Title

Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer

Journal

Cell, 174(3)

ISSN

0092-8674

Authors

Quigley, David A

Dang, Ha X

Zhao, Shuang G

et al.

Publication Date

2018-07-01

DOI

10.1016/j.cell.2018.06.039

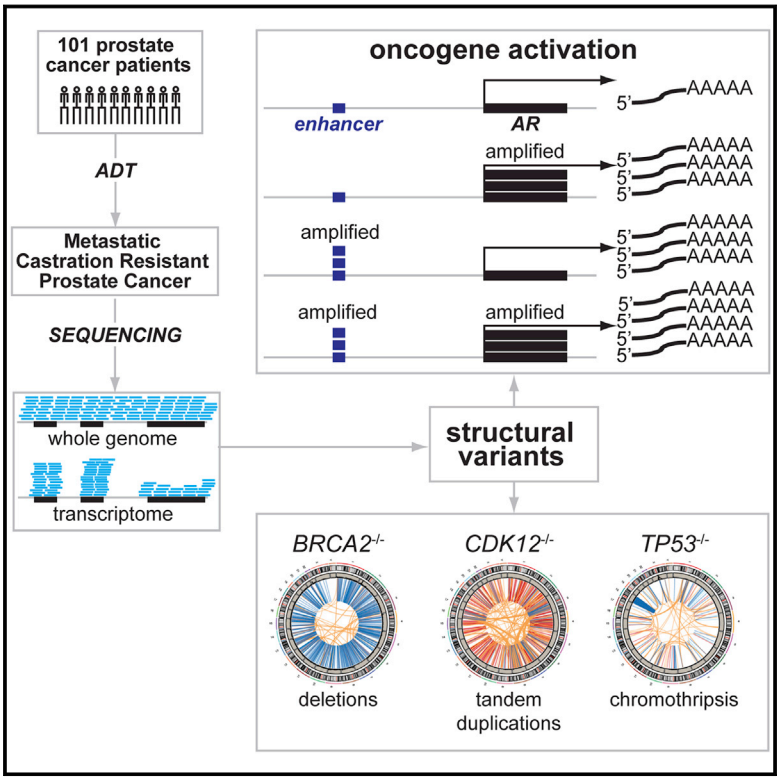
Peer reviewed

Resource

Cell

Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer

Graphical Abstract



Highlights

- Deep whole-genome and -transcriptome sequencing of 101 prostate cancer metastases
- Tandem duplication affects intergenic regulatory loci upstream of *AR* and *MYC*
- Inactivation of *CDK12*, *TP53*, and *BRCA2* affect distinct classes of structural variants
- Androgen receptor is affected by mutation or structural variation in 85% of mCRPC

Authors

David A. Quigley, Ha X. Dang, Shuang G. Zhao, ..., Christopher A. Maher, Eric J. Small, Felix Y. Feng

Correspondence

arul@med.umich.edu (A.M.C.), christophermaher@wustl.edu (C.A.M.), eric.small@ucsf.edu (E.J.S.), felix.feng@ucsf.edu (F.Y.F.)

In Brief

Integrative whole-genome and -transcriptome sequencing provides a comprehensive view of structural variations that affect major regulators in prostate cancer and would escape detection by exome-based approaches.

Quigley et al., 2018, Cell *174*, 1–12
July 26, 2018
https://doi.org/10.1016/j.cell.2018.06.039

CellPress

Please cite this article in press as: Quigley et al., Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer, Cell (2018), https://doi.org/10.1016/j.cell.2018.06.039

Resource

Cell

Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer

David A. Quigley,[1,2,38] Ha X. Dang,[3,4,38] Shuang G. Zhao,[5,38] Paul Lloyd,[6] Rahul Aggarwal,[6] Joshi J. Alumkal,[7,8] Adam Foye,[6] Vishal Kothari,[9] Marc D. Perry,[9] Adina M. Bailey,[6] Denise Playdle,[6] Travis J. Barnard,[9] Li Zhang,[6] Jin Zhang,[10,11] Jack F. Youngren,[6] Marcin P. Cieslik,[12,13] Abhijit Parolia,[12,13] Tomasz M. Beer,[7] George Thomas,[7,14] Kim N. Chi,[15,16] Martin Gleave,[15] Nathan A. Lack,[15] Amina Zoubeidi,[15] Robert E. Reiter,[17,18] Matthew B. Rettig,[17] Owen Witte,[19] Charles J. Ryan,[20] Lawrence Fong,[6] Won Kim,[6] Terence Friedlander,[6] Jonathan Chou,[6] Haolong Li,[9] Rajdeep Das,[9] Hui Li,[9] Ruhollah Moussavi-Baygi,[9] Hani Goodarzi,[21,22] Luke A. Gilbert,[1,22] Primo N. Lara, Jr.,[23,24] Christopher P. Evans,[24,25] Theodore C. Goldstein,[6,26] Joshua M. Stuart,[26] Scott A. Tomlins,[12,27] Daniel E. Spratt,[5] R. Keira Cheetham,[28] Donavan T. Cheng,[28] Kyle Farh,[28] Julian S. Gehring,[28] Jörg Hakenberg,[28] Arnold Liao,[28] Philip G. Febbo,[28] John Shon,[28] Brad Sickler,[28] Serafim Batzoglou,[28] Karen E. Knudsen,[29] Housheng H. He,[30] Jiaoti Huang,[31] Alexander W. Wyatt,[15] Scott M. Dehm,[32,33] Alan Ashworth,[1,6] Arul M. Chinnaiyan,[12,13,34,35,36,37,39,*] Christopher A. Maher,[3,4,39,*] Eric J. Small,[1,6,39,*] and Felix Y. Feng[1,6,9,22,39,40,*]

[1]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco (UCSF), San Francisco, CA, USA
[2]Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA, USA
[3]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO, USA
[4]Department of Internal Medicine, Washington University in St. Louis, St. Louis, MO, USA
[5]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA
[6]Division of Hematology and Oncology, Department of Medicine, UCSF, San Francisco, CA, USA
[7]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA
[8]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA
[9]Department of Radiation Oncology, UCSF, San Francisco, CA, USA
[10]Cancer Biology Division, Department of Radiation Oncology, Washington University in St. Louis, MO USA
[11]Institute for Informatics (I^2), Washington University in St. Louis, MO
[12]Department of Pathology, University of Michigan, Ann Arbor, MI, USA
[13]Michigan Center for Translational Pathology, Ann Arbor, MI, USA
[14]Department of Pathology, Oregon Health and Science University, Portland, OR, USA
[15]Vancouver Prostate Centre, Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada
[16]British Columbia Cancer Agency, Vancouver Centre, Vancouver, BC, Canada
[17]Jonsson Comprehensive Cancer Center, Department of Urology, UCLA, Los Angeles, CA, USA
[18]VA Greater Los Angeles Healthcare System, Department of Medicine, Los Angeles, CA, USA
[19]Department of Microbiology, Immunology, and Molecular Genetics at the David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[20]Division of Hematology, Oncology, and Transplant, Department of Medicine, University of Minnesota, Minneapolis, MN, USA
[21]Department of Biophysics and Biochemistry, UCSF, San Francisco, CA, USA
[22]Department of Urology, UCSF, San Francisco, CA, USA
[23]Division of Hematology Oncology, Department of Internal Medicine, University of California Davis, Sacramento, CA, USA
[24]Comprehensive Cancer Center, University of California Davis, Sacramento, CA, USA
[25]Department of Urologic Surgery, University of California Davis, Sacramento, CA, USA
[26]UC Sant Cruz Genome Institute and Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA
[27]Department of Urology, University of Michigan, Ann Arbor, MI, USA
[28]Illumina, Inc., San Diego, CA, USA
[29]Department of Cancer Biology, Thomas Jefferson University, Philadelphia, PA, USA
[30]Princess Margaret Cancer Centre/University Health Network, Toronto, ON, Canada
[31]Department of Pathology, Duke University, Durham, NC, USA
[32]Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA
[33]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA
[34]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[35]Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI, USA
[36]Department of Urology, University of Michigan, Ann Arbor, MI, USA
[37]Rogel Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA
[38]These authors contributed equally
[39]Senior author
[40]Lead Contact
*Correspondence: arul@med.umich.edu (A.M.C.), christophermaher@wustl.edu (C.A.M.), eric.small@ucsf.edu (E.J.S.), felix.feng@ucsf.edu (F.Y.F.)
https://doi.org/10.1016/j.cell.2018.06.039

Please cite this article in press as: Quigley et al., Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer, Cell (2018), https://doi.org/10.1016/j.cell.2018.06.039

SUMMARY

While mutations affecting protein-coding regions have been examined across many cancers, structural variants at the genome-wide level are still poorly defined. Through integrative deep whole-genome and -transcriptome analysis of 101 castration-resistant prostate cancer metastases (109X tumor/38X normal coverage), we identified structural variants altering critical regulators of tumorigenesis and progression not detectable by exome approaches. Notably, we observed amplification of an intergenic enhancer region 624 kb upstream of the androgen receptor (*AR*) in 81% of patients, correlating with increased *AR* expression. Tandem duplication hotspots also occur near *MYC*, in lncRNAs associated with post-translational *MYC* regulation. Classes of structural variations were linked to distinct DNA repair deficiencies, suggesting their etiology, including associations of *CDK12* mutation with tandem duplications, *TP53* inactivation with inverted rearrangements and chromothripsis, and *BRCA2* inactivation with deletions. Together, these observations provide a comprehensive view of how structural variations affect critical regulators in metastatic prostate cancer.

INTRODUCTION

Prostate cancer represents a common and clinically heterogeneous disease entity. While over 160,000 American men are diagnosed with prostate cancer each year, <20% of patients will experience progression to the lethal form of the disease, termed metastatic castration-resistant prostate cancer (mCRPC) (Siegel et al., 2018). A major barrier to studying mCRPC has been the difficulty in obtaining tumor samples, as clinical biopsies of metastatic lesions are not routinely performed. mCRPC has recently been evaluated by targeted or whole-exome sequencing (Armenia et al., 2018; Beltran et al., 2013; Grasso et al., 2012; Robinson et al., 2015; Zehir et al., 2017). These studies identified alterations in pathways involving androgen signaling, DNA repair, and phosphoinositide 3-kinase (PI3K) signaling, as well as recurrent mutations in genes such as *SPOP*, *FOXA1*, and *IDH1*. However, the exome represents <2% of the genome, and outside of small case series (Gundem et al., 2015; Wedge et al., 2018), the complete genomic landscape of mCRPC remains largely unexplored.

Genomic structural variants (SVs) include genomic deletions, insertions, tandem duplications, inversion rearrangements, and inter-chromosomal translocations. SVs are prevalent in prostate cancer, with gene fusions involving the E26 transformation-specific (ETS) family of transcription factors identified in 40%–60% of cases (Maher et al., 2009; Tomlins et al., 2005, 2007). A recent study in localized prostate cancer demonstrated clusters of genomic rearrangements each occurring in 5%–6% of samples (Fraser et al., 2017). In addition, previous studies have demonstrated that SVs may define subtypes of ovarian, pancreatic, and breast cancers (Nik-Zainal et al., 2016; Patch et al., 2015; Waddell et al., 2015; Wang et al., 2017). Of note, the majority of SVs involve intergenic or intronic noncoding regions of the genome and are not captured by exome sequencing or transcriptome analysis. A key advantage of whole-genome sequencing (WGS) over exome sequencing is that WGS allows the identification of SVs that alter the activity of key driver genes, tumor suppressors, and regulatory elements.

To comprehensively investigate the genomic drivers of mCRPC, we interrogated the whole genomes and transcriptomes of mCRPC samples from over 100 patients at a mean depth of 109X in tumors, a depth 2–3 times greater than that achieved in previous large WGS studies in cancer. Deep sequencing of a large patient cohort permitted us to discover novel recurrent SVs and define the prevalence of these variations in mCRPC. We discovered previously unidentified recurrent SVs modulating tumor suppressors or oncogenes, identified new rearrangements coupling noncoding genes to known cancer drivers, and uncovered novel global associations between DNA repair alterations and SVs.

