Published in final edited form as:

Nature. 2011 February 10; 470(7333): 214–220. doi:10.1038/nature09744.

The genomic complexity of primary human prostate cancer

Michael F. Berger[1,17,#], **Michael S. Lawrence**[1,17], **Francesca Demichelis**[2,3,17], **Yotam Drier**[4,17], **Kristian Cibulskis**[1], **Andrey Y. Sivachenko**[1], **Andrea Sboner**[5,6], **Raquel Esgueva**[2], **Dorothee Pflueger**[2], **Carrie Sougnez**[1], **Robert Onofrio**[1], **Scott L. Carter**[1], **Kyung Park**[2], **Lukas Habegger**[6], **Lauren Ambrogio**[1], **Timothy Fennell**[1], **Melissa Parkin**[1], **Gordon Saksena**[1], **Douglas Voet**[1], **Alex H. Ramos**[1,7], **Trevor J. Pugh**[1,7,8], **Jane Wilkinson**[1], **Sheila Fisher**[1], **Wendy Winckler**[1], **Scott Mahan**[1], **Kristin Ardlie**[1], **Jennifer Baldwin**[1], **Jonathan W. Simons**[9], **Naoki Kitabayashi**[2], **Theresa Y. MacDonald**[2], **Philip W. Kantoff**[7,8], **Lynda Chin**[1,7,8,10], **Stacey B. Gabriel**[1], **Mark B. Gerstein**[5,6,11], **Todd R. Golub**[1,12,13,14], **Matthew Meyerson**[1,7,8,14], **Ashutosh Tewari**[15], **Eric S. Lander**[1,7,16,18], **Gad Getz**[1,18], **Mark A. Rubin**[2,18,*], and **Levi A. Garraway**[1,7,8,14,18,*]

[1] The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[2] Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York, 10065, USA

[3] Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York, 10021, USA

[4] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, 76100, Israel

[5] Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, Connecticut, 06520, USA

[6] Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, 06520, USA

[7] Harvard Medical School, Boston, Massachusetts, 02115, USA

[*]Correspondence and requests for materials should be addressed to M.A.R. (rubinma@med.cornell.edu) or L.A.G. (levi_garraway@dfci.harvard.edu).

[17]These authors contributed equally to this work.

[18]These authors contributed equally to this work.

[#]Current address: Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York, 10065, USA.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions M.F.B, E.S.L., G.G., M.A.R., and L.A.G. designed the study, analyzed the data, and wrote the paper. M.S.L., F.D., and Y.D. performed analysis of mutations, copy number, rearrangements, and ChIP-Seq associations. K.C. and A.Y.S. performed analysis of mutations and indels. R.E., D.P., N.K., A.T., and M.A.R. contributed to the procurement of tumor tissue and preparation of DNA. C.S, R.O., W.W., S.M., and K.A. participated in DNA sample processing, quality control, and SNP microarray experiments. L.A., J.W., S.F., J.B., and S.B.G. generated the DNA sequence data. A.S., S.L.C., L.H., T.F., G.S., D.V., A.H.R., and T.J.P. provided additional bioinformatic analyses. K.P., T.Y.M., and M.A.R. performed FISH experiments. M.F.B, M.S.L., R.O., M.P., W.W., G.G., and L.A.G. validated candidate rearrangements. J.W.S., P.W.K., L.C., S.B.G., M.B.G., T.R.G., and M.M. contributed to the study design and interpretation of data.

Author Information All Illumina sequence data have been deposited in dbGaP (http://www.ncbi.nlm.nih.gov/gap) and are available at accession phs000330.v1.p1. Reprints and permissions information is available at www.nature.com/reprints.

The authors declare no competing financial interests.

[8] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, 02115, USA

[9] The Prostate Cancer Foundation, Santa Monica, California 90401

[10] Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts, 02115, USA

[11] Department of Computer Science, Yale University, New Haven, Connecticut, 06520, USA

[12] Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, 02115, USA

[13] Howard Hughes Medical Institute, Chevy Chase, MD, 20815, USA

[14] Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts, 02115, USA

[15] Department of Urology, Institute of Prostate Cancer and Lefrak Center of Robotic Surgery, Weill Cornell Medical College and New York Presbyterian Hospitals, New York, New York, 10065, USA

[16] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Abstract

Prostate cancer is the second most common cause of male cancer deaths in the United States. Here we present the complete sequence of seven primary prostate cancers and their paired normal counterparts. Several tumors contained complex chains of balanced rearrangements that occurred within or adjacent to known cancer genes. Rearrangement breakpoints were enriched near open chromatin, androgen receptor and ERG DNA binding sites in the setting of the ETS gene fusion *TMPRSS2-ERG*, but inversely correlated with these regions in tumors lacking ETS fusions. This observation suggests a link between chromatin or transcriptional regulation and the genesis of genomic aberrations. Three tumors contained rearrangements that disrupted *CADM2*, and four harbored events disrupting either *PTEN* (unbalanced events), a prostate tumor suppressor, or *MAGI2* (balanced events), a PTEN interacting protein not previously implicated in prostate tumorigenesis. Thus, genomic rearrangements may arise from transcriptional or chromatin aberrancies to engage prostate tumorigenic mechanisms.

