Wistuba II, Wang W. DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. *Bioinformatics*. 2013 Aug 1; 29(15):1865–71. [PubMed: 23712657]
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, et al. Signatures of mutational processes in human cancer. *Nature*. 2013 Aug 22; 500(7463):415–21. [PubMed: 23945592]
- Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, Benlloch S, Hazelett DJ, Wang Z, Saunders E, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet*. 2014 Oct; 46(10):1103–9. [PubMed: 25217961]
- An J, Wang C, Deng Y, Yu L, Huang H. Destruction of full-length androgen receptor by wild-type SPOP, but not prostate-cancer-associated mutants. *Cell Rep*. 2014 Feb 27; 6(4):657–69. [PubMed: 24508459]
- Antonarakis ES, Lu C, Wang H, Lubner B, Nakazawa M, Roeser JC, Chen Y, Mohammad TA, Chen Y, Fedor HL, et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med*. 2014 Sep 11; 371(11):1028–38. [PubMed: 25184630]
- Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013 Apr 25; 153(3):666–77. [PubMed: 23622249]

- Baena E, Shao Z, Linn DE, Glass K, Hamblen MJ, Fujiwara Y, Kim J, Nguyen M, Zhang X, Godinho FJ, et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev.* 2013 Mar 15; 27(6):683–98. [PubMed: 23512661]
- Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet.* 2012 May 20; 44(6):685–9. [PubMed: 22610119]
- Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, et al. The genomic complexity of primary human prostate cancer. *Nature.* 2011 Feb 10; 470(7333):214–20. [PubMed: 21307934]
- Blattner M, Lee DJ, O'Reilly C, Park K, MacDonald TY, Khani F, Turner KR, Chiu YL, Wild PJ, Dolgalev I, et al. SPOP mutations in prostate cancer across demographically diverse patient cohorts. *Neoplasia.* 2014 Jan; 16(1):14–20. [PubMed: 24563616]
- Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, Cimernancic P, Ule J, Peterlin BM. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* 2011 Oct 15; 25(20):2158–72. [PubMed: 22012619]
- Börno ST, Fischer A, Kerick M, Fälth M, Laible M, Brase JC, Kuner R, Dahl A, Grimm C, Sayanjali B, et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. *Cancer Discov.* 2012 Nov; 2(11):1024–35. [PubMed: 22930729]
- Boutros PC, Fraser M, Harding NJ, de Borja R, Trudel D, Lalonde E, Meng A, Hennings-Yeomans PH, McPherson A, Sabelnykova VY, et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet.* 2015 May 25.
- Bowyer SE, Rao AD, Lyle M, Sandhu S, Long GV, McArthur GA, Raleigh JM, Hicks RJ, Millward M. Activity of trametinib in K601E and L597Q BRAF mutation-positive metastatic melanoma. *Melanoma Res.* 2014 Oct; 24(5):504–8. [PubMed: 24933606]
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014 Jul 31; 511(7511):543–50. [PubMed: 25079552]
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012 May; 30(5):413–21. [PubMed: 22544022]
- Chandran UR, Ma C, Dhir R, Bisceglia M, Lyons-Weiler M, Liang W, Michalopoulos G, Becich M, Monzon FA. Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer.* 2007 Apr 12; 7:64. [PubMed: 17430594]
- Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandath C, Gao J, Socci ND, Solit DB, Olshen AB, Schultz N, Taylor BS. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotech.* In press.
- Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, Kremeyer B, Butler A, Lynch AG, Camacho N, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet.* 2015 Apr; 47(4):367–72. [PubMed: 25730763]
- Cooperberg MR, Broering JM, Carroll PR. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. *J Natl Cancer Inst.* 2009 Jun 16; 101(12):878–87. [PubMed: 19509351]
- Dahlman KB, Xia J, Hutchinson K, Ng C, Hucks D, Jia P, Atefi M, Su Z, Branch S, Lyle PL, et al. BRAF(L597) mutations in melanoma are associated with sensitivity to MEK inhibitors. *Cancer Discov.* 2012 Sep; 2(9):791–7. [PubMed: 22798288]
- D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ, Wein A. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA.* 1998 Sep 16; 280(11):969–74. [PubMed: 9749478]
- Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, Tindall DJ. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res.* 2008 Jul 1; 68(13):5469–77. [PubMed: 18593950]