RESULTS

A multi-institutional consortium conducted a prospective IRB-approved study (NCT02432001) that obtained and profiled metastatic tumor biopsies from prostate cancer patients with castration-resistant disease (Aggarwal et al., 2016). Image-guided core biopsies were obtained (Holmes et al., 2017) and fresh-frozen. Tumor tissue was centrally processed and banked. Laser capture microdissection was used to isolate samples enriched for cancer, and sequencing of RNA was performed. Whole-genome DNA sequencing was performed from frozen sections for tumor and from peripheral blood for matched normal samples, obtaining a mean depth of 109X in tumor and 38X in normal samples (Figure S1). Paired end mRNA libraries were sequenced to a median depth of 114 M paired reads. This report includes results from 101 patients, including mCRPC lesions from bone (n = 42), lymph node (n = 40), liver (n = 11), or other soft tissue sites (n = 8) (clinical summary in Table 1, sample-level features related to sequencing, molecular analysis, and biopsy site in Table S1). Of these patients, 64% had received second-generation anti-androgen therapy (abiraterone: 47%, enzalutamide: 37%, both: 20%).

Structural Variations Disrupt Key Driver Genes

The frequency of genomic copy number alterations in our mCRPC tumors was consistent with previous exome sequencing reports (Armenia et al., 2018; Beltran et al., 2013; Grasso et al., 2012; Robinson et al., 2015) (Figures 1A and S2A). The percent of the genome altered in each sample ranged between 7% and 47% (median 23%; Table S1). The median mutation frequency was 4.1 mutations/Mb, slightly lower than the 4.4 mutations/Mb reported previously in mCRPC (Robinson et al., 2015), but greater than the 0.53 mutations/Mb reported in primary prostate cancer (Fraser et al., 2017). Approximately 40% of tumors were triploid (Figure S2B; Table S1). Triploid status was associated

Table 1. Clinical Characteristics of the Patient Cohort

Category	Value
Median age (range)	71 (45–90)
Race/ethnicity	
White	85
Black	5
Asian	4
Unknown	7
Gleason grade at diagnosis	
6	11
7	28
≥8	52
Unknown	10
Site of biopsy	
Bone	43
Lymph node	39
Liver	11
Other soft tissue	8
Prior therapy	
Abiraterone	27
Enzalutamide	17
Both	20
Neither	37
Visceral metastases	
Yes	31
No	70
Median lab values at biopsy	
PSA, ng/mL (range)	65.1 (0.4–1874.5)
Alkaline phosphatase, IU/L (range)	92 (49–1506)
Lactate dehydrogenase, IU/L (range)	187 (31–856)
Hemoglobin, g/dL (range)	12.8 (8.0–15.7)

with more translocations and mutations overall ($p < 0.007$, Figures S2C and S2D).

We systematically identified loci most frequently affected by structural variations by counting SVs within 1 Mb windows genome-wide (SV per window 9.6 ± 5.1; mean ± SD, listed in Table S2). The frequency of SVs is plotted in concert with copy number alteration frequencies in Figure 1A. The loci most frequently affected by SV (>3 SD from mean) contained key drivers of prostate cancer, underscoring the importance of structural variation in this disease. This included *AR*, the transmembrane serine protease 2 (*TMPRSS2*), and ETS transcription factor (*ERG*) genes that produce the TMPRSS2/ERG fusion protein, the oncogene *MYC*, Forkhead Box protein A1 (*FOXA1*), and phosphatase and tensin homolog (*PTEN*). This analysis also identified clusters of deletions affecting genes at fragile sites previously identified in more than one cancer type (Bignell et al., 2010; Glover et al., 2017).

An integrated analysis of SVs and mRNA expression levels was then used to define cases where SVs were predicted to inactivate tumor suppressors. *PTEN* was affected by biallelic alterations in 36% of tumors and monoallelic alterations in 26% of tumors (Figure 1B). The *PTEN* sequence or promoter was frequently interrupted by a translocation (7% of cases) or inverted rearrangement (5% of cases, Figure S3A). SV were essential to assigning biallelic *PTEN* alteration status in 8% of cases and mono-allelic *PTEN* alteration status in 5% of cases (Figure 1B, left). *TP53* was affected by biallelic somatic alterations in 46% of tumors and monoallelic alterations in 30% of tumors, with 11% of the biallelic assignments due to SV gene disruption. SVs also contributed to biallelic inactivation of *RB1* (12% biallelic, 3% by SV), *CDKN1B* (7% biallelic, 1% by SV), and *CHD1* (7% biallelic, 2% by SV) (Figure 1B). There was a significant association between the number of inactivated alleles and mRNA levels of *PTEN*, *TP53*, *CDKN1B*, *RB1*, and *CHD1* (Figure 1B, right), suggesting monoallelic alterations impacted expression levels of these genes.

Novel Gene Fusions Predicted to Activate Oncogenes

We then determined cases where structural variants were predicted to activate driver genes by integrating SV data, mRNA expression levels, and predicted mRNA fusions. A majority of prostate cancers harbor fusions from the juxtaposition of the 5′ regulatory region of the androgen-responsive gene *TMPRSS2* upstream of *ERG* (Tomlins et al., 2005). We observed mutually exclusive fusions activating the ETS family members *ERG*, *ETV1*, *ETV4*, and *ETV5* in 59% of our cohort (Figure 1C; fusions listed in Table S3). In four cases, an ETS family member fused to a gene not previously reported in mCRPC, including *ETV1* fusions driven by the solute carrier *SLC30A4* and *ETV4* fusions driven by transmembrane and coiled-coil domain family 2 (*TMCC2*), clathrin heavy chain (*CLTC*), and cell division cycle 6 (*CDC6*). We also identified novel fusions between coding and non-coding genes, exemplified by *SCHLAP1*, a lncRNA highly enriched in a subset of aggressive prostate cancers (Prensner et al., 2013). The PI3K pathway member *PIK3CA* was expressed at very low levels except for a single sample bearing a translocation that placed the first exon of *SCHLAP1* immediately downstream of the *PIK3CA* 5′ UTR, resulting in the overexpression of a full-length *PIK3CA* transcript (Figure S3B). In two other cases, *ETV1* was translocated to chromosome 14 between *FOXA1* and mirror-image polydactyly 1 (*MIPOL1*). The lncRNA *RP11-356O9.1* (also annotated as *AL121790.1*) lies in this region. Previously published data showed that in normal tissues, *RP11-356O9.1* is expressed exclusively in prostate (Figure S4). In these two cases, the first exon of *RP11-356O9.1* was fused to exon 4 or exon 5 of *ETV1* (Figure 1D). Fusions between *ETV1* and this region have been previously reported in the prostate cancer cell line MDA-PCa 2B (Tomlins et al., 2007) and in a single patient sample (Abeshouse et al., 2015).

Multiple low-frequency gene fusions involving oncogenes, including *AXL*, *BRAF*, and *MYC*, were also noted (Figures 1E and S3B). A gene fusion joined prostatic acid phosphatase (*ACPP*) residue 380 (NM_001009) to the transmembrane receptor tyrosine kinase *AXL* at residue 429 (NM_001699), producing an in-frame transcript. Review of an independent cohort of patients with high risk primary prostate cancer identified a similar *ACPP-AXL* fusion, demonstrating these fusions are a

Please cite this article in press as: Quigley et al., Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer, Cell (2018), https://doi.org/10.1016/j.cell.2018.06.039

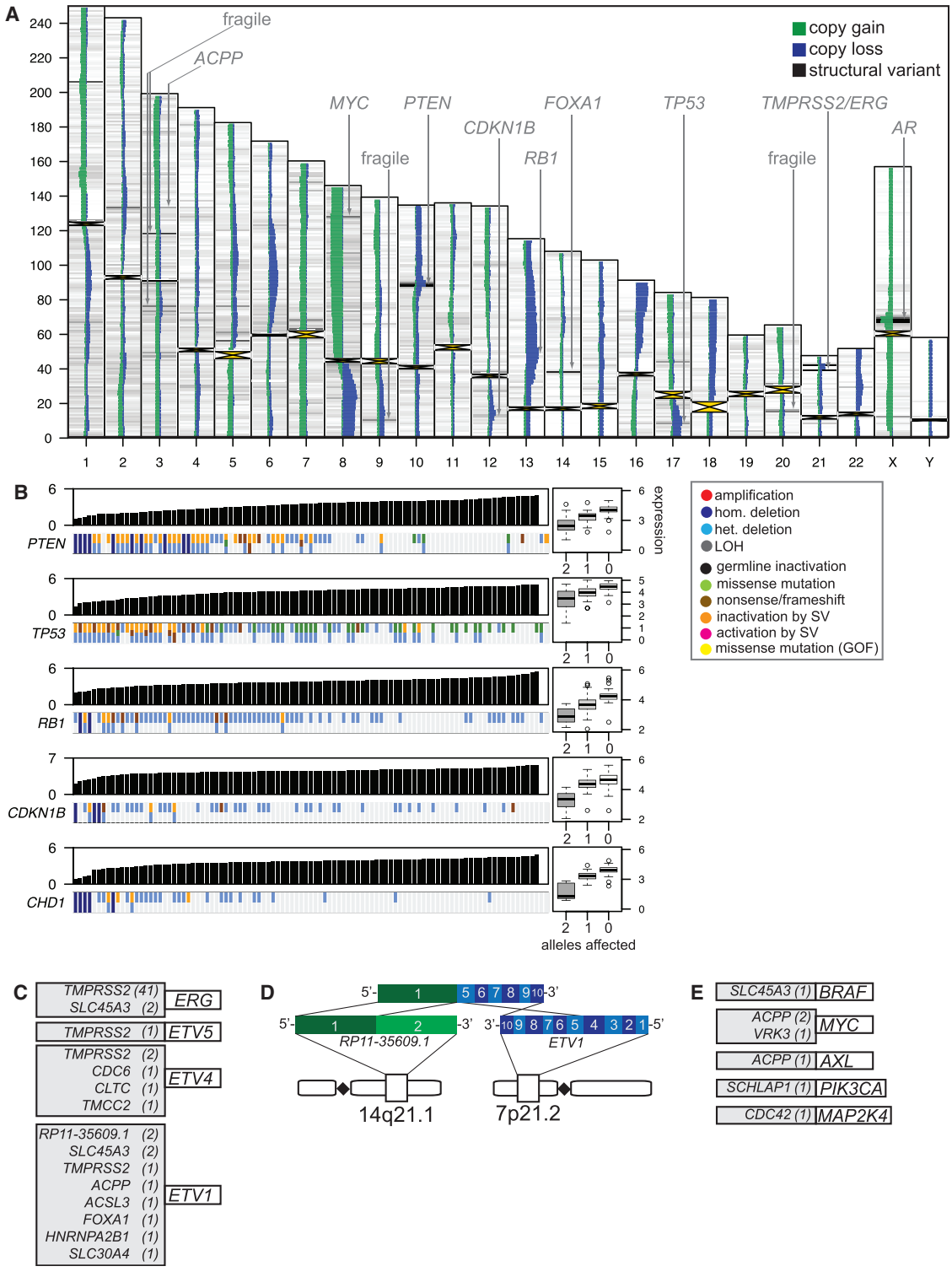


Figure 1. Structural Variants Disrupt Tumor Suppressors and Activate Oncogenes

(A) SV and copy number frequency plotted on scaled chromosomes. Wider green/blue bars indicate more frequent copy gain/loss. Darker black bars indicate more frequent SV.

(B) Top: expression levels of *PTEN*, *TP53*, *RB1*, *CDKN1B*, and *CHD1* in individual samples reported as (log[1+(TPM × 10^6)]). Bottom: somatic events affecting each sample. Right: box and whisker plots showing expression for samples with 0, 1, or 2 alleles affected; horizontal bar indicates median. Each gene was sorted independently by expression level. See also Figure S3.

(legend continued on next page)

repeated finding (Figure S3C). In a case lacking high level *MYC* DNA copy number amplification, *ACPP* was fused to *MYC* within 150 nt of the *MYC* 5′ untranslated region, originating within the second and third *ACPP* exons. Collectively, these novel, low frequency gene fusions could represent therapeutic targets in mCRPC.

Duplications Target *AR*, *MYC*, and *FOXA1*

Genomic duplication events are a mechanism of genome evolution (Ohno, 1970) and are known to alter specific drivers important in cancer, such as *FLT3* in acute myeloid leukemia and *BRAF* in pilocytic astrocytoma (Jones et al., 2008; Nakao et al., 1996). Unbiased analysis identified a region ~624 kb upstream of *AR* as the most frequent site of structural variation in mCRPC (Table S2). *AR* amplification occurred in 70% of cases and was associated with significantly elevated *AR* mRNA expression ($p = 9 \times 10^{-8}$, Figures 2A, top, and 2B). Our result is consistent with earlier findings that *AR* amplification is rare in primary prostate cancer (Abeshouse et al., 2015) but common in mCRPC (Robinson et al., 2015) and is a major mechanism of resistance to androgen deprivation therapy (Visakorpi et al., 1995). The region of peak amplification upstream of *AR* at 66.94 Mb was amplified in 81% of cases, 11% more frequently than *AR* itself (Figure 2A, middle). Tumors frequently amplified both *AR* and the upstream peak (68 cases), but in 13 cases the upstream peak alone was amplified (Figure 2B). DNA copy gain at the upstream peak in cases that lacked *AR* amplification was significantly associated with elevated *AR* expression ($p = 0.003$, Figure 2B), indicating that amplification of the upstream peak was independently associated with *AR* expression levels. Cases with amplification of both the upstream peak and *AR* had significantly higher expression than cases where only the upstream peak was amplified ($p = 0.01$, Figure 2B), consistent with additive effects.