Among men in the United States, prostate cancer accounts for more than 200,000 new cancer cases and 32,000 deaths annually[1]. Although androgen deprivation therapy yields transient efficacy, most patients with metastatic prostate cancer eventually die of their disease. These aspects underscore the critical need to articulate both genetic underpinnings and novel therapeutic targets in prostate cancer.

Recent years have heralded a marked expansion in our understanding of the somatic genetic basis of prostate cancer. Of considerable importance has been the discovery of recurrent gene fusions that render ETS transcription factors under the control of androgen-responsive or other promoters[2–5]. These findings suggest that genomic rearrangements may comprise a

major mechanism driving prostate carcinogenesis. Other types of somatic alterations also engage important mechanisms[6–8]; however, the full spectrum of prostate cancer genomic alterations remains incompletely characterized. Moreover, although the androgen signaling axis represents an important therapeutic focal point[9,10], relatively few additional drug targets have yet been elaborated by genetic studies of prostate cancer[11]. To discover additional genomic alterations that may underpin lethal prostate cancer, we performed paired-end, massively parallel sequencing on tumor and matched normal genomic DNA obtained from seven patients with “high-risk” primary prostate cancer.

Landscape of genomic alterations

All patients harbored tumors of stage T2c or greater, and Gleason grade 7 or higher. Serum prostate-specific antigen (PSA) levels ranged from 2.1–10.2 ng/ml (Supplementary Table 1). Three tumors contained chromosomal rearrangements involving the *TMPRSS2-ERG* loci as determined by fluorescence in situ hybridization (FISH) and RT-PCR[2] (Table 1 and Supplementary Table 1). We obtained approximately 30-fold mean sequence coverage for each sample, and reliably detected somatic mutations in more than 80% of the genome (described in Supplementary Information). Circos plots[12] indicating genomic rearrangements and copy number alterations for each prostate cancer genome are shown in Figure 1.

We identified a median of 3,866 putative somatic base mutations (range: 3,192–5,865) per tumor; the estimated mean mutation frequency was 0.9 per megabase (see Supplementary Methods). This mutation rate is similar to that observed in acute myeloid leukemia and breast cancer[13–16] but 7–15 fold lower than rates reported for small cell lung cancer and melanoma[17–19]. The mutation rate at CpG dinucleotides was more than 10-fold higher than at all other genomic positions (Supplementary Fig. 1). A median of 20 non-synonymous base mutations per sample were called within protein-coding genes (range: 13–43; Supplementary Table 3). We also identified six high-confidence coding indels (4 deletions, 2 insertions) ranging from 1 to 9 base pairs (bp) in length, including a 2bp frameshift insertion in the tumor suppressor gene, *PTEN* (Supplementary Table 4, Supplementary Fig. 2).

Two genes (*SPTA1* and *SPOP*) harbored mutations in 2/7 tumors. *SPTA1* encodes a scaffold protein involved in erythroid cell shape specification, while *SPOP* encodes a modulator of Daxx-mediated ubiquitination and transcriptional regulation[20]. The *SPOP* mutations exceeded the expected background rate in these tumors (Q = 0.055), Moreover, *SPOP* was also found significantly mutated in a separate study of prostate cancer[21]. Interestingly, the chromatin modifiers *CHD1*, *CHD5*, and *HDAC9* were mutated in 3/7 prostate cancers. These genes regulate embryonic stem cell pluripotency, gene regulation, and tumor suppression[22–24]. Members of the HSP-1 stress response complex (*HSPA2*, *HSPA5*, and *HSP90AB1*) were also mutated in 3/7 tumors. The corresponding proteins form a chaperone complex targeted by several anticancer drugs in development[25]. Furthermore, we found the KEGG pathway “Antigen processing and presentation” to be significantly mutated out of 616 diverse gene sets corresponding to gene families and known pathways (Q = 0.0021). This result is intriguing given the clinical benefit associated with immunotherapy for

prostate cancer[26,27]. Other known cancer genes were mutated in single tumors, including *PRKCI* and *DICER*. Thus, some coding mutations may contribute to prostate tumorigenesis and suggest possible therapeutic interventions.