- Delahaye-Sourdeix M, Anantharaman D, Timofeeva MN, Gaborieau V, Chabrier A, Vallée MP, Lagiou P, Holcátová I, Richiardi L, Kjaerheim K, Agudo A, et al. A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *J Natl Cancer Inst.* 2015 Apr 2;(5):107. pii: djv037.
- Demichelis F, Setlur SR, Beroukhim R, Perner S, Korbel JO, Lafargue CJ, Pflueger D, Pina C, Hofer MD, Sboner A, et al. Distinct genomic aberrations associated with ERG rearranged prostate cancer. *Genes Chromosomes Cancer.* 2009 Apr; 48(4):366–80. [PubMed: 19156837]
- Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, Wadum L, Johnson K, Mentlick J, Tavtigian SV, et al. Functional assays for classification of BRCA2 variants of uncertain significance. *Cancer Res.* 2008 May 1; 68(9):3523–31. [PubMed: 18451181]
- Ferguson LR, Chen H, Collins AR, Connell M, Damia G, Dasgupta S, Malhotra M, Meeker AK, Amedei A, Amin A, et al. Genomic instability in human cancer: Molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. *Semin Cancer Biol.* 2015 Apr 10.
- Ferlay, J.; Soerjomataram, I.; Ervik, M., et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013.
- Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell.* 2010; 18:553–67. [PubMed: 21130701]
- Friedlander TW, Roy R, Tomlins SA, Ngo VT, Kobayashi Y, Azameera A, Rubin MA, Pienta KJ, Chinnaiyan A, Ittmann MM, et al. Common structural and epigenetic changes in the genome of castration-resistant prostate cancer. *Cancer Res.* 2012 Feb 1; 72(3):616–25. [PubMed: 22158653]
- Gasi D, van der Korput HA, Douben HC, de Klein A, de Ridder CM, van Weerden WM, Trapman J. Overexpression of full-length ETV1 transcripts in clinical prostate cancer due to gene translocation. *PLoS One.* 2011 Jan 26; 6(1):e16332. [PubMed: 21298110]
- Geng C, He B, Xu L, Barbieri CE, Eedunuri VK, Chew SA, Zimmermann M, Bond R, Shou J, Li 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.* 2013 Apr 23; 110(17):6997–7002. [PubMed: 23559371]
- Geng C, Rajapakshe K, Shah SS, Shou J, Eedunuri VK, Foley C, Fiskus W, Rajendran M, Chew SA, Zimmermann M, et al. Androgen receptor is the key transcriptional mediator of the tumor suppressor SPOP in prostate cancer. *Cancer Res.* 2014 Oct 1; 74(19):5631–43. [PubMed: 25274033]
- Ghiam AF, Cairns RA, Thoms J, Dal Pra A, Ahmed O, Meng A, Mak TW, Bristow RG. IDH mutation status in prostate cancer. *Oncogene.* 2012 Aug 16; 31(33):3826. [PubMed: 22120718]
- Gopalan A, Leversha MA, Satagopan JM, Zhou Q, Al-Ahmadie HA, Fine SW, Eastham JA, Scardino PT, Scher HI, Tickoo SK, et al. TMPRSS2-ERG gene fusion is not associated with outcome in patients treated by prostatectomy. *Cancer Res.* 2009 Feb 15; 69(4):1400–6. [PubMed: 19190343]
- Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature.* 2012 Jul 12; 487(7406):239–43. [PubMed: 22722839]
- Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JM, Papaemmanuil E, Brewer DS, Kallio HM, Högnäs G, Annala M, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature.* 2015 Apr 16; 520(7547):353–7. [PubMed: 25830880]
- Guo G, Sun X, Chen C, Wu S, Huang P, Li Z, Dean M, Huang Y, Jia W, Zhou Q, et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet.* 2013 Dec; 45(12):1459–63. [PubMed: 24121792]
- Heemers HV, Tindall DJ. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev.* 2007; 28(7):778–808. [PubMed: 17940184]

- Hieronymus H, Lamb J, Ross KN, Peng XP, Clement C, Rodina A, Nieto M, Du J, Stegmaier K, Raj SM, et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell*. 2006 Oct; 10(4):321–30. [PubMed: 17010675]
- Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, Heguy A, Huberman K, Bernstein M, Assel M, et al. Copy number alteration burden predicts prostate cancer relapse. *Proc Natl Acad Sci U S A*. 2014 Jul 29; 111(30):11139–44. [PubMed: 25024180]
- Hovelson DH, McDaniel AS, Cani AK, Johnson B, Rhodes K, Williams PD, Bandla S, Bien G, Choppa P, Hyland F, et al. Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. *Neoplasia*. 2015 Apr; 17(4):385–99. [PubMed: 25925381]
- Huether R, Dong L, Chen X, Wu G, Parker M, Wei L, Ma J, Edmonson MN, Hedlund EK, Rusch MC, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun*. 2014 Apr 8.5:3630. [PubMed: 24710217]
- Jin HJ, Zhao JC, Ogdan I, Bergan RC, Yu J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res*. 2013 Jun 15; 73(12):3725–36. [PubMed: 23539448]
- Kang MR, Kim MS, Oh JE, Kim YR, Song SY, Seo SI, Lee JY, Yoo NJ, Lee SH. Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. *Int J Cancer*. 2009 Jul 15; 125(2):353–5. [PubMed: 19378339]
- Karanika S, Karantanos T, Li L, Corn PG, Thompson TC. DNA damage response and prostate cancer: defects, regulation and therapeutic implications. *Oncogene*. 2014 Aug.18:0.
- Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst*. 1998 May 20; 90(10):766–71. [PubMed: 9605647]
- Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science*. 2013 Oct 4.342(6154):1235587. [PubMed: 24092746]
- Kim JH, Dhanasekaran SM, Prensner JR, Cao X, Robinson D, Kalyana-Sundaram S, Huang C, Shankar S, Jing X, Iyer M, et al. Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res*. 2011 Jul; 21(7):1028–41. [PubMed: 21724842]
- Kluth M, Hesse J, Heintz A, Krohn A, Steurer S, Sirma H, Simon R, Mayer PS, Schumacher U, Grupp K, et al. Genomic deletion of MAP3K7 at 6q12-22 is associated with early PSA recurrence in prostate cancer and absence of TMPRSS2:ERG fusions. *Mod Pathol*. 2013 Jul; 26(7):975–83. [PubMed: 23370768]
- Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, Sherlock G. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res*. 2011 Jul; 21(7):1017–27. [PubMed: 21521786]
- Kunju LP, Carskadon S, Siddiqui J, Tomlins SA, Chinnaiyan AM, Palanisamy N. Novel RNA hybridization method for the in situ detection of ETV1, ETV4, and ETV5 gene fusions in prostate cancer. *Appl Immunohistochem Mol Morphol*. 2014 Sep; 22(8):e32–40. [PubMed: 25203299]
- Lalonde E, Ishkanian AS, Sykes J, Fraser M, Ross-Adams H, Erho N, Dunning MJ, Halim S, Lamb AD, Moon NC, et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet Oncol*. 2014 Dec; 15(13):1521–32. [PubMed: 25456371]
- Lapointe J, Li C, Giacomini CP, Salari K, Huang S, Wang P, Ferrari M, Hernandez-Boussard T, Brooks JD, Pollack JR. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res*. 2007 Sep 15; 67(18):8504–10. [PubMed: 17875689]
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013 Jul 11; 499(7457):214–8. [PubMed: 23770567]
- Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014 Jan 23; 505(7484):495–501. [PubMed: 24390350]

- Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, Rosenfeld MG. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell*. 2009 Dec 11; 139(6):1069–83. [PubMed: 19962179]
- Lindberg J, Kristiansen A, Wiklund P, Grönberg H, Egevad L. Tracking the origin of metastatic prostate cancer. *Eur Urol*. 2015 May; 67(5):819–22. [PubMed: 25246081]
- Mahapatra S, Klee EW, Young CY, Sun Z, Jimenez RE, Klee GG, Tindall DJ, Donkena KV. Global methylation profiling for risk prediction of prostate cancer. *Clin Cancer Res*. 2012 May 15; 18(10):2882–95. [PubMed: 22589488]
- Mani RS, Tomlins SA, Callahan K, Ghosh A, Nyati MK, Varambally S, Palanisamy N, Chinnaiyan AM. Induced chromosomal proximity and gene fusions in prostate cancer. *Science*. 2009 Nov 27; 326(5957):1230. [PubMed: 19933109]
- Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*. 2009 Sep 10; 361(11):1058–66. [PubMed: 19657110]
- Martin ST, Matsubayashi H, Rogers CD, Philips J, Couch FJ, Brune K, Yeo CJ, Kern SE, Hruban RH, Goggins M. Increased prevalence of the BRCA2 polymorphic stop codon K3326X among individuals with familial pancreatic cancer. *Oncogene*. 2005 May 19; 24(22):3652–6. [PubMed: 15806175]
- Mateo J, Hall E, Sandhu S, Omlin AG, Miranda S, Carreira S, Goodall J, Gillman A, Mossop H, Ralph C, et al. LBA20 - Antitumour activity of the PARP inhibitor olaparib in unselected sporadic castration-resistant prostate cancer (CRPC) in the TOPARP trial. *Annals of Oncology*. 2014; 25(5):1–41.
- Mostaghel EA, Geng L, Holcomb I, Coleman IM, Lucas J, True LD, Nelson PS. Variability in the androgen response of prostate epithelium to 5alpha-reductase inhibition: implications for prostate cancer chemoprevention. *Cancer Res*. 2010 Feb 15; 70(4):1286–95. [PubMed: 20124490]
- Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. 2010 May 18; 17(5):510–22. [PubMed: 20399149]
- Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, Suleman K, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med*. 2010 Jul; 16(7):793–8. [PubMed: 20526349]
- Parikh C, Janakiraman V, Wu WI, Foo CK, Kljavin NM, Chaudhuri S, Stawiski E, Lee B, Lin J, Li H, et al. Disruption of PH-kinase domain interactions leads to oncogenic activation of AKT in human cancers. *Proc Natl Acad Sci U S A*. 2012 Nov 20; 109(47):19368–73. [PubMed: 23134728]
- Paris PL, Andaya A, Fridlyand J, Jain AN, Weinberg V, Kowbel D, Brebner JH, Simko J, Watson JE, Volik S, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum Mol Genet*. 2004 Jul 1; 13(13):1303–13. [PubMed: 15138198]
- Penney KL, Stampfer MJ, Jahn JL, Sinnott JA, Flavin R, Rider JR, Finn S, Giovannucci E, Sesso HD, Loda M, et al. Gleason grade progression is uncommon. *Cancer Res*. 2013 Aug 15; 73(16):5163–8. [PubMed: 23946472]
- Pettersson A, Graff RE, Bauer SR, Pitt MJ, Lis RT, Stack EC, Martin NE, Kunz L, Penney KL, Ligon AH, et al. The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2012; 21(9):1497–509. [PubMed: 22736790]
- Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, MacDonald TY, et al. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. *Genome Res*. 2011 Jan; 21(1):56–67. [PubMed: 21036922]
- Prandi D, Baca SC, Romanel A, Barbieri CE, Mosquera JM, Fontugne J, Beltran H, Sboner A, Garraway LA, Rubin MA, Demichelis F. Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biol*. 2014 Aug 26; 15(8):439. [PubMed: 25160065]
- Pritchard CC, Morrissey C, Kumar A, Zhang X, Smith C, Coleman I, Salipante SJ, Milbank J, Yu M, Grady WM, et al. Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. *Nat Commun*. 2014 Sep 25; 5:4988. [PubMed: 25255306]