Tandem duplications at the upstream peak corresponding to copy number gain break points were observed in 36% of all cases and in 44% of the 81 cases bearing copy gain at this region. Focal tandem duplication of the upstream peak region was almost exclusive to patients lacking or with low *AR* amplification ($p < 0.0007$, hypergeometric test, Figure 2C), consistent with tandem duplication at the peak being a sufficient alternative to *AR* amplification. The presence of amplification at this peak was not associated with previous treatment with the second-line hormone therapies abiraterone or enzalutamide (Table S4). We assessed the frequency of H3K27ac occupancy within the upstream peak, as H2K27ac enrichment is associated with potential enhancer activity (Heintzman et al., 2009). Previously published data from 19 primary prostate tumors revealed that the minimally targeted region at the upstream peak was enriched for H3K27ac histone modifications (Kron et al., 2017) (Figure 2A, middle and bottom). Collectively, these data support the detection of an enhancer, amplified in 81% of castration-resistant metastatic patients, that can act independently of *AR* locus amplification to increase expression of *AR* in response to first-line ADT.

Intergenic regions near *MYC* at 8q24 and *FOXA1* at 14q13.3 were also frequent targets of SVs (Table S2). We observed distinct tandem duplication peaks 700 and 300 kb. upstream of *MYC*, with duplication frequencies of 25% and 23%, respectively (Figure 2D, top, middle). The farther region included three long non-coding RNAs: prostate cancer associated transcript -1 and -2 (*PCAT-1*, *PCAT-2*), and prostate cancer associated non-coding RNA 1 (*PRNCR1*). The degree of *MYC* copy number amplification was modestly associated with *MYC* mRNA expression levels (rho = 0.28, $p = 0.005$). Although *PRNCR1* is unlikely to be implicated in mCRPC pathogenesis (Prensner et al., 2014), *PCAT-1* has been shown to upregulate cMyc protein levels post-translationally (Prensner et al., 2014b). The nearer region included additional non-coding genes, as well as the rs6983267 and rs1447295 germline variants associated with prostate cancer risk (Amundadottir et al., 2006; Yeager et al., 2007). Tandem duplications overlapping *FOXA1* and/or the adjacent gene mirror-image polydactyl 1 (*MIPOL1*) were present in 14% of samples (Figure 2E). These events were less frequent than *AR* or *MYC* events described above, precluding nomination of a candidate local peak, but several sites in this region had H3K27ac enrichment (Figure 2E). Three of the 14 samples bearing tandem duplications in this region also bore *FOXA1* mutations. Observations of *ETV1* translocations into this region by us (Figure 1D) and others (Abeshouse et al., 2015; Tomlins et al., 2007) suggest that SVs at this locus play a role in prostate cancer. These observations collectively demonstrate that unbiased analysis of tandem duplications by whole-genome sequencing identifies loci that are selected for amplification near driver genes such *AR*, *MYC*, and *FOXA1* in metastatic prostate cancer, and this selection potentially drives disease progression.

DNA Repair Defects Drive SVs

To explore the etiology of SVs in prostate cancer, we identified alterations associated with SV frequency. The number of SVs identified in individual tumors ranged between 103 and 923 (337 ± 166, mean ± SD, Figure 3A). Deletion frequency was significantly higher in tumors with biallelic *BRCA2* mutations ($p = 4 \times 10^{-6}$) (Figures 3B, left, and 3C). Additionally, we observed that biallelic inactivation of *CDK12* was associated with a significant increase in tandem duplications with a bimodal length distribution (Figures 3B center, 3C, and S5) ($p = 0.003$). These results were consistent with results previously reported in ovarian cancer (Popova et al., 2016).

We noted that the number of inverted rearrangements and deletions observed in each sample was significantly correlated (rho = 0.54, $p < 4 \times 10^{-10}$, Figure 3D). Tumors bearing large numbers of both deletions and inverted rearrangements had

(C) Schematic diagrams of ETS family fusions indicating previously observed and novel partners. See also Figure S3.
(D) Schematic diagram of *ETV1* activation via RP11-35609.1 fusions. See also Figure S4.
(E) Schematic diagrams of oncogene fusions showing previously observed and novel partners.
See also Figure S2 and Tables S1, S2, S3, and S5.

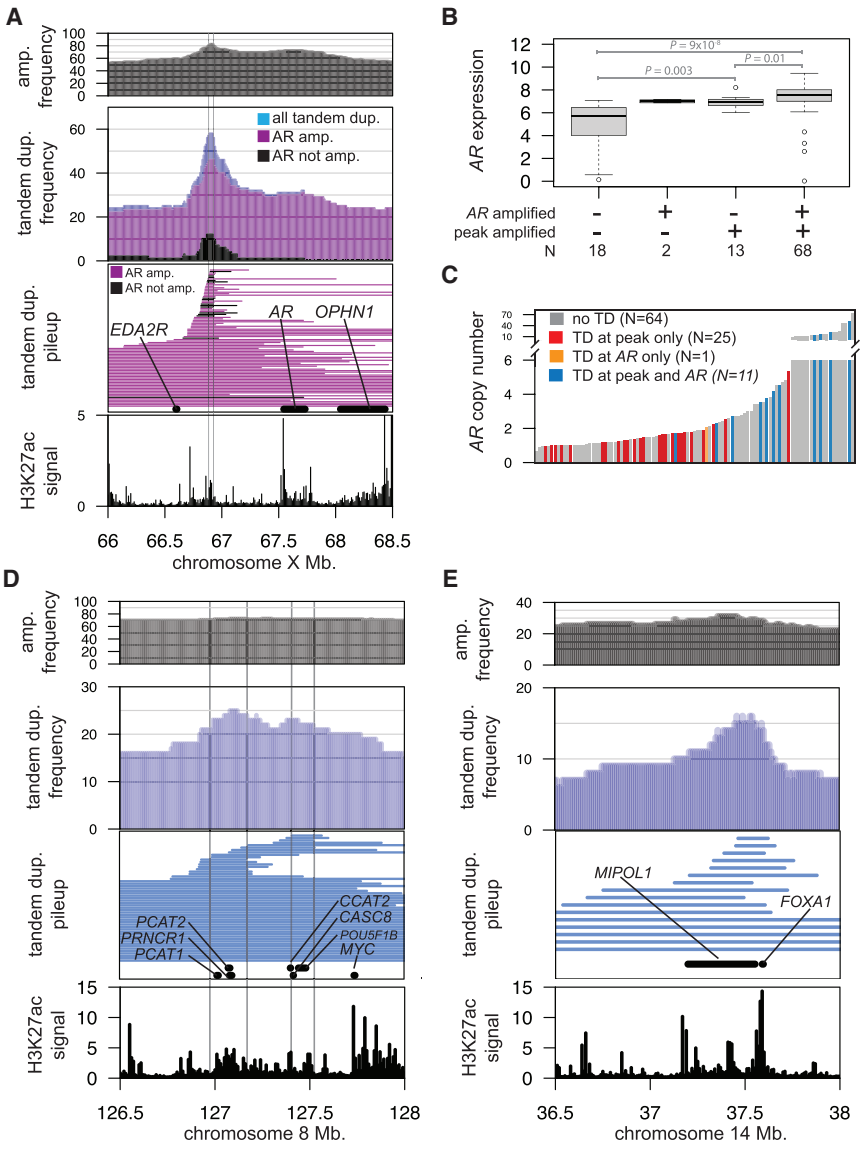


Figure 2. Tandem Duplication Target Enhancers near *AR*, *MYC*, and *FOXA1*

(A) Aligned tracks showing the DNA amplification frequency (top), tandem duplication frequency (middle), tandem duplication bounds (middle), and H3K27ac average read coverage (bottom, from Kron et al., 2017) at the *AR* locus.

(B) Box and whisker plot showing *AR* expression in the presence/absence of DNA amplification at *AR* or at the peak.

(C) Samples with tandem duplication of the peak in (A) but not *AR* (red) more frequently had *AR* unamplified or amplified at low levels.

(D and E) aligned tracks showing tandem duplications near *MYC* (D) and *FOXA1* (E) as in (A).

See also Table S4.

all undergone chromothripsis, the shattering and subsequent reconstruction of a single chromosome (Figure 3C, right; Figure 3D, orange points) (Fraser et al., 2017; Maher and Wilson, 2012; Stephens et al., 2011; Zack et al., 2013) (Figure 3B, right). We identified chromothripsis in 23% of mCRPC (Figure 3A, 3D; samples listed in Table S1), compared with 20% reported in non-indolent primary prostate tumors (Fraser et al., 2017). Biallelic *TP53* inactivation was the event most significantly associated with elevated inverted rearrangement frequency (median 57 versus 79 inversion rearrangements, p = 0.0004) and with the presence of chromothripsis (19 of the 23 cases with chromothripsis versus 28 of the 78 cases lacking chromothripsis, p = 0.0004). No locus was preferentially targeted by chromothripsis, consistent with a stochastic process. No tumor with biallelic loss of *BRCA2* also exhibited chromothripsis (Figures 3A and 3D). As observed in a previous pan-cancer analysis, chromothripsis was not associated with an elevated mutation frequency genome-wide (p > 0.05, Figure 3E) (Zack et al., 2013). In contrast, *BRCA2* loss had the strongest statistical association with tumor mutational burden (median 7.0 versus 4.0 mutations/Mb, p = 0.0002, Figure 3E).

Chromoplexy, a balanced interweaving of interchromosomal translocations, has been observed in prostate cancer (Baca et al., 2013). We identified chromoplexy in 50% of samples (Table S1). Of the 23 samples with chromothripsis, 12 (52%) also showed chromoplexy, as expected if there were neither positive nor negative enrichment for chromothripsis in samples that had undergone chromoplexy. The presence of somatic *TP53* alterations was not associated with either translocation frequency or with the presence of chromoplexy. Our analysis therefore identified biallelic inactivation of *CDK12*, *BRCA2*, and *TP53* as strongly linked to three forms of SV in mCRPC, with the link between *TP53* inactivation and inversion rearrangements further linked to chromothripsis.