Complex patterns of balanced rearrangements

Given the importance of oncogenic gene fusions in prostate cancer, we next characterized the spectrum of chromosomal rearrangements. We identified a median of 90 rearrangements per genome (range: 43–213) supported by ≥3 distinct read pairs (Supplementary Table 5). This distribution of rearrangements was similar to that previously described for breast cancer[28]. We examined 594 candidate rearrangements by multiplexed PCR followed by massively parallel sequencing, and validated 78% of events by this approach (see Supplementary Methods). Three genes disrupted by rearrangements also harbored non-synonymous mutations in another sample: *ZNF407*, *CHD1*, and *PTEN*. Notably, the chromatin modifier *CHD1*, which contains a validated splice site mutation in prostate tumor PR-1701 (as indicated above), also harbored intragenic breakpoints in two additional samples (PR-0508 and PR-1783). These rearrangements predict truncated proteins, raising the possibility that dysregulated CHD1 may contribute to a block in differentiation in some prostate cancer precursor cells[22].

In 88% of cases, the fusion point could be mapped to base pair resolution (Supplementary Methods). The most common type of fusion involved a precise join, with neither overlapping nor intervening sequence at the rearrangement junction. In a minority of cases, an overlap (microhomology) of 1 base pair (bp) or more was observed. The rearrangement frequency declined by approximately 4-fold for each base of microhomology. This result differed from the patterns seen in breast tumors, in which the most common junction involved a microhomology of 2–3 bp[28]. Thus, mechanisms by which rearrangements are generated may differ between prostate and breast cancer.

Detailed examination of these chromosomal rearrangements revealed a distinctive pattern of balanced breaking and rejoining not previously observed in solid tumors: several genomes contained complex inter- and intra-chromosomal events involving an exchange of “breakpoint arms.” A mix of chimeric chromosomes was thereby generated, without concomitant loss of genetic material (e.g., all breakpoints produced balanced translocations; illustrated conceptually in Fig. 2a).

This “closed chain” pattern of breakage and rejoining was evident in each of the *TMPRSS2-ERG* fusion-positive prostate cancers. In two such cases, both the *TMPRSS2* and *ERG* genomic loci were involved in a closed chain of breakpoints. For example, the *TMPRSS2-ERG* gene fusion in PR-1701 was produced by a closed quartet of balanced translocations on chromosomes 21 and 1 (Fig. 2b). The *TMRPSS2-ERG* gene fusion in PR-0581 occurred within a closed trio of intrachromosomal rearrangements involving *C21ORF45*, *ERG*, and *TMPRSS2* (Supplementary Fig. 3).

One noteworthy closed chain of rearrangements harbored breakpoints situated independently of *TMPRSS2-ERG* (Supplementary Fig. 4) but in close proximity to multiple known cancer genes or orthologues. This chain (found in sample PR-2832) contained breakpoint pairs at

the following loci: (1) 60 bp from exon 6 of TANK binding kinase 1 (*TBK1* or "NF-kB-activating kinase")[29]; (2) within the first intron of *TP53* (7 kb upstream of translation start); (3) 51 kb from *MAP2K4* (a kinase recently shown to induce anchorage-independent growth via mutations[21]); and (4) 3 kb from the *ABL1* protooncogene (Fig. 2c). This striking phenomenon suggests that complex translocations may dysregulate multiple genes in parallel to drive prostate tumorigenesis.

Association of rearrangements and epigenetic marks

The closed chain pattern of chromosomal breakpoints also raised the possibility that multiple genomic regions might become spatially co-localized prior to undergoing rearrangement. Conceivably, such a phenomenon could reflect migration to "transcription factories"—preassembled nuclear subcompartments that contain RNA polymerase II holoenzyme[30]. In prostate cells, androgen signaling has been shown to induce co-localization of *TMPRSS2* and *ERG*, thereby allowing double-strand breaks to facilitate gene fusion formation[31–33]. A role for transcription in the genesis of *TMPRSS2-ERG* in PR-1701 seems plausible, as genomic sequences of up to 240 bp are duplicated at the resulting fusion junctions (Fig. 2b). Alternatively, chains of breakpoints might reflect the clustering of active and inactive chromatin within the recently demonstrated fractal globule structure of nuclear architecture[34]. Stimulated by these models, we considered whether the genomic regions involved in prostate cancer rearrangements exhibited similarities in terms of either transcriptional patterns or chromatin marks. Here, we employed published chromatin immunoprecipitation and massively parallel sequencing (ChIP-seq) data from VCaP, an androgen-sensitive prostate cancer cell line that harbors the *TMPRSS2-ERG* gene fusion[35].