- Quon G, Haider S, Deshwar AG, Cui A, Boutros PC, Morris Q. Computational purification of individual tumor gene expression profiles leads to significant improvements in prognostic prediction. *Genome Med.* 2013 Mar 28.5(3):29. [PubMed: 23537167]
- Robinson R, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell.* In press.
- Rodrigues LU, Rider L, Nieto C, Romero L, Karimpour-Fard A, Loda M, Lucia MS, Wu M, Shi L, Cimic A, et al. Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. *Cancer Res.* 2015 Mar 15; 75(6):1021–34. [PubMed: 25770290]
- Rohle D, Popovici-Muller J, Palaskas N, Turcan S, Grommes C, Campos C, Tsoi J, Clark O, Oldrini B, Komisopoulou E, et al. An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science.* 2013 May 3; 340(6132):626–30. [PubMed: 23558169]
- Sboner A, Habegger L, Pflueger D, Terry S, Chen DZ, Rozowsky JS, Tewari AK, Kitabayashi N, Moss BJ, Chee MS, et al. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol.* 2010; 11(10):R104. [PubMed: 20964841]
- Schaeffer EM, Marchionni L, Huang Z, Simons B, Blackman A, Yu W, Parmigiani G, Berman DM. Androgen-induced programs for prostate epithelial growth and invasion arise in embryogenesis and are reactivated in cancer. *Oncogene.* 2008 Dec 4; 27(57):7180–91. [PubMed: 18794802]
- Schwartz S, Wongvipat J, Trigwell CB, Hancox U, Carver BS, Rodrik-Outmezguine V, Will M, Yellen P, de Stanchina E, Baselga J, et al. Feedback suppression of PI3K α signaling in PTEN-mutated tumors is relieved by selective inhibition of PI3K β . *Cancer Cell.* 2015 Jan 12; 27(1):109–22. [PubMed: 25544636]
- Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics.* 2009 Nov 15; 25(22):2906–12. [PubMed: 19759197]
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin.* 2015 Jan-Feb;65(1):5–29. [PubMed: 25559415]
- Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell.* 2010 Jul 13; 18(1):11–22. [PubMed: 20579941]
- Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, Zhang Z, Lapouble E, Grossetête-Lalami S, Rusch M, et al. Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* 2014 Nov; 4(11):1342–53. [PubMed: 25223734]
- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005; 310(5748):644–8. [PubMed: 16254181]
- Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature.* 2007 Aug 2; 448(7153):595–9. [PubMed: 17671502]
- Tomlins SA, Bjartell A, Chinnaiyan AM, Jenster G, Nam RK, Rubin MA, Schalken JA. ETS gene fusions in prostate cancer: from discovery to daily clinical practice. *Eur Urol.* 2009; 56(2):275–86. [PubMed: 19409690]
- Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, Zhao S, Haddad Z, Den RB, Dicker AP, Trock BJ, et al. Characterization of 1577 Primary Prostate Cancers Reveals Novel Biological and Clinicopathologic Insights into Molecular Subtypes. *Eur Urol.* 2015 May 8. pii: S0302-2838(15)00375-9.
- van Dekken H, Paris PL, Albertson DG, Alers JC, Andaya A, Kowbel D, van der Kwast TH, Pinkel D, Schröder FH, Vissers KJ, et al. Evaluation of genetic patterns in different tumor areas of intermediate-grade prostatic adenocarcinomas by high-resolution genomic array analysis. *Genes Chromosomes Cancer.* 2004 Mar; 39(3):249–56. [PubMed: 14732926]
- Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* 2010 Oct.38(18):e178. [PubMed: 20802226]

- Wang XS, Shankar S, Dhanasekaran SM, Ateeq B, Sasaki AT, Jing X, Robinson D, Cao Q, Prensner JR, Yocum AK, et al. Characterization of KRAS rearrangements in metastatic prostate cancer. *Cancer Discov.* 2011 Jun; 1(1):35–43. [PubMed: 22140652]
- Watson PA, Chen YF, Balbas MD, Wongvipat J, Socci ND, Viale A, Kim K, Sawyers CL. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proc Natl Acad Sci U S A.* 2010 Sep 28; 107(39): 16759–65. [PubMed: 20823238]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