Mutational Signatures of DNA Damage

Cells bearing homologous recombination repair defects develop genomic scars (reviewed in Lord and Ashworth, 2016), including deletions with homology at both ends of the deleted region. These cells rely on microhomology-mediated end joining to repair double strand DNA breaks, also known as alternative nonhomologous end-joining (Davies et al., 2017; Nik-Zainal et al., 2012, 2016; Tutt et al., 2001). Tumors bearing biallelic loss of *BRCA2* had elevated levels of deletions with flanking microhomology (Figure 4A). Tumors with biallelic inactivation of *CDK12* or *ATM*, or with monoallelic alterations in *BRCA1* or *BRCA2,* lacked this phenotype, confirming previously published observations (Polak et al., 2017). We fitted published mutation signature profiles to somatic single nucleotide variations and

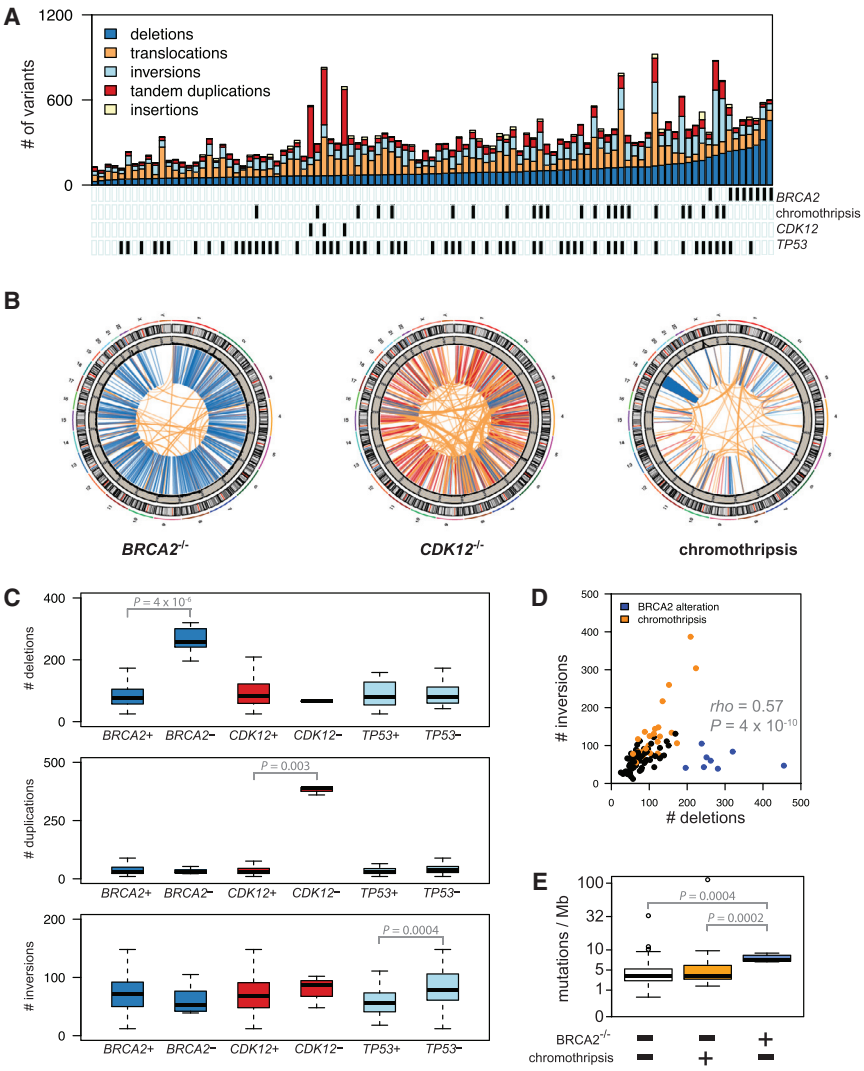


Figure 3. DNA Repair Alterations Are Associated with Structural Variation Frequency

(A) Top: structural variant frequency by sample, sorted by deletion frequency. Bottom: presence of chromothripsis or biallelic inactivating alterations in *BRCA2*, *CDK12*, or *TP53*.

(B) Circos plots illustrating *BRCA2* inactivation (left), *CDK12* inactivation (center), and chromothripsis (right). Colors as in (A).

(C) Box and whiskers plots showing association between biallelic inactivating alterations in *BRCA2*, *CDK12*, or *TP53* and the frequencies of deletions, tandem duplications, and inverted rearrangements respectively. See also Figure S5.

(D) Counts of inverted rearrangements and deletions per sample. Samples with biallelic *BRCA2* loss drawn in blue, samples bearing chromothripsis drawn in orange.

(E) Box and whisker plots showing mutation frequency in the presence of biallelic loss of *BRCA2* and chromothripsis.

See also Table S1.

performed *de novo* mutational profile signature analysis using non-negative matrix factorization (Alexandrov et al., 2013). A solution including eight *de novo* signatures provided the optimal balance between variance explained and parsimonious modeling. Signature *de novo 8* was strongly associated with samples bearing biallelic *BRCA2* inactivation (Figures 4A and S6A) and closely resembled COSMIC signatures 3 and 8 (Figures 4A and S6B), previously associated with defects in homologous recombination DNA repair (HRD) (Alexandrov et al., 2013; Nik-Zainal et al., 2016). COSMIC 3 signature fit was significantly elevated in samples bearing biallelic loss of *BRCA2,* consistent with previous reports in breast, ovarian, and prostate cancer ($p = 4 \times 10^{-7}$, Figures 4A and 4B).

A sample with heterozygous mutation of both *BRCA1* and *BRCA2* lacked an elevated microhomology deletion frequency, but nevertheless showed strong *de novo* 8 and COSMIC 3 signature scores. In all, 6% of cases harbored compound *BRCA1/2* heterozygosity, either by single copy DNA loss (n = 5) or somatic mutation (n = 1). These samples had significantly elevated mutation frequency, statistically indistinguishable from that observed in $BRCA2^{-/-}$ samples (Figure 4B). Compound heterozygous samples had COSMIC 3 signature scores intermediate between cases with biallelic *BRCA2* inactivation and cases with one or zero *BRCA1/2* alleles affected (Figure 4C), but the difference in signature fit was not statistically significant.

The other robust *de novo* signatures identified in this cohort recapitulated known signatures (Alexandrov et al., 2013). These included *de novo 1*, likely identical to COSMIC signature 1 associated with spontaneous deamination of 5-methylcytosine associated with age at tumor diagnosis (Alexandrov et al., 2015), and de *novo 5*, present in a hypermutated sample with deep deletion of mutS homolog 2 (*MSH2*) and *MSH6*, mismatch repair genes 300 kb apart on chromosome 2. This signature bore the strongest similarity to COSMIC 6 (associated with defective MMR) and COSMIC 9 (activation-induced deaminase activity during hypermutation). These data confirmed that DNA repair defects in mismatch repair and homologous recombination can produce genomic scars in metastatic prostate cancer and showed that *BRCA1/2* compound heterozygosity produces a mutational phenotype distinct from that of biallelic *BRCA2* inactivation.

A Landscape of Mutations and Structural Alterations

Somatic alterations and structural variants for 44 key prostate cancer genes are shown in Figure 5, and listed in Table S5. The somatic mutation frequencies were consistent with previous reports in mCRPC (Armenia et al., 2018; Robinson et al., 2015). In total, 85% of mCRPC samples carried either pathogenic activating *AR* mutations, amplifications of *AR*, or putative *AR*

Please cite this article in press as: Quigley et al., Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer, Cell (2018), https://doi.org/10.1016/j.cell.2018.06.039

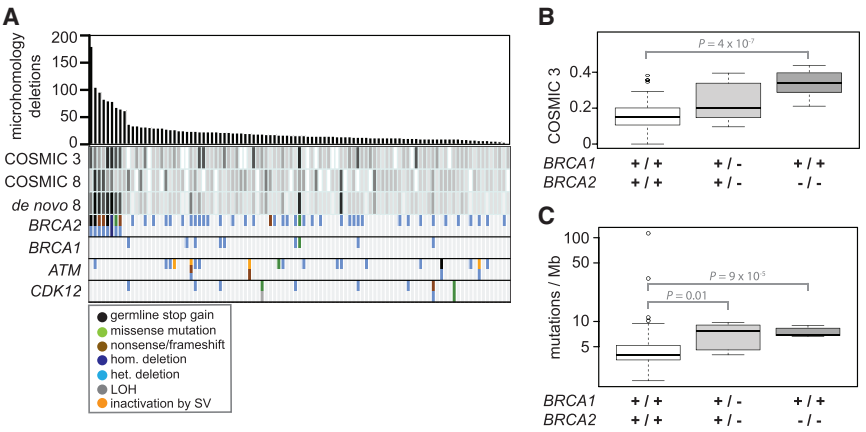


Figure 4. Mutational Signatures of DNA Damage in mCRPC

(A) From top to bottom: the frequency of deletions bearing two or more nucleotides of microhomology; fit of mutation signatures COSMIC 3 and 8 and *de novo* 8; alterations associated with DNA repair by homologous recombination. See also Figure S6.

(B) Box and whisker plots showing mutation frequency in samples bearing either biallelic loss of *BRCA2* or compound *BRCA1-BRCA2* heterozygosity, compared to samples lacking either of these alterations.

(C) Box and whisker plots showing COSMIC signature 3 fit in tumors bearing biallelic loss of *BRCA2* and samples bearing compound *BRCA1-BRCA2* heterozygosity.

enhancer region amplifications, an increase over the 63% of cases identified as carrying *AR* alterations in a benchmark exome study of comparable size (Robinson et al., 2015). ETS family genes were activated by fusions in 59% of cases. We observed MAPK driver mutations in *HRAS* (p.Q61K, 2%) and *BRAF* (p.G469A, 1%). Putative dominant negative *SPOP* mutations were present in 5% of cases (Barbieri et al., 2012; Blattner et al., 2017). ETS gene family activations were mutually exclusive with activating alterations in the RAS/MAPK pathway members ($p = 0.01$, Fisher's exact test) and with inactivation of *SPOP* and *CHD1* (Barbieri et al., 2012; Burkhardt et al., 2013; Huang et al., 2012). A single *IDH1* mutation at the previously reported p.R132C hotspot was observed (Abeshouse et al., 2015). Additionally, mutually exclusive alterations affecting genes that modulate the *AR* pathway (*FOXA1*, *NCOR1*, *NCOR2*, and *ASXL2*) were present in 29% of cases. Alterations in WNT pathway members *CTNNB1*, *APC*, and *ZNRF3* that were predicted to activate WNT signaling were mutually exclusive in all but one of the 17% of cases where they were present. Previously unreported inactivating events targeting *HDAC4* were present in 6% of cases. No somatic alteration was significantly associated with tissue biopsy site after accounting for multiple testing correction. We searched for recurrent point mutations affecting the promoter, enhancer, and UTR regions of 574 known cancer driver genes (Table S5). This analysis identified 101 mutations of unknown significance; no variant was significantly associated with expression or structural variation phenotypes.

We next assessed the frequency of mutations in genes responsible for DNA damage repair. Inactivating germline alterations were present in the DNA repair genes (*BRCA2* and *ATM*) in 4% of samples, a slightly lower frequency than the ~10% frequency observed in a large study of metastatic prostate tumors (Pritchard et al., 2016). Somatic alterations alone accounted for five of the eight cases of biallelic *BRCA2* inactivation and all three tumors carrying biallelic *CDK12* inactivation. Biallelic *BRCA2*, *CDK12*, and *ATM* inactivating mutations were mutually exclusive, and the total frequency of biallelic *BRCA2*, *CDK12*, and *ATM* inactivation was 15%. Two hypermutated samples were present, consistent with the reported 3% frequency of mismatch repair defects in mCRPC (Robinson et al., 2015). One hypermutated sample bore deep deletion in mutS homolog 2 (*MSH2*) and *MSH6*, mismatch repair genes 300 kb apart on chromosome 2, an alteration predicted to abrogate mismatch repair.

DISCUSSION

In contrast to previously published large-scale analyses of primary and metastatic prostate cancer that have largely focused on the coding genome (Abeshouse et al., 2015; Armenia et al., 2018; Barbieri et al., 2012; Beltran et al., 2013; Fraser et al., 2017; Robinson et al., 2015; Taylor et al., 2010; Wedge et al., 2018), we have performed whole-genome analysis of metastases from 101 mCRPC patients at 109x depth of coverage in tumor samples. This coverage, 2-fold deeper than previous efforts in this space (Wedge et al., 2018) and performed on a large patient cohort, has produced a unique resource for dissecting structural variation in metastatic prostate cancer samples. Our data emphasize that structural variations may inactivate tumor suppressors by disrupting the coding region of these genes (Patch et al., 2015; Waddell et al., 2015), whereas both fusions and alterations affecting intergenic regulatory elements appear to activate driver genes. Fusions driving proteins such as *AXL* or *BRAF* that can be targeted therapeutically may open directions for new treatments in mCRPC. We derived insight into the etiology of structural variation, associating *BRCA2*, *CDK12*, and *TP53* with deletions, tandem duplications, and chromothripsis. Our novel observation that non-coding RNAs such as *SCHLAP1* and *RP11-356O9.1* drive oncogene expression highlights the under-explored role of non-coding genes in mCRPC and will serve as the foundation for further studies of the non-coding genome.