Interestingly, the location of rearrangement breakpoints from the *TMPRSS2-ERG* fusion positive tumor PR-2832 showed significant spatial correlation with various marks of open chromatin in VCaP cells (Fig. 3 and Supplementary Fig. S5). These marks included ChIP-seq peaks corresponding to RNA polymerase II (pol II, $p = 1.0\times 10^{-15}$), histone H3K4 trimethylation (H3K4me3, $p = 3.1\times 10^{-7}$), histone H3K36 trimethylation (H3K36me3, $p = 3.5\times 10^{-12}$), and histone H3 acetylation (H3ace, $p = 9.5 \times 10^{-12}$) (Fig. 3). Similar statistical correlations were observed for peaks corresponding to AR ($p = 1.1\times 10^{-5}$), and ERG binding sites ($p = 4.9\times 10^{-14}$) (Fig. 3 and Supplementary Table 6), consistent with the substantial overlap between AR and ERG binding locations in VCaP cells[35]. (We did not observe significant enrichment of either AR or ERG binding site sequences in the vicinity of these breakpoints.) In the other ERG fusion-positive tumors (PR-0581 or PR-1701), the correlations between breakpoints and ChIP-seq peaks were intermittently apparent, albeit much less significant.

Surprisingly, rearrangement breakpoints from all four ETS fusion-negative tumors were inversely correlated with these same marks of open chromatin and AR/ERG binding (Fig. 3 and Supplementary Fig. S5). In fact, breakpoints from two of four ETS-negative tumors were significantly correlated with marks of histone H3K27 trimethylation (H3K27me3) in VCaP cells, which denote inactive chromatin and transcriptional repression (Fig. 3). This result suggested that somatic rearrangements might occur within closed chromatin in some tumor cells, or that the epigenetic architecture or transcriptional program of some

TMPRSS2-ERG fusion positive cells differs markedly from that of ERG fusion-negative cells. In support of the former, we observed a similar enrichment of PR-2832 rearrangements and depletion of fusion-negative rearrangements near marks of active transcription profiled in several additional cell lines, including fusion-negative prostate cancer cell lines LNCaP and PC-3 as well as three cell lines derived from non-prostate lineages (Supplementary Fig. S5)[35–37].

Based on these intriguing results, we performed similar analyses comparing the chromatin state in VCaP cells to rearrangement patterns of other cancer types. No statistically significant correlations or inverse correlations were observed between VCaP ChIP-seq data and rearrangement breakpoints obtained from a melanoma cell line[18], a small-cell lung cancer cell line[17], or a primary non-small cell lung tumor[38] (Supplementary Fig. S5 and Supplementary Table 6). However, rearrangements from 16 of 18 breast tumors and cell lines examined[28] exhibited a pattern of association similar to that observed in prostate tumor PR-2832 (Supplementary Fig. S6). Notably, breakpoints in these tumors were also strongly associated with estrogen receptor (ER) binding sites derived from the breast cancer cell line MCF-7[39]. Furthermore, we observed a strong association between ER ChIP-seq peaks from MCF-7 and all VCaP ChIP-seq peaks corresponding to open chromatin, AR, and ERG binding ($p < 10^{-90}$; Supplementary Fig. S6). Thus, patterns of open chromatin may be highly overlapping in some hormone-driven cancer cells. Such regions may correlate significantly with sites of somatic rearrangement in cancers of the prostate, breast, and possibly other tissues.

To examine whether processes linked to chromatin reorganization and DNA rearrangement are also associated with increased mutation frequency, we tested for enrichment of point mutations near regions of ChIP-seq peaks and rearrangement breakpoints. We observed a significantly reduced prevalence of point mutations near marks of VCaP active transcription —and slight enrichment of mutations in closed chromatin—in all 7 prostate tumors (Supplementary Fig. S7). This pattern is consistent with both negative selection and transcription-coupled DNA repair. Additionally, we observed a significant enrichment of mutations near rearrangement breakpoints in 5 of 7 prostate tumors (Supplementary Fig. S7). Although the increased rate of mutations near rearrangements may conceivably reflect activation-induced cytodine deaminase (AID) in the double strand break repair process[31,40], we did not observe a significant overrepresentation of any one class of mutation among those located near breakpoints.

Recurrent rearrangements involving *CADM2*

Sixteen genes harbored a somatic rearrangement in at least 2 prostate tumors (Supplementary Table 7), and four contained rearrangements in 3 of 7 tumors. In addition to *TMPRSS2* and *ERG*, the latter included *CSMD3* and *CADM2*. These genes were rearranged at a frequency beyond that expected by chance, even after correcting for gene size (Supplementary Table 8). *CSMD3* encodes a giant gene that contains multiple CUB and sushi repeats. However, we did not observe additional *CSMD3* rearrangements by fluorescence in situ hybridization (FISH) in an independent analysis of 94 prostate tumors (Supplementary Fig. S8).