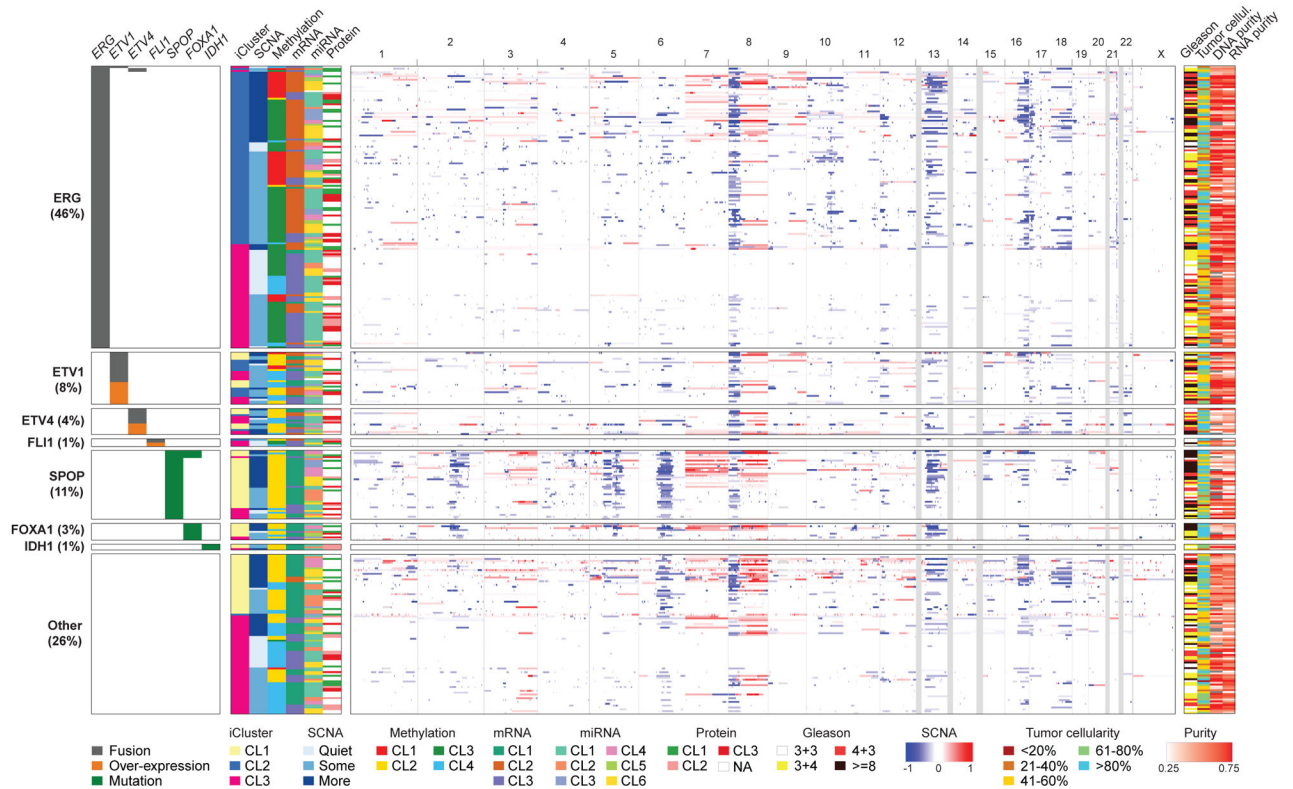


Figure 1. The molecular taxonomy of primary prostate cancer

Comprehensive molecular profiling of 333 primary prostate cancer samples revealed seven genomically distinct subtypes, defined (top to bottom) by ERG fusions (46%), ETV1/ETV4/FLI1 fusions or overexpression (8%, 4%, 1%, respectively), or by SPOP (11%), FOXA1 (3%), and IDH1 (1%) mutations. A subset of these subtypes was correlated with clusters computationally derived from the individual characterization platforms (somatic copy-number alterations, methylation, mRNA, microRNA, and protein levels from reverse phase protein arrays). The heatmap shows DNA copy-number for all cases, with chromosomes shown from left to right. Regions of loss are indicated by shades of blue, and gains by shades of red.

See also Figures S1, S2, S3, S4, S5, S6, S7, and Tables S1A, S1B, S1E, S2.