All of the men in this study had developed resistance following front-line treatment with androgen deprivation therapy (ADT). A key finding made possible by our integrated analysis of the whole genome and transcriptome across a large population of mCRPC patients is that amplification of a putative enhancer region 624 kb distant from *AR* was present in 81% of men and 85% had either amplification or pathogenic activating *AR* mutation. Our data support the model that amplification at the putative enhancer locus results in increased *AR* expression. In 13% of

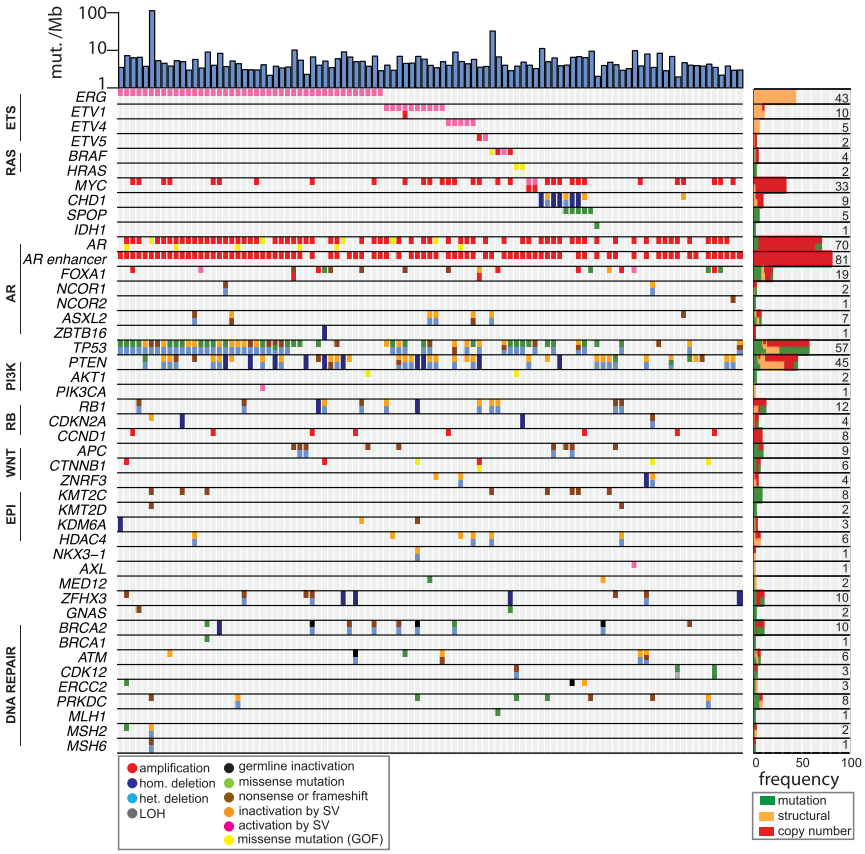


Figure 5. Landscape of Somatic and Structural Alterations in mCRPC

Mutation frequency (top) and germline or somatic alterations in key genes where such alterations were predicted to be functionally meaningful. Alteration frequency shown at right.
See also Tables S3 and S5.

men, putative enhancer amplification was present without alterations in *AR* itself. This finding suggests that DNA copy gain affecting this locus, commonly by tandem duplication, may be a frequent mechanism by which prostate tumor cells initially develop ADT resistance (Karantanos et al., 2013). Observations of tandem duplication at putative enhancers near *AR*, *MYC*, and *FOXA1* underline the value of whole-genome analysis, even in diseases where exome analysis has been performed in large cohorts of patients.

We observed chromothripsis in 23% of mCRPC patients and demonstrated that chromothripsis was significantly associated with *TP53* alterations. This observation supports the proposed but unproven mechanistic association between *TP53* alteration and chromothripsis (Rausch et al., 2012; reviewed in Maher and Wilson, 2012). However, *TP53* alterations cannot be the sole driver of chromothripsis, as chromothripsis is not widespread in other tumors with high rates of *TP53* inactivation such as high grade serous ovarian carcinoma (Zack et al., 2013). In our study, chromothripsis was mutually exclusive with biallelic inactivation of *BRCA2*, inconsistent with a model where cells lacking the ability to perform homologous recombination repair of double-strand DNA breaks would be predisposed to chromothripsis.

Our study linked biallelic inactivation of *BRCA2* but not *ATM* or *CDK12* with deletions that manifest flanking microhomology (Figure 4A). It is not yet clear what combination of genotype and genomic data will best identify the patients who will benefit from PARP inhibitor therapy (Mateo et al., 2015). The 6% of samples with compound *BRCA1*/*BRCA2* heterozygosity lacked deletions with flanking microhomology but had significantly increased mutation rates not statistically distinguishable from biallelic *BRCA2* tumors. Dissecting the functional consequences of these alterations will have implications for patient selection when considering treatment with a PARP inhibitor (Lord and Ashworth, 2016).

Our study demonstrates the utility of whole-genome analysis across a clinically relevant metastatic tumor cohort, as our analysis led to multiple discoveries that eluded existing exome-centric genomic investigations in the advanced disease setting. We have provided the first landscape of structural variants in mCRPC, a substantial mutational class in this disease that will serve as a repository for other researchers to continue exploring their biological and clinical significance. Our data also provides the foundation for further dissection of the non-coding genome through complementary profiling efforts (e.g., epigenetics) and subsequent preclinical studies that may have translational impact in prostate cancer patients.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient Cohort
- METHOD DETAILS
  - Sample Preparation and DNA Sequencing
  - RNA Sequencing
  - Whole-Genome Sequencing Data Analysis
  - Mutation signature analysis
  - Evaluation of deletions with flanking microhomology
  - Evaluation of chromothripsis and chromoplexy
  - Noncoding mutation analysis
  - Data visualization and reporting
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical Analysis

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and five tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.06.039.

ACKNOWLEDGMENTS

We thank the patients who selflessly contributed samples to this study and without whom this research would not have been possible. This research was primarily supported by a Stand Up To Cancer-Prostate Cancer Foundation Prostate Cancer Dream Team Award (SU2C-AACR-DT0812 to E.J.S.) and by the Movember Foundation. Stand Up To Cancer is a division of the Entertainment Industry Foundation. This research grant was administered by the American Association for Cancer Research, the scientific partner of SU2C. This project was also supported by the following awards: Goldberg-Benioff Research Fund for Prostate Cancer Translational Biology (to F.Y.F.), Stand Up To Cancer-Prostate Cancer Foundation Prostate Cancer Dream Team Award (SU2C-AACR-DT0712 to A.M.C.), several Prostate Cancer Foundation Challenge grants (to C.A.M., F.Y.F., S.M.D., and L.F.), V Foundation Scholar Grant (to F.Y.F.), BRCA Foundation Young Investigator Award (to D.A.Q.), Department of Defense (DOD) (W81XWH-16-1-0747 to F.Y.F., W81XWH-15-1-0562 to A.M.C., PC160429 to M.P.C., and W81XWH-17-1-0192 to Hui Li), Early Detection Research Network (U01 CA214170 to A.M.C.), Prostate SPORE (P50 CA186786 and P50 CA097186 to A.M.C.), NIH (R01CA174777 to S.M.D.) and NCI T32 training grant (CA108462 to J.C.). D.A.Q., S.G.Z., R.A., M.P.C., W.K., and V.K. are supported by Prostate Cancer Foundation Young Investigator Awards.

AUTHOR CONTRIBUTIONS

D.A.Q., H.X.D., S.G.Z., P.L., M.P., J.Z., M.C., A.P., N.A.L., A.Z., R.M.-B., H.G., L.A.G., T.C.G., J.M.S., S.A.T., D.E.S., R.K.C., D.T.C., K.F., J.S.G., J. H., A.L., J.S., S.B., K.E.K., H.H.H., A.W.W., S.M.D., C.A.M., and F.Y.F. contributed with algorithm development or analysis of genomic data. D.A.Q., H.X.D., S.G.Z., and C.A.M., oversaw bioinformatics analyses. R.A., J.J.A., A.F., A.M.B., D.P., T.J.B., L.Z., J.F.Y., T.M.B., G.T., K.N.C., M.G., R.E.R., M.B.R., O.W., C.J.R., L.F., W.K., T.F., J.C., P.N.L., C.P.E., P.F., J. Hakenberg, J. Huang, and E.J.S. developed the patient cohort, obtained tumor biopsies, performed molecular testing for metastatic cases, and carried out data interpretation of the clinical cohort. A.F., V.K., D.P., Hui Li, R.D., and Haolong Li contributed to sample processing. F.Y.F., C.A.M., A.M.C., and E.J.S. conceived the study concept. D.A.Q., H.X.D., S.G.Z., A.A., A.M.C., C.A.M., E.J.S., and F.Y.F. wrote the manuscript, which all authors reviewed.

DECLARATION OF INTERESTS

A.M.C. is on the scientific advisory board of Tempus. F.Y.F. is co-founder of PFS Genomics. R.K.C., D.T.C., K.F., J.S.G., J. Hakenberg, A.L., J.S., S.B., and P.F. are employees of Illumina, Inc. and hold stock in the company. The University of Michigan has been issued a patent on ETS gene fusions in prostate cancer on which A.M.C. and S.A.T. are co-inventors. The diagnostic field of use has been licensed to Hologic/Gen-Probe, Inc., which has sublicensed rights to Roche/Ventana Medical Systems. S.A.T. has an unrelated sponsored research agreement with Astellas. S.A.T. has served as a consultant for and received honoraria from Almac Diagnostics, Janssen, and Astellas/Medivation. S.A.T. is a co-founder of, consultant for, and Laboratory Director of Strata Oncology. K.C. receives consultant, honoraria, and research funding from Janssen and Astellas.

Received: May 15, 2018
Revised: June 11, 2018
Accepted: June 21, 2018
Published: July 19, 2018

REFERENCES

Abeshouse, A., Ahn, J., Akbani, R., Ally, A., and Amin, S.; Cancer Genome Atlas Research Network (2015). The molecular taxonomy of primary prostate cancer. Cell *163*, 1011–1025.

Aggarwal, R., Beer, T.M., Gleave, M., Stuart, J.M., Rettig, M., Evans, C.P., Youngren, J., Alumkal, J.J., Huang, J., Thomas, G., et al. (2016). Targeting adaptive pathways in metastatic treatment-resistant prostate cancer: update on the Stand Up 2 Cancer/Prostate Cancer Foundation-supported West Coast Prostate Cancer Dream Team. Eur. Urol. Focus *2*, 469–471.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nat. Genet. *47*, 1402–1407.

Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz, J., et al. (2006). A common variant associated with prostate cancer in European and African populations. Nat. Genet. *38*, 652–658.

Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., et al.; PCF/SU2C International Prostate Cancer Dream Team; PCF/SU2C International Prostate Cancer Dream Team (2018). The long tail of oncogenic drivers in prostate cancer. Nat. Genet. *50*, 645–651.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., et al. (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat. Genet. *44*, 685–689.

Beltran, H., Yelensky, R., Frampton, G.M., Park, K., Downing, S.R., MacDonald, T.Y., Jarosz, M., Lipson, D., Tagawa, S.T., Nanus, D.M., et al. (2013). Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. Eur. Urol. *63*, 920–926.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., et al. (2010). Signatures of mutation and selection in the cancer genome. Nature *463*, 893–898.

Blattner, M., Liu, D., Robinson, B.D., Huang, D., Poliakov, A., Gao, D., Nataraj, S., Deonarine, L.D., Augello, M.A., Sailer, V., et al. (2017). SPOP mutation drives prostate tumorigenesis in vivo through coordinate regulation of PI3K/mTOR and AR signaling. Cancer Cell *31*, 436–451.

Burkhardt, L., Fuchs, S., Krohn, A., Masser, S., Mader, M., Kluth, M., Bachmann, F., Huland, H., Steuber, T., Graefen, M., et al. (2013). CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. Cancer Res. *73*, 2795–2805.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012a). Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front. Genet. *3*, 35.

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012b). A program for annotating and predicting the

effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

Davies, H., Glodzik, D., Morganella, S., Yates, L.R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A.M., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat. Med. *23*, 517–525.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*. Published online January 1, 2017. https://doi.org/10.1093/database/bax028.

Fraser, M., Sabelnykova, V.Y., Yamaguchi, T.N., Heisler, L.E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A.P., et al. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. Nature *541*, 359–364.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Gehring, J.S., Fischer, B., Lawrence, M., and Huber, W. (2015). SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics *31*, 3673–3675.

Glover, T.W., Wilson, T.E., and Arlt, M.F. (2017). Fragile sites in cancer: more than meets the eye. Nat. Rev. Cancer *17*, 489–501.

Grasso, C.S., Wu, Y.-M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. Nature *487*, 239–243.

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil, E., Brewer, D.S., Kallio, H.M.L., Högnäs, G., Annala, M., et al.; ICGC Prostate Group (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353–357.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Holmes, M.G., Foss, E., Joseph, G., Foye, A., Beckett, B., Motamedi, D., Youngren, J., Thomas, G.V., Huang, J., Aggarwal, R., et al. (2017). CT-guided bone biopsies in metastatic castration-resistant prostate cancer: factors predictive of maximum tumor yield. J. Vasc. Interv. Radiol. *28*, 1073–1081.

Huang, S., Gulzar, Z.G., Salari, K., Lapointe, J., Brooks, J.D., and Pollack, J.R. (2012). Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness. Oncogene *31*, 4164–4170.