CADM2 encodes a nectin-like member of the immunoglobulin-like cell adhesion molecules. Several nectin-like proteins exhibit tumor suppressor properties in various contexts. Analysis of SNP array-derived copy number profiles of tumors and cell lines[41,42] suggests that *CADM2* does not reside near a fragile site (Supplementary Fig. S9). At the same time, the complexity of *CADM2* rearrangements (Fig. 4a) suggested that a simple FISH validation approach might prove insufficient to determine the overall frequency of *CADM2* disruption. Nevertheless, we screened an independent cohort of 90 additional prostate tumors using a "break-apart" FISH assay designed to query the *CADM2* locus (Supplementary Fig. S8). *CADM2* aberrations were detected in 6/90 samples (5 rearrangements and 1 copy gain; Fig. 4b). These results confirmed that *CADM2* is recurrently disrupted in prostate cancer, although they likely represent a lower bound for the true prevalence of *CADM2* alteration in this malignancy.

Rearrangements disrupting *PTEN* and *MAGI2*

Two prostate tumors contained breakpoints within the *PTEN* tumor suppressor gene[6] (Fig. 4c). In both cases, the rearrangements generated heterozygous deletions that were confirmed by FISH analysis (Supplementary Fig. S10). In one tumor (PR-0581), *PTEN* rearrangement co-occurred with a dinucleotide insertion within the *PTEN* coding sequence (described above).

Two additional tumors harbored rearrangements disrupting the *MAGI2* gene, which encodes a PTEN-interacting protein[43,44] (Fig. 4c). In one tumor (PR-0508), two independent but closely aligned inversion events (marking both ends of a 450-kilobase inverted sequence) affected the *MAGI2* locus. In the other tumor (PR-2832), two long-range intrachromosomal inversions were observed, raising the possibility of heterogeneous subclones harboring independent *MAGI2* rearrangements. Thus, 4 of 7 tumors harbored rearrangements predicted to inactivate *PTEN* or *MAGI2*, including all three tumors harboring *TMPRSS2-ERG* rearrangements. Although a tumor suppressor function for *MAGI2* has not been established previously, this gene was recently shown to undergo rearrangement in the genome of a melanoma cell line[18], another tumor type in which PTEN loss is prevalent. In principle, genomic rearrangements that subvert PTEN function either directly or indirectly (e.g., through loss of *MAGI2*) might dysregulate the PI3 kinase pathway in prostate cancer.

Whereas both *PTEN* rearrangements involved chromosomal copy loss, the *MAGI2* rearrangements were balanced events (Supplementary Fig. S11). Like *CSMD3* and *CADM2*, *MAGI2* does not appear to reside near a fragile site (Supplementary Fig. S9). We screened 88 independent prostate tumors using FISH inversion probes and identified 3 additional samples harboring similar inversions, each of which was wild type for PTEN disruption (Fig. 4d and Supplementary Fig. S8). As with *CADM2* above, these FISH findings may underestimate the true frequency of *MAGI2* disruption in prostate cancer.

We further analyzed the *PTEN* and *MAGI2* loci using high-density SNP arrays obtained from 66 primary prostate cancers. As shown in Supplementary Figure S11, focal somatic deletions affecting the *PTEN* locus were commonly observed in these tumors, as expected. Interestingly, no SCNAs were observed at the *MAGI2* locus in either prostate tumor found to

contain *MAGI2* rearrangements by genome sequencing (Supplementary Fig. S11). Conceivably, this region may also harbor genes whose loss would be deleterious to prostate cancer cells. More generally, these findings suggest that extensive shotgun paired-end sequencing (as opposed to lower-resolution approaches) may be required to elaborate the compendium of genes targeted by somatic alterations in prostate cancer.

Discussion

This study represents the first whole genome sequencing analysis of human prostate cancer. Systematic genome characterization efforts have often focused primarily on gene-coding regions to identify "driver" or "druggable" alterations[45–47]. In contrast, the high prevalence of recurrent gene fusions has highlighted chromosomal rearrangements as critical initiating events in prostate cancer[2,3]. Genome sequencing data indicate that complex rearrangements may enact pivotal gain- and loss-of-function driver events in primary prostate carcinogenesis. Moreover, many rearrangements may occur preferentially in genes that are spatially localized together with transcriptional or chromatin compartments, perhaps initiated by DNA strand breaks and erroneous repair. The complexity of "closed chain" and other rearrangements suggests that complete genome sequencing—as opposed to approaches focused on exons or gene fusions—may be required to elaborate the spectrum of mechanisms directing prostate cancer genesis and progression.

A positive correlation exists between the location of breakpoints in ERG-positive tumor cells and open chromatin in VCaP cells, and also between breakpoints present in ERG-negative cells and VCaP regions of closed chromatin. This suggests that breakpoints may preferentially occur within regions of open chromatin in some ERG-positive tumor cells while raising alternate possibilities for the genesis of breakpoints in ERG-negative cells. Conceivably, somatic rearrangements may occur within regions of closed chromatin in ERG-negative tumor cells. Alternately, ERG-negative tumor cells may have distinct transcriptional or chromatin patterns, with many regions that are closed in VCaP being open in these cells. Clustering of breakpoints within active regions might also reflect selection for functionally consequential rearrangements during tumorigenesis. The relative contribution of these aspects to tumorigenesis will likely be informed by additional integrative analyses of epigenetic and structural genomic datasets across many tumor types.