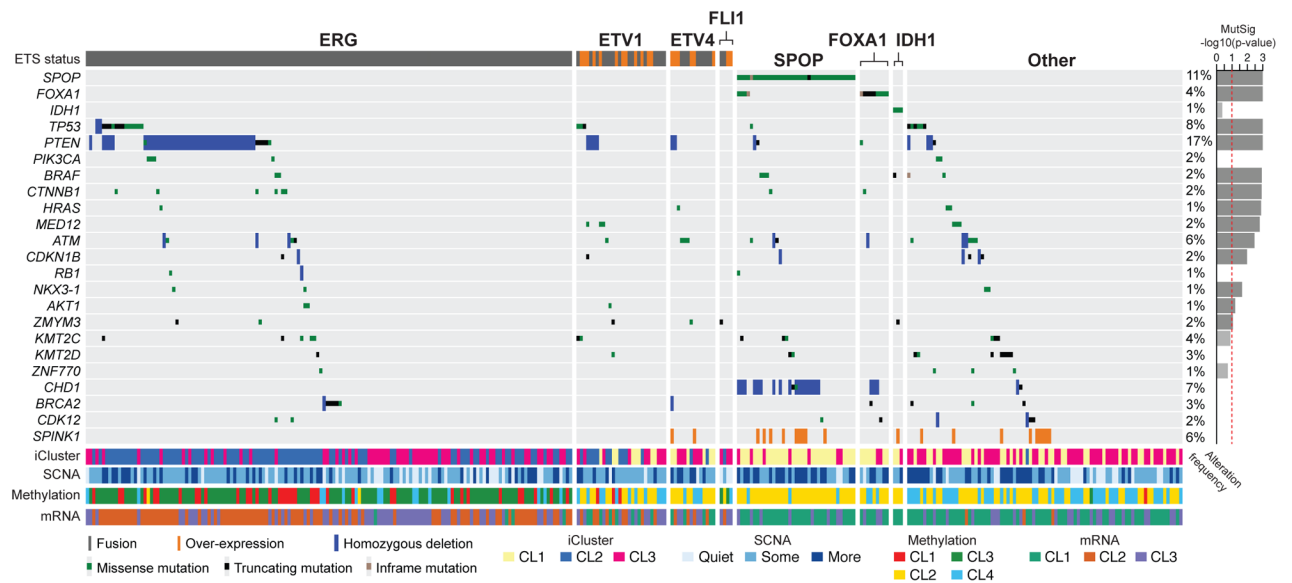


Figure 2. Recurrent alterations in primary prostate cancer

The spectrum and type of recurrent alterations and genes (mutations, fusions, deletions, and overexpression) in the cohort are shown (left to right) grouped by the molecular subtypes defined in Figure 1. On the right, the statistical significance of individual mutant genes (MutSig q-value) is shown. Mutations in *IDH1*, *PIK3CA*, *RB1*, *KMT2D*, *CHD1*, *BRCA2*, and *CDK12* are also shown, despite their not being statistically significant. *SPINK1* overexpression is shown for reference.

See also Tables S1B, S1C, S1D, S1E.