Jones, D.T.W., Kocialkowski, S., Liu, L., Pearson, D.M., Bäcklund, L.M., Ichimura, K., and Collins, V.P. (2008). Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. Cancer Res. *68*, 8673–8677.

Karantanos, T., Corn, P.G., and Thompson, T.C. (2013). Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. Oncogene *32*, 5501–5511.

Kron, K.J., Murison, A., Zhou, S., Huang, V., Yamaguchi, T.N., Shiah, Y.J., Fraser, M., van der Kwast, T., Boutros, P.C., Bristow, R.G., and Lupien, M. (2017). TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nat. Genet. *49*, 1336–1345.

Lord, C.J., and Ashworth, A. (2016). BRCAness revisited. Nat. Rev. Cancer *16*, 110–120.

Maher, C.A., and Wilson, R.K. (2012). Chromothripsis and human disease: piecing together the shattering process. Cell *148*, 29–32.

Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., and Chinnaiyan, A.M. (2009). Transcriptome sequencing to detect gene fusions in cancer. Nature *458*, 97–101.

Mateo, J., Carreira, S., Sandhu, S., Miranda, S., Mossop, H., Perez-Lopez, R., Nava Rodrigues, D., Robinson, D., Omlin, A., Tunariu, N., et al. (2015). DNA-repair defects and olaparib in metastatic prostate cancer. N. Engl. J. Med. *373*, 1697–1708.

Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T., and Misawa, S. (1996). Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia *10*, 1911–1918.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Ohno, S. (1970). Evolution by Gene Duplication (London: George Allen and Unwin).

Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.

Patch, A.M., Christie, E.L., Etemadmoghadam, D., Garsed, D.W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P.J., et al.; Australian Ovarian Cancer Study Group (2015). Whole-genome characterization of chemoresistant ovarian cancer. Nature *521*, 489–494.

Polak, P., Kim, J., Braunstein, L.Z., Karlic, R., Haradhavala, N.J., Tiao, G., Rosebrock, D., Livitz, D., Kübler, K., Mouw, K.W., et al. (2017). A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat. Genet. *49*, 1476–1486.

Popova, T., Manié, E., Boeva, V., Battistella, A., Goundiam, O., Smith, N.K., Mueller, C.R., Raynal, V., Mariani, O., Sastre-Garau, X., and Stern, M.H. (2016). Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. Cancer Res. *76*, 1882–1891.

Prensner, J.R., Iyer, M.K., Sahu, A., Asangani, I.A., Cao, Q., Patel, L., Vergara, I.A., Davicioni, E., Erho, N., Ghadessi, M., et al. (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat. Genet. *45*, 1392–1398.

Prensner, J.R., Sahu, A., Iyer, M.K., Malik, R., Chandler, B., Asangani, I.A., Poliakov, A., Vergara, I.A., Alshalalfa, M., Jenkins, R.B., et al. (2014). The lncRNAs PCGEM1 and PRNCR1 are not implicated in castration resistant prostate cancer. Oncotarget *5*, 1434–1438.

Prensner, J.R., Chen, W., Han, S., Iyer, M.K., Cao, Q., Kothari, V., Evans, J.R., Knudsen, K.E., Paulsen, M.T., Ljungman, M., et al. (2014b). The long noncoding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. Neoplasia *16*, 900–908.

Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N. Engl. J. Med. *375*, 443–453.

R Core Team (2018). R: A Language and Environment for Statistical Computing (Vienna: R Foundation for Statistical Computing).

Raczy, C., Petrovski, R., Saunders, C.T., Chorny, I., Kruglyak, S., Margulies, E.H., Chuang, H.Y., Källberg, M., Kumar, S.A., Liao, A., et al. (2013). Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics *29*, 2041–2043.

Rausch, T., Jones, D.T., Zapatka, M., Stütz, A.M., Zichner, T., Weischenfeldt, J., Jäger, N., Remke, M., Shih, D., Northcott, P.A., et al. (2012). Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell *148*, 59–71.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate cancer. Cell *161*, 1215–1228.

Roller, E., Ivakhno, S., Lee, S., Royce, T., and Tanner, S. (2016). Canvas: versatile and scalable detection of copy number variants. Bioinformatics *32*, 2375–2377.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31.

Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics *28*, 1811–1817.

Siegel, R.L., Miller, K.D., and Jemal, A. (2018). Cancer statistics, 2018. CA Cancer J. Clin. *68*, 7–30.

Spritzer, C.E., Afonso, P.D., Vinson, E.N., Turnbull, J.D., Morris, K.K., Foye, A., Madden, J.F., Roy Choudhury, K., Febbo, P.G., and George, D.J. (2013). Bone marrow biopsy: RNA isolation with expression profiling in men with metastatic castration-resistant prostate cancer–factors affecting diagnostic success. Radiology *269*, 816–823.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.

Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. Cancer Cell *18*, 11–22.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644–648.

Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature *448*, 595–599.

Tutt, A., Bertwistle, D., Valentine, J., Gabriel, A., Swift, S., Ross, G., Griffin, C., Thacker, J., and Ashworth, A. (2001). Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. EMBO J. *20*, 4704–4716.

Visakorpi, T., Hyytinen, E., Koivisto, P., Tanner, M., Keinänen, R., Palmberg, C., Palotie, A., Tammela, T., Isola, J., and Kallioniemi, O.P. (1995). In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat. Genet. *9*, 401–406.

Waddell, N., Pajic, M., Patch, A.-M., Chang, D.K., Kassahn, K.S., Bailey, P., Johns, A.L., Miller, D., Nones, K., Quek, K., et al.; Australian Pancreatic Cancer Genome Initiative (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. Nature *518*, 495–501.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Wang, Y.K., Bashashati, A., Anglesio, M.S., Cochrane, D.R., Grewal, D.S., Ha, G., McPherson, A., Horlings, H.M., Senz, J., Prentice, L.M., et al. (2017). Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. Nat. Genet. *49*, 856–865.

Wedge, D.C., Gundem, G., Mitchell, T., Woodcock, D.J., Martincorena, I., Ghori, M., Zamora, J., Butler, A., Whitaker, H., Kote-Jarai, Z., et al.; CAMCAP Study Group; TCGA Consortium (2018). Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. Nat. Genet. *50*, 682–692.

Wyatt, A.W., Mo, F., Wang, K., McConeghy, B., Brahmbhatt, S., Jong, L., Mitchell, D.M., Johnston, R.L., Haegert, A., Li, E., et al. (2014). Heterogeneity in the inter-tumor transcriptome of high risk prostate cancer. Genome Biol. *15*, 426.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat. Genet. *39*, 645–649.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.

Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat. Med. *23*, 703–713.

Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R package for Circos 2D track plots. BMC Bioinformatics *14*, 244.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
QIAamp Fast DNA Tissue Kit	QIAGEN	51404
QIAamp DNA Blood Mini Kit	QIAGEN	51104
Quant-iTTM PicoGreen dsDNA Reagent	ThermoFisher	P11496
Illumina DNA Sample Preparation HT Kit	Illumina	FC-121-2003
Agencourt AMPure XP Beads	Beckman Coulter	A63880
Illumina HiSeq X HD Paired End Cluster Kit	Illumina	FC-501-2501
Illumina HiSeq X HD SBS Kit	Illumina	FC-401-4002
Agilent Absolutely RNA Nano Prep	Agilent	400753
Agilent Bioanalyzer RNA 6000 Pico	Agilent	5067-1513
NuGEN Ovation RNA-Seq System V2	NuGEN	7102-08
NuGEN Ovation Ultralow System V2	NuGEN	0344NB-08
High Output 150 cycle V2 reagents	Illumina	FC-404-2002
Deposited Data		
Raw data	this study	dbGAP: phs001648.v1.p1
Human Reference Genome NCBI GRCh38 PAR-masked with decoys hs38d1	Illumina	https://support.illumina.com/sequencing/sequencing_software/igenome.html
CHIP-seq analysis of primary prostate cancer	Kron et al., 2017	GEO: GSE96652
COSMIC Cancer Gene Census	Futreal et al., 2004	https://cancer.sanger.ac.uk/cosmic/census
Software and Algorithms		
R v.3.3.3	R Core Team, 2018	https://www.r-project.org
Whole Genome Sequencing app v7.0.1	Illumina BaseSpace	https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/whole-genome-sequencing.html
Isaac v04.17.06.15	Raczy et al., 2013	https://github.com/Illumina/Isaac4
Strelka v2.8.0	Saunders et al., 2012	https://github.com/Illumina/strelka
Mutect v1.1.7	Cibulskis et al., 2013	http://archive.broadinstitute.org/cancer/cga/mutect
Manta v1.1.1	Chen et al., 2016	https://github.com/Illumina/manta
Canvas v1.28.0-O01073	Roller et al., 2016	https://github.com/Illumina/canvas
CopyCat	https://github.com/chrisamiller/copyCat	https://github.com/chrisamiller/copyCat
DNAcopy v1.54.0	Olshen et al., 2004	https://bioconductor.org/packages/release/bioc/html/DNAcopy.html
RNaseq alignment app v1.1.0	Illumina Inc.	https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/rna-seq-alignment.html
Integrated Genomics Viewer	Robinson et al., 2011	https://software.broadinstitute.org/software/igv
Illumina Variant Interpreter	Illumina Inc.	https://variantinterpreter.informatics.illumina.com
snpSift v4.3	Cingolani et al., 2012b	http://snpeff.sourceforge.net/SnpEff.html
deconstructSigs v.1.8.0	Rosenthal et al., 2016	https://github.com/raerose01/deconstructSigs
SomaticSignatures v.2.16.0	Gehring et al., 2015	http://bioconductor.org/packages/release/bioc/html/SomaticSignatures.html
ChainFinder v1.0.1	Baca et al., 2013	http://archive.broadinstitute.org/cancer/cga/chainfinder
RCircos v1.2.0	Zhang et al., 2013	https://cran.r-project.org/web/packages/RCircos/index.html
ANNOVAR v2018Apr16	Wang et al., 2010	http://annovar.openbioinformatics.org/en/latest/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Felix Feng (Felix.Feng@ucsf.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient Cohort

Patient tissue samples were obtained through the Stand Up 2 Cancer/Prostate Cancer Foundation–funded West Coast Prostate Cancer Dream Team project, a multi-center study that acquired biopsies of metastases from men with mCRPC. All patients were male and ranged in age from 45-90 years when biopsied. See also Table 1 for additional clinical details. Samples were obtained by image-guided core needle biopsy of metastatic lesions in bone, soft tissue, or an organ. Fresh-frozen tissue and peripheral blood drawn at the time of biopsy was shipped to a central facility at UCSF for laser-capture microdissection and DNA and RNA extraction. Human studies were approved and overseen by the UCSF Institutional Review Board. All individuals provided written informed consent to obtain fresh tumor biopsies and to perform comprehensive molecular profiling of tumor and germline samples.

METHOD DETAILS

Sample Preparation and DNA Sequencing

Biopsies identified by histological assessment (H&E, serial section) to contain at least 50% tumor were selected for genomic DNA (gDNA) isolation through microdissection of frozen sections (200-500 μm total section depth, QIAGEN QIAamp Fast DNA Tissue Kit, Cat. 51404). Matched normal gDNA was extracted from peripheral blood drawn at time of biopsy (QIAGEN QIAamp DNA Blood Mini Kit, Cat. 51104). Tumor and Normal DNA were quantified prior to library construction using PicoGreen (Quant-iTTM PicoGreen dsDNA Reagent, ThermoFisher Scientific, Catalog #P11496). Quantifications were measured using a Spectromax Gemini XPS (Molecular Devices). PCR-free paired-end libraries were generated by automated liquid handlers using 500-1000 ng input gDNA and the Illumina DNA Sample Preparation HT Kit. Pre-fragmentation gDNA cleanup was performed using paramagnetic sample purification beads (Agencourt AMPure XP reagents, Beckman Coulter). Samples were fragmented, and libraries size-selected following fragmentation and end-repair using paramagnetic sample purification beads to enrich for short insert sizes. Final libraries were quantified by qPCR and evaluated for quality using gel electrophoresis separation. DNA libraries were denatured, diluted and clustered onto patterned flow cells using the Illumina cBot system with Illumina HiSeq X HD Paired End Cluster Kit reagents. Clustered patterned flow cells were loaded onto HiSeq X instruments and sequenced on 151 bp paired-end, non- indexed runs on independent lanes, using HiSeq X HD SBS Kit reagents. Illumina HiSeq Control Software (HCS), and Real-Time Analysis (RTA) wren used with the HiSeq X sequencers for real-time image analysis, and base calling.