Previous studies of genetically engineered mouse models have shown that the combination of ERG dysregulation and PTEN loss triggers the formation of aggressive prostate tumors[48,49]. This same combination identifies a subtype of human prostate cancer characterized by poor prognosis[50]. The discovery of *MAGI2* genomic rearrangements in prostate cancer suggests that interrogating both the *PTEN* and *MAGI2* loci might improve prognostication and patient stratification for clinical trials of PI3 kinase pathway inhibitors. Additional mutated genes discovered in this study also suggest interesting therapeutic avenues. For example, the presence of point mutations involving chromatin modifying genes and the HSP70/HSP90 chaperone complex raises the possibility that these cellular processes may represent targetable dependencies in some prostate tumors. Overall, complete genome sequencing of large numbers of relapsing primary and metastatic prostate cancers promises

to define a genetic cartography that assists in tumor classification, defines mechanisms of carcinogenesis, and identifies new targets for therapeutic intervention.

Methods Summary

The complete genomes of seven prostate tumors and patient-matched normal samples were sequenced to approximately 30-fold haploid coverage on an Illumina GA II sequencer. DNA was extracted from patient blood and from tumors following radical prostatectomy, and was subjected to extensive quality control procedures to monitor DNA structural integrity, genotype concordance, and tumor purity and ploidy. Standard paired-end libraries (~400bp inserts) were sequenced as 101bp paired-end reads. Raw sequencing data were processed by Illumina software and passed to the Picard pipeline, which produced a single BAM file for each sample storing all reads with well-calibrated quality scores together with their alignments to the reference genome. BAM files for each tumor/normal sample pair were analyzed by the Firehose pipeline to characterize the full spectrum of somatic mutations in each tumor, including base pair substitutions, short insertions and deletions, and large-scale structural rearrangements. A subset of base pair mutations and rearrangements were validated using independent technologies in order to assess the specificity of the detection algorithms. Fluorescence in situ hybridization (FISH) was also performed for selected recurrent rearrangements. The locations of all rearrangement breakpoints were compared to previously published chromatin immunoprecipitation (ChIP) binding peaks from related cell types to test for global associations between rearrangements and a range of epigenetic marks.

A complete description of the materials and methods is provided in the Supplementary Information. All Illumina sequence data have been deposited in dbGaP (http://www.ncbi.nlm.nih.gov/gap) and are available at accession phs000330.v1.p1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Robert Leung and all members of the Broad Institute Sequencing Platform. This work was supported by the Prostate Cancer Foundation/Movember (T.R.G., M.A.R., L.A.G.), the Howard Hughes Medical Institute (T.R.G.), the National Human Genome Research Institute (S.B.G., E.S.L.), the Kohlberg Foundation (P.W.K., L.A.G.), the National Cancer Institute (F.D., M.A.R., M.M.), the National Institutes of Health (L.A.G.), the Department of Defense (F.D.), the Dana-Farber/Harvard Cancer Center Prostate Cancer SPORE grant 2 P50 CA090381-11, and the Starr Cancer Consortium (M.F.B., F.D., M.A.R., L.A.G.).

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 2010; 60:277–300. [PubMed: 20610543]
2. Tomlins SA, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005; 310:644–648. [PubMed: 16254181]
3. Tomlins SA, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007; 448:595–599. [PubMed: 17671502]