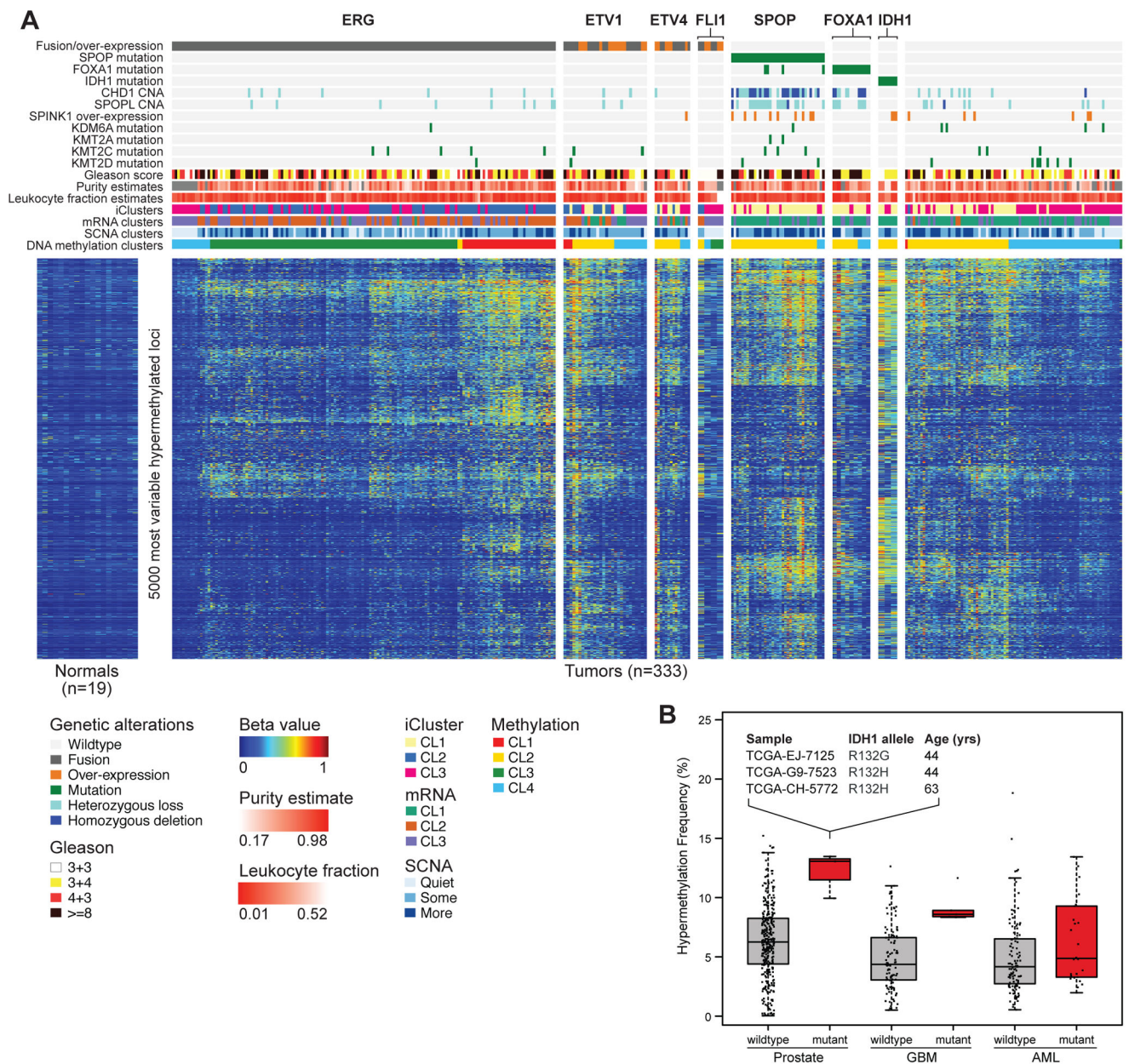


Figure 3. Hypermethylation is common across primary prostate cancer

A. Primary prostate cancers show diverse methylation changes compared to normal prostate samples (left). Unsupervised clustering was performed on the beta-values of the 5,000 most hypermethylated loci, and the results mapped to the genomic subtypes. *ERG*-positive tumors had a high diversity of methylation changes, with a distinct subgroup (cluster 1) nearly unique to this group. *SPOP* and *FOXA1* mutant tumors also exhibited global hypermethylation. **B.** *IDH1*-mutant prostate cancers, which are associated with younger age, are among the most hypermethylated tumors, as in Glioblastoma (GBM) and Acute Myeloid Leukemia (AML).

See also Figure S4 and Table S1F.