RNA Sequencing

Tumor cores were fresh-frozen in OCT for gene expression analysis. Laser capture microdissection was performed on frozen sections to enrich for tumor content (Spritzer et al., 2013). Total RNA was isolated (Agilent Absolutely RNA Nano Prep, Cat. 400753) and samples of sufficient quality (Agilent Bioanalyzer RNA 6000 Pico, Cat. 5067-1513) were amplified using NuGEN Ovation RNA-Seq System V2. cDNA fragmentation was performed on a Covaris M220 sonicator to 200bp. Libraries were generated using NuGEN Ovation Ultralow System V2 for Illumina sequencing. RNA-Seq for 88 samples was performed on an Illumina NextSeq500 (High Output 150 cycle V2 reagents, Cat. FC-404-2002) in 2x76bp paired-end runs; 13 additional samples were sequenced on an Illumina HiSeq2500 at 2x100bp (10), 2x101bp (2) and 2x50bp (1) paired-end runs.

Whole-Genome Sequencing Data Analysis

Whole-genome FASTQ files were uploaded to Illumina BaseSpace Sequence Hub (https://basespace.illumina.com). The Whole Genome Sequencing (WGS) app version 7.0.1 was used to coordinate sample alignment to the NCBI GRCh38 PAR-masked with decoys hs38d1 reference genome (hg38-decoy) and subsequent analytical steps. Reads were aligned against hg38-decoy using the Isaac aligner version 04.17.06.15 (Raczy et al., 2013). Germline mutation analysis was performed using Strelka version 2.8.0 (Saunders et al., 2012) filtered to require an assignment of PASS and snpEff version 4.3g (Cingolani et al., 2012b) labels of "pathogenic," "splice_donor," "splice_acceptor," "stop_gain," or "frameshift." Somatic mutation analysis was performed with Strelka and Mutect version 1.1.7 (Cibulskis et al., 2013), excluding samples lacking a PASS designation. DNA structural variants were identified using Manta version 1.1.1 (Chen et al., 2016), requiring calls to bear a PASS or MGE10kb designation, tumor split read + tumor paired read $\geq$ 10, matched normal split reads = 0, and matched normal paired reads = 0. DNA copy number variants were identified using Canvas version 1.28.0-O01073 (Roller et al., 2016) and CopyCat (https://github.com/chrisamiller/copyCat) using the *runPairedSampleAnalysis* method with default parameters and performing GC correction. Copy number ratios were segmented by Circular Binary Segmentation implemented in the DNAcopy package (Olshen et al., 2004). RNA-seq analysis was performed using the Illumina RNaseq alignment app v1.1.0, aligning RNA FASTQ files to hg38-decoy using STAR version 2.5.0b (Dobin et al., 2013).

Manual review of DNA and RNA data was performed using the Integrated Genomics Viewer (Robinson et al., 2011) (https://software.broadinstitute.org/software/igv) and the Illumina Variant Interpreter (https://variantinterpreter.informatics.illumina.com).

Mutation signature analysis

To perform per-sample mutation counting, all somatic mutations that were not excluded by quality filtering steps were counted. For mutation signature analysis this list was filtered using snpSift version 4.3g, including all alterations designated with the call "SNP" (Cingolani et al., 2012a). Evaluation of COSMIC mutation signatures was performed using the *deconstructSigs* package (Rosenthal et al., 2016), using the BSgenome.Hsapiens.UCSC.hg38 reference, the *signatures.cosmic* comparison set, a *signature.cutoff* value of 0.06, and a *tri.counts.method* parameter of "default." De novo mutation signatures were derived using non-negative matrix factorization implemented in the *SomaticSignatures* R package (Gehring et al., 2015).

Evaluation of deletions with flanking microhomology

Deletions bearing microhomology were identified by a script counting deletions with two or more nucleotides of identical sequence between either 1) the 5′ end of the deleted region (determined from the HG38 genome reference) and the 3′ end following the deleted region or 2) the 3′ end of the deleted region and the 5′ end immediately preceding the deletion.

Evaluation of chromothripsis and chromoplexy

Chromothripsis was evaluated by counting the number of insertions, deletions, and copy number alterations within a moving 20 Mb. window positioned at 10 Kb. intervals along the entire genome, excluding telomeres and centromeres. Windows bearing at least 15 inversion rearrangements, 15 alternating copy number switches, and 10 deletions were called positive for chromothripsis. Chromoplexy was evaluated by applying the ChainFinder application version 1.0.1 (Baca et al., 2013) obtained from http://archive.broadinstitute.org/cancer/cga/chainfinder, using a deletion threshold of −0.278 and a significance threshold of 0.05. The presence of chromoplexy was defined by the presence of a chromoplexy chain connecting at least three chromosomes.

Noncoding mutation analysis

Recurrent promoter and untranslated region (UTR) point mutations were nominated by identifying mutations with variant allele frequency of at least 10% that were present in gene untranslated regions, enhancers, or promoters. UTR and promoter annotation were performed using ANNOVAR v2018Apr16 (Wang et al., 2010). Promoter regions were defined as 1 Kb upstream of the transcription start site. Enhancer regions were nominated by intersecting regions predicted by GeneHancer (Fishilevich et al., 2017) with regions enriched for H3K27ac histone modification identified by CHIPseq in (Kron et al., 2017). A peak in any of 19 samples in that dataset was considered sufficient for inclusion in this analysis. Analysis of recurrent mutations in noncoding regions was restricted to regulatory regions predicted to affect any of the 574 genes listed in Tier 1 of the COSMIC Cancer Gene Census v85, obtained from https://cancer.sanger.ac.uk/cosmic/census (Futreal et al., 2004).

Data visualization and reporting

Circos plots were generated using the *RCircos* R package (Zhang et al., 2013).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis

All statistical analysis was performed using R (v3.3.3) (R Core Team, 2018). Between-group comparisons of continuous variables were performed with the Wilcoxon rank sum test. Contingency table tests were performed with Fisher's exact test. Correlation was assessed with Spearman's correlation. All tests were two-sided.

DATA AND SOFTWARE AVAILABILITY

Sequencing Data

The accession number for the raw sequencing data reported in this paper is dbGAP: phs001648.v1.p1.

Cell

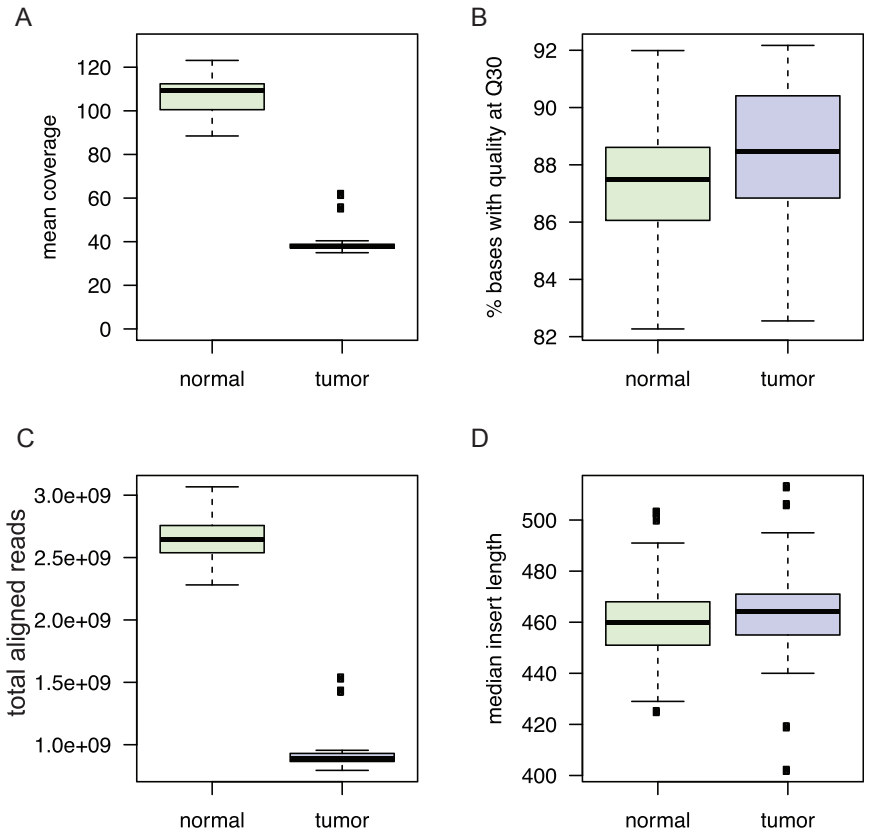


Figure S1. DNA Sequencing Depth and Quality Metrics in Normal and Tumor Samples, Related to STAR Methods

(A) Box and whisker plots showing distribution of mean coverage in normal, tumor.

(B) Percentage of bases at sequencing quality $\geq$ Q30.

(C) Total number of aligned reads.

(D) Median insert length.