4. Helgeson BE, et al. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. Cancer Res. 2008; 68:73–80. [PubMed: 18172298]
5. Tomlins SA, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res. 2006; 66:3396–3400. [PubMed: 16585160]
6. Li J, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science. 1997; 275:1943–1947. [PubMed: 9072974]
7. Visakorpi T, et al. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet. 1995; 9:401–406. [PubMed: 7795646]
8. Taylor BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11–22. [PubMed: 20579941]
9. Tran C, et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. Science. 2009; 324:787–790. [PubMed: 19359544]
10. Attard G, et al. Phase I clinical trial of a selective inhibitor of CYP17, abiraterone acetate, confirms that castration-resistant prostate cancer commonly remains hormone driven. J Clin Oncol. 2008; 26:4563–4571. [PubMed: 18645193]
11. Palanisamy N, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med. 2010; 16:793–798. [PubMed: 20526349]
12. Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645. [PubMed: 19541911]
13. Ley TJ, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008; 456:66–72. [PubMed: 18987736]
14. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009; 461:809–813. [PubMed: 19812674]
15. Mardis ER, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009; 361:1058–1066. [PubMed: 19657110]
16. Ding L, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010; 464:999–1005. [PubMed: 20393555]
17. Pleasance ED, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–190. [PubMed: 20016488]
18. Pleasance ED, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010; 463:191–196. [PubMed: 20016485]
19. Berger MF, et al. Integrative analysis of the melanoma transcriptome. Genome Res. 2010; 20:413–427. [PubMed: 20179022]
20. Kwon JE, et al. BTB domain-containing speckle-type POZ protein (SPOP) serves as an adaptor of Daxx for ubiquitination by Cul3-based ubiquitin ligase. J Biol Chem. 2006; 281:12664–12672. [PubMed: 16524876]
21. Kan Z, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010; 466:869–873. [PubMed: 20668451]
22. Gaspar-Maia A, et al. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature. 2009; 460:863–868. [PubMed: 19587682]
23. Zhang CL, McKinsey TA, Olson EN. Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation. Mol Cell Biol. 2002; 22:7302–7312. [PubMed: 12242305]
24. Bagchi A, et al. CHD5 is a tumor suppressor at human 1p36. Cell. 2007; 128:459–475. [PubMed: 17289567]
25. Pearl LH, Prodromou C, Workman P. The Hsp90 molecular chaperone: an open and shut case for treatment. Biochem J. 2008; 410:439–453. [PubMed: 18290764]
26. Kantoff PW, et al. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. N Engl J Med. 2010; 363:411–422. [PubMed: 20818862]
27. Kantoff PW, et al. Overall survival analysis of a phase II randomized controlled trial of a Poxviral-based PSA-targeted immunotherapy in metastatic castration-resistant prostate cancer. J Clin Oncol. 2010; 28:1099–1105. [PubMed: 20100959]

28. Stephens PJ, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009; 462:1005–1010. [PubMed: 20033038]
29. Barbie DA, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009; 462:108–112. [PubMed: 19847166]
30. Osborne CS, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet. 2004; 36:1065–1071. [PubMed: 15361872]
31. Lin C, et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell. 2009; 139:1069–1083. [PubMed: 19962179]
32. Mani RS, et al. Induced chromosomal proximity and gene fusions in prostate cancer. Science. 2009; 326:1230. [PubMed: 19933109]
33. Haffner MC, et al. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. Nat Genet. 2010; 42:668–675. [PubMed: 20601956]
34. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]
35. Yu J, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer Cell. 2010; 17:443–454. [PubMed: 20478527]
36. Lin B, et al. Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor. PLoS One. 2009; 4:e6589. [PubMed: 19668381]
37. Birney E, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816. [PubMed: 17571346]
38. Lee W, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature. 2010; 465:473–477. [PubMed: 20505728]
39. Krum SA, et al. Unique ERalpha cistromes control cell type-specific gene regulation. Mol Endocrinol. 2008; 22:2393–2406. [PubMed: 18818283]
40. Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA. Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast Saccharomyces cerevisiae. PLoS Genet. 2008; 4:e1000264. [PubMed: 19023402]
41. Beroukhim R, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905. [PubMed: 20164920]
42. Bignell GR, et al. Signatures of mutation and selection in the cancer genome. Nature. 2010; 463:893–898. [PubMed: 20164919]
43. Wu X, et al. Evidence for regulation of the PTEN tumor suppressor by a membrane-localized multi-PDZ domain containing scaffold protein MAGI-2. Proc Natl Acad Sci U S A. 2000; 97:4233–4238. [PubMed: 10760291]
44. Vazquez F, et al. Phosphorylation of the PTEN tail acts as an inhibitory switch by preventing its recruitment into a protein complex. J Biol Chem. 2001; 276:48627–48630. [PubMed: 11707428]
45. Sjoblom T, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006; 314:268–274. [PubMed: 16959974]
46. Parsons DW, et al. An integrated genomic analysis of human glioblastoma multiforme. Science. 2008; 321:1807–1812. [PubMed: 18772396]
47. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]
48. Carver BS, et al. Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. Nat Genet. 2009; 41:619–624. [PubMed: 19396168]
49. King JC, et al. Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. Nat Genet. 2009; 41:524–526. [PubMed: 19396167]
50. Han B, et al. Fluorescence in situ hybridization study shows association of PTEN deletion with ERG rearrangement during prostate cancer progression. Mod Pathol. 2009; 22:1083–1093. [PubMed: 19407851]

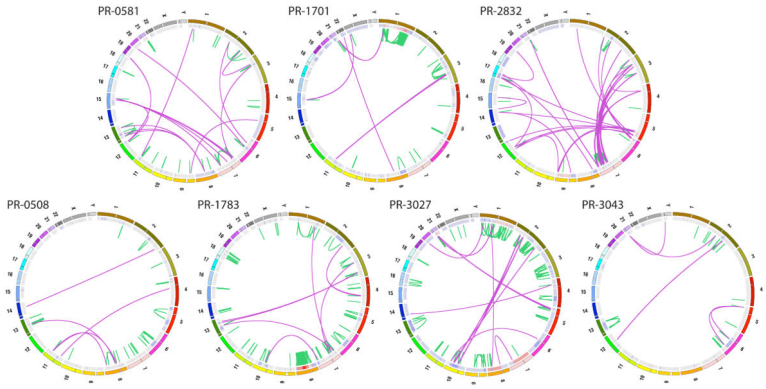

Figure 1.