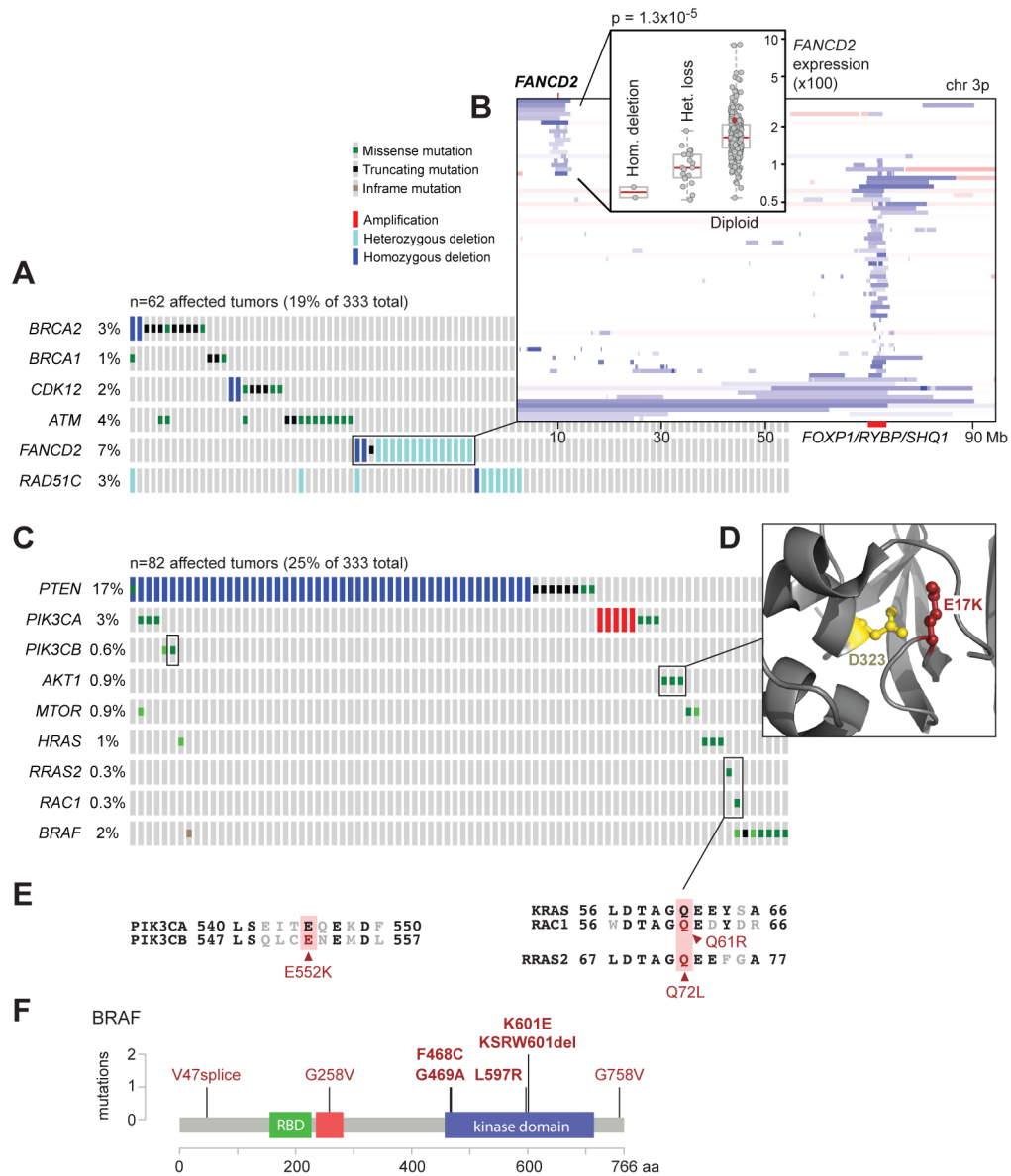


Figure 5. Alterations in clinically relevant pathways

A. Alterations in DNA repair genes were common in primary prostate cancer, affecting almost 20% of samples through mutations or deletions in *BRCA2*, *BRCA1*, *CDK12*, *ATM*, *FANCD2*, or *RAD51C*. **B.** Focal deletions of *FANCD2* were found in 7% of samples and were associated with reduced mRNA expression of *FANCD2*. **C.** The RAS or PI-3-Kinase pathways were altered in about a quarter of tumors, mostly through deletion or mutation of *PTEN*, but also through rare mutations in other pathway members. **D.** *AKT1* mutations were found in three samples. Two of them were the known activating E17K, and the third one affected the D323 residue, which is adjacent to E17 in the protein structure. **E.** One of the observed *PIK3CB* mutations, E552K, is paralogous to the known activating E545K mutation in *PIK3CA*, and the *RAC1* Q61 and *RRAS2* Q72 mutations are paralogous to the Q61

mutations in KRAS. **F.** BRAF mutations were found in 2% of samples, mostly in known non-V600E hotspots in the kinase domain.
See also Figure S3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

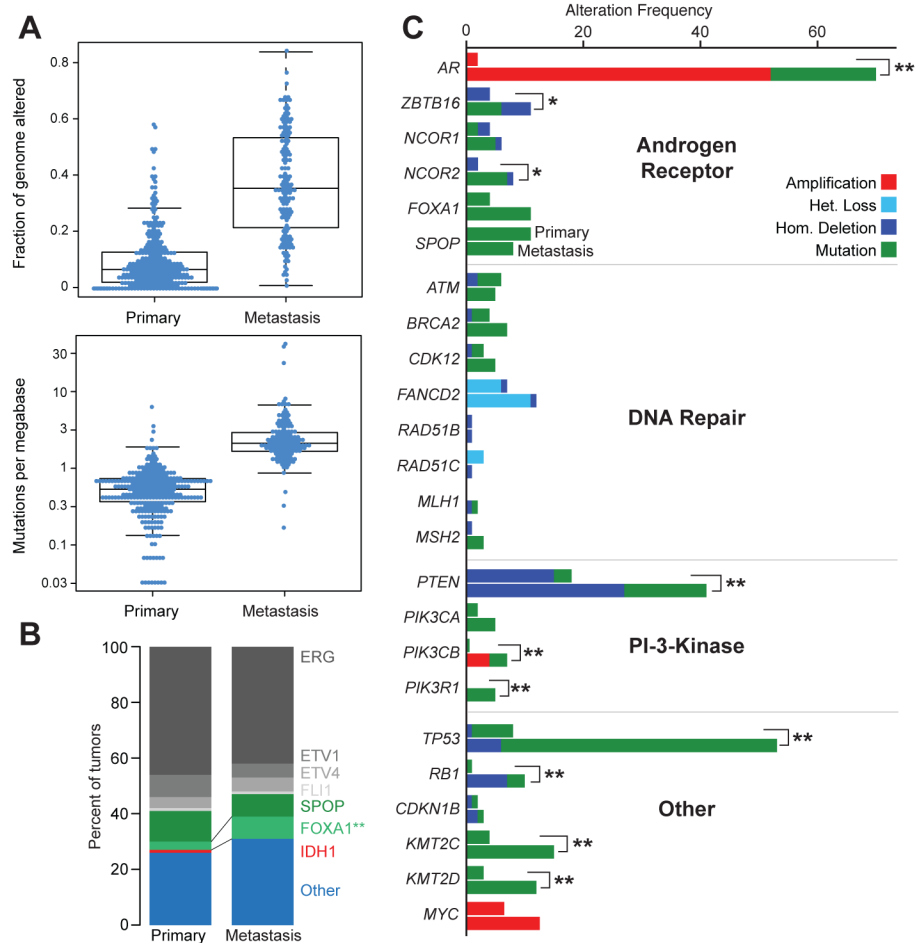


Figure 6. Comparison of primary with metastatic prostate cancer

A. Metastatic prostate cancer samples have more copy-number alterations (top panel, measured as fraction of genome altered) and mutations (bottom panel). **B.** The relative distribution of main subtypes (ERG, ETV1/4, FLI1, SPOP, FOXA1, IDH1, other) is similar in primary and metastatic samples. **C.** The alteration frequencies of several genes and pathways are higher in metastatic samples. The upper bar for each gene indicates the alteration frequency in primary samples, the lower bar for metastatic samples. The most notable differences in alteration frequencies involve the Androgen Receptor pathway, the PI3K pathway, and TP53.

See also Table S3.

Table 1

Cohort Characteristics

Clinical feature	
Age	61 (43–76)
Pre-operative PSA	7.4 (1.6–87.0)
Gleason Score	
3+3	65
3+4	102
4+3	78
8	88
Tumor Cellularity (pathology)	
<20%	7
21–40%	40
41–60%	84
61–80%	115
81–100%	87
Pathologic Stage	
pT2a/b	18
pT2c	111
pT3a	110
pT3b	82
pT4	6
PSA Recurrence	
Yes	33
No ^[a]	248
Not Available	47
Margin Status	
Positive	69
Negative	193
Not Available	71
Ethnicity	
Caucasian	270
African descent	43
Asian	8
Not Available	12

^[a]Either no evidence of recurrence or insufficient follow-up