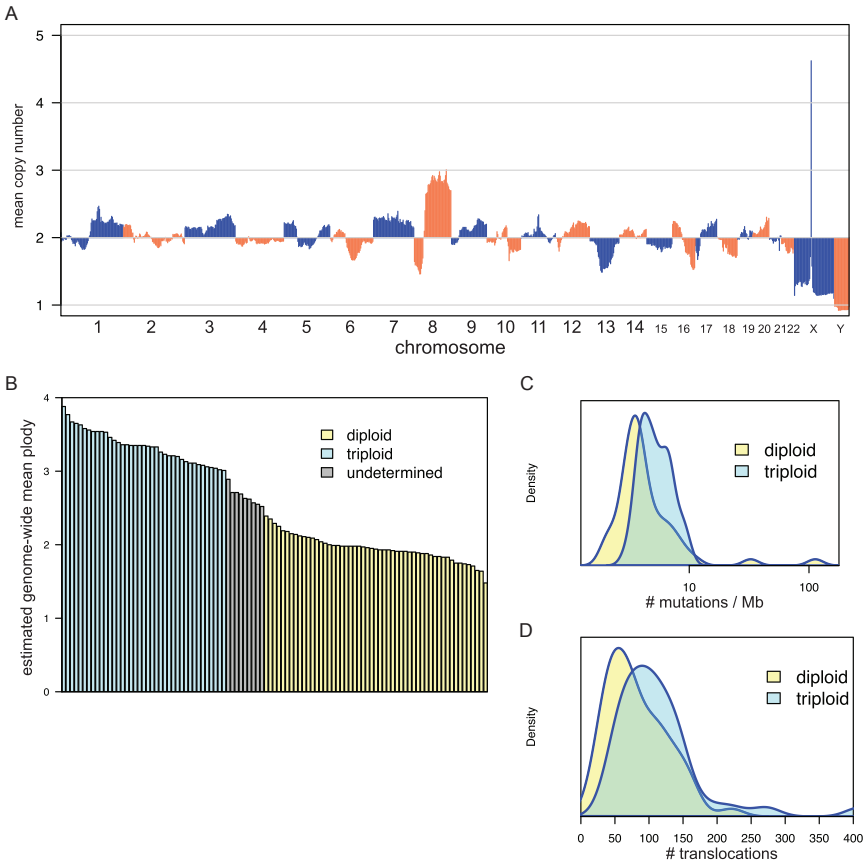


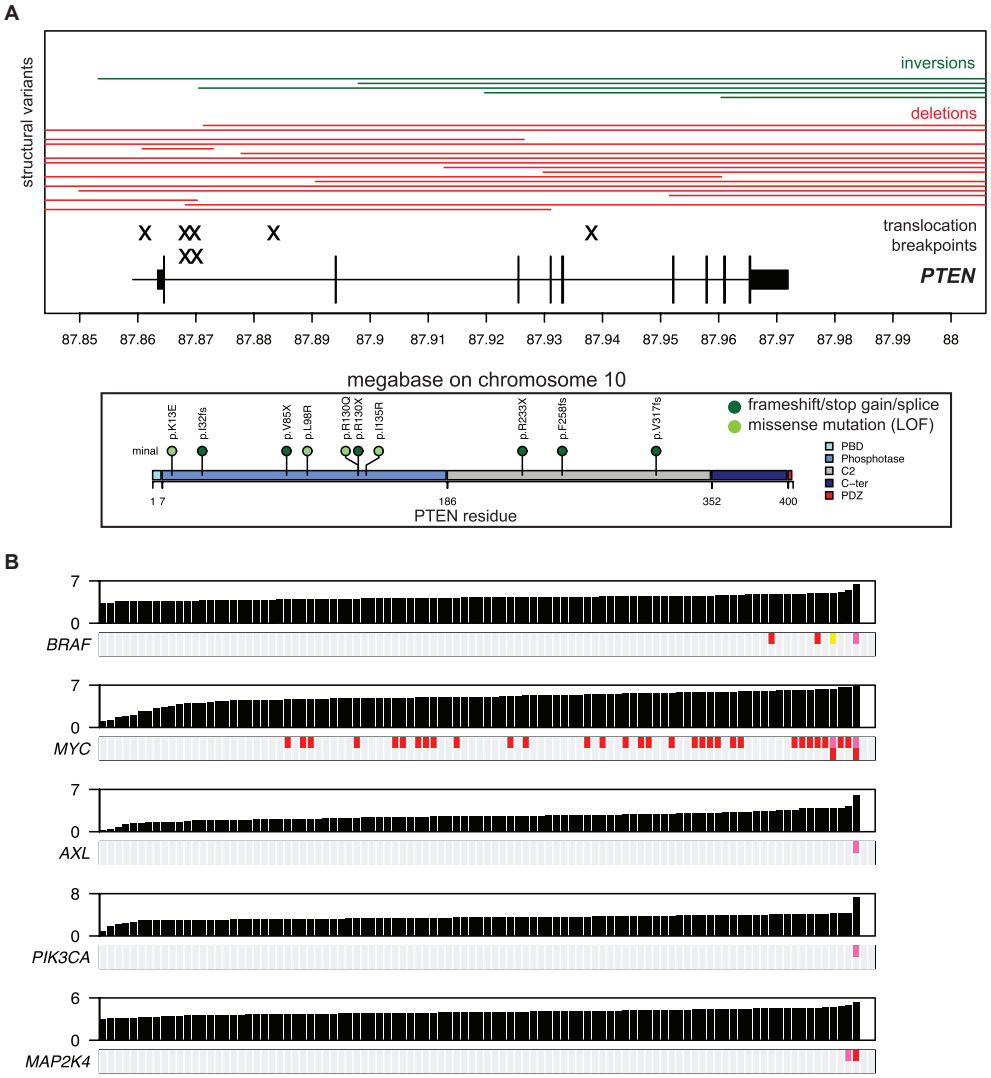
Figure S2. Genome-wide Assessment of Copy Number and Ploidy, Related to Figures 1 and S1 and Table S1

(A) Mean copy number. Baseline copy number for chromosomes X and Y is 1 copy.

(B) Mean genome-wide ploidy estimates. Estimated ploidy values for each sample are listed in Table S1.

(C) Density plot for the number of mutations per megabase in tumors assigned diploid or triploid status as in (B).

(D) Density plot for the number of translocations per megabase in tumors assigned diploid or triploid status as in (B).



C

***ACPP-AXL* fusion**

CGGAATTGTACTTTGAGAAGGGGGAGTACTTTGTGGAGATGTACTATCGGAATGAGACGCAGCACGAGCCGTATCCCCTCATGCTACCTGGCTGCAGCCCCAGCTGTCCTC
TGGAGAGGTTTGCTGAGCTGGTTGGCCCTGTGATCCCTCAAGACTGGTCCACGGAGTGTATGACCACAAACAGCCATCAA**GTGAAGGAACCTTCAACTCCTGCCTTCTCGT**
GGCCCTGGTGGTATGTACTGCTAGGAGCAGTCGTGGCCGCTGCCTGTGTCCTCATCTTGGCTCTCTTCCTTGTCCACCGGCGAAAGAAGGAGACCCGTTATGGAGAAGTGT
TTGAACCAACAGTGGAAAGAGGTGAACTGGTAGTCAGGTACCGCGTGCGCAAGTCCTACAGTCGTCGGACCACTGAAGCTACCTTGAACAGCCTGGGCATCAGTGAAGAGC
TGAAGGA

ACPP sequence in plain text
***AXL* sequence in bold text**

Figure S3. Structural Variants Affecting Oncogenes and *PTEN*, Related to Figure 1

(A) Schematic illustration of the *PTEN* gene locus, DNA copy loss (red lines), intersecting inversion rearrangements (green lines), translocations (black X). Below, locations of pathogenic missense or nonsense mutations (light/dark green circles) are indicated. Each variation was identified in a separate sample.

(B) Expression of fusions activating oncogenes, sorted in increasing order by gene expression level, with somatic alterations noted below each gene. Red: amplification; Yellow: missense mutation; pink: gene fusion. See also Figure 1E listing upstream fusion partners. RNA expression measurements were available for 99 of 101 samples. RNA values are expressed as log(1+(TPM x 106).

(C) RNA sequence of an ACPP-AXL fusion observed independently in a patient seen at the Vancouver Prostate Center; RNA generated from fresh frozen tumor tissue (radical prostatectomy) with high risk primary prostate cancer and methodology described in Wyatt et al. (2014).

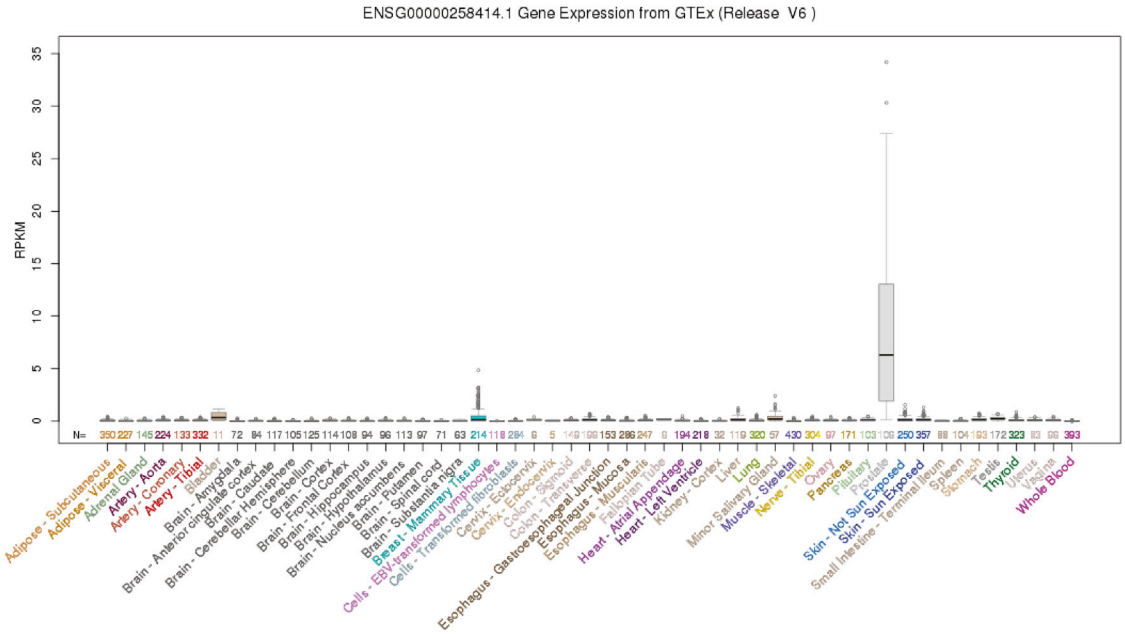


Figure S4. RP11-356O9.1 Is Exclusively Expressed in Prostate, Related to Figure 1

mRNA expression data expressed in RPKM obtained from GTEx as viewed on the UCSC genome browser (http://genome.ucsc.edu).

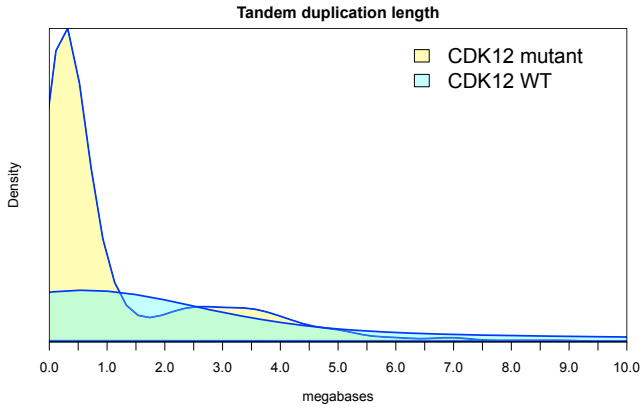


Figure S5. Tandem Duplication Length Has a Bimodal Distribution in *CDK12*-Inactivated Tumors, Related to Figure 3

Density plots of tandem duplication length, contrasting the three *CDK12* mutant samples (yellow) with the 98 *CDK12*-WT samples (teal).

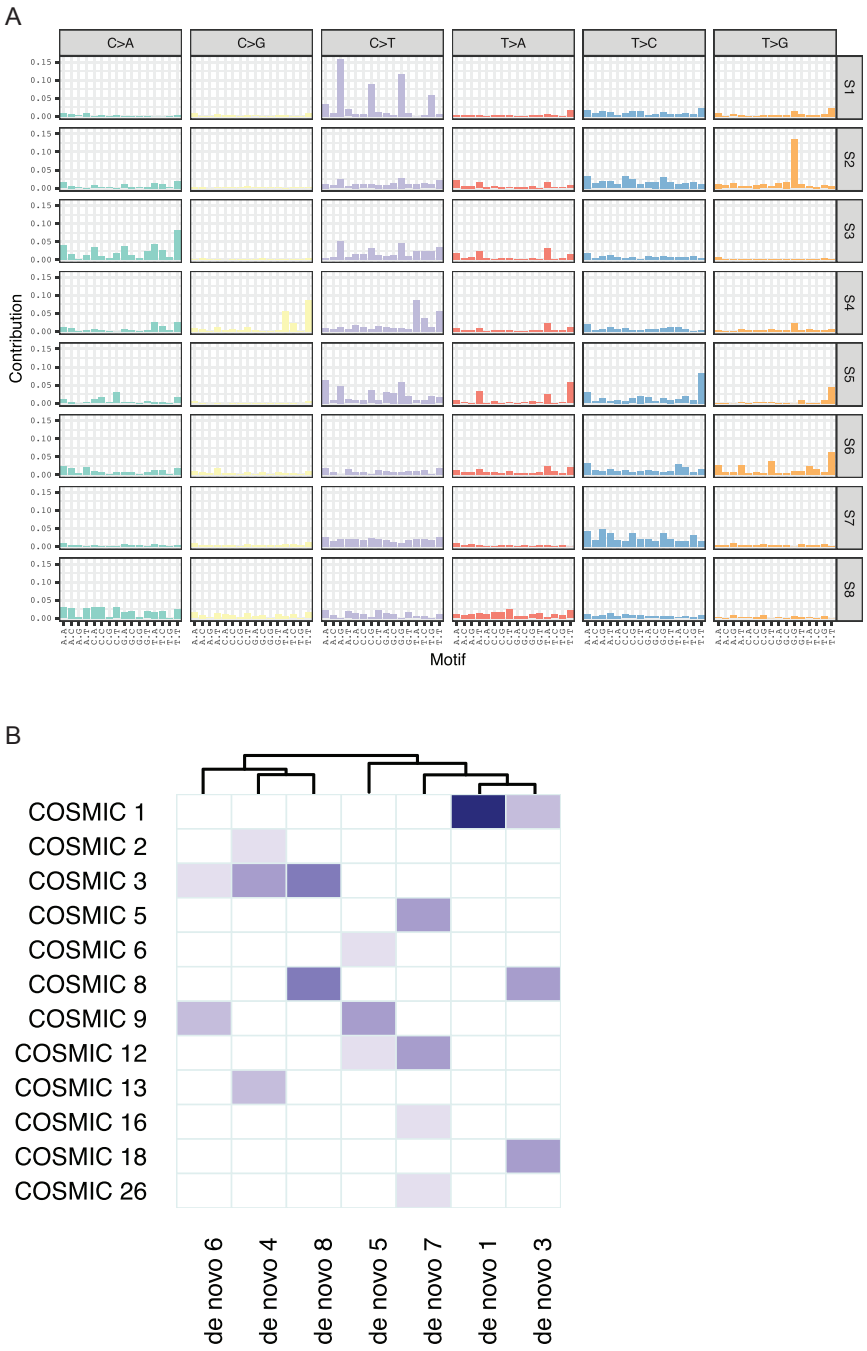


Figure S6. Somatic Signatures Identified in mCRPC, Related to Figure 4

(A) Trinucleotide context of signatures *de novo* 1 through 8. Signature *de novo* 2 was most likely due to a technical artifact, and was not considered during analysis.

(B) COSMIC versus *de novo* comparison of mutation signature fit. Results were hierarchically clustered.