Graphical representation of 7 prostate cancer genomes. Each Circos plot[12] depicts the genomic location in the outer ring and chromosomal copy number in the inner ring (red = copy gain; blue = copy loss). Interchromosomal translocations and intrachromosomal rearrangements are shown in purple and green, respectively. Genomes are organized according to the presence (top row) or absence (bottom row) of the *TMPRSS2-ERG* gene fusion.

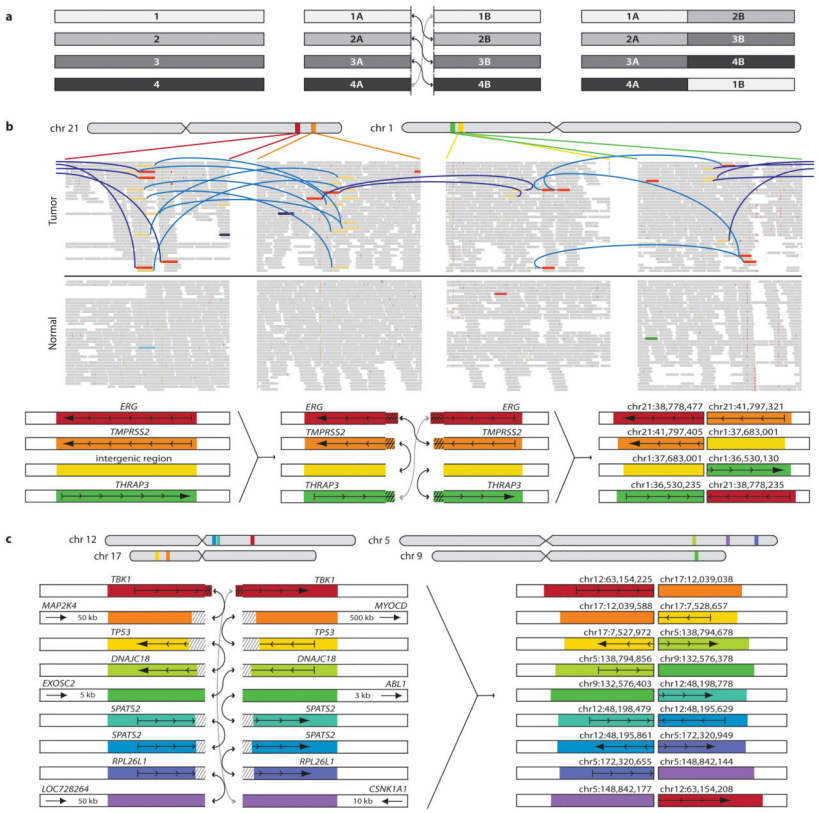

Figure 2.

Complex structural rearrangements in prostate cancer. **(a)** Schematic of "closed chain" pattern of chromosomal breakage and rejoining. Breaks are induced in a set of loci (left), followed by an exchange of free ends without loss of chromosomal material (middle), leading to the observed pattern of balanced (copy neutral) translocations involving a closed set of breakpoints (right). **(b)** Complex rearrangement in prostate PR-1701. *TMPRSS2-ERG* is produced by a closed quartet of balanced rearrangements involving 4 loci on chromosomes 1 and 21. *Top*: Each rearrangement is supported by the presence of discordant read pairs in the tumor genome but not the normal genome (colored bars connected by blue lines). Thin bars represent sequence reads; directionality represents mapping orientation on the reference genome. Figures are based on the Integrative Genomics Viewer (http://www.broadinstitute.org/igv). *Bottom*: Schematic of breakpoints and balanced translocations. Hatched lines indicate sequences that are duplicated in the derived chromosomes at the resulting fusion junctions. **(c)** Complex rearrangement in prostate PR-2832 involving breakpoints and fusions at 9 distinct genomic loci. Hatched lines indicate sequences that are duplicated or deleted in the derived chromosomes at the resulting fusion junctions. For breakpoints in intergenic regions, the nearest gene in each direction is shown. In addition to the sheer number of regions involved, this complex rearrangement is notable for the abundance of breakpoints in or near cancer related genes, such as *TBK1*, *MAP2K4*, *TP53*, and *ABL1*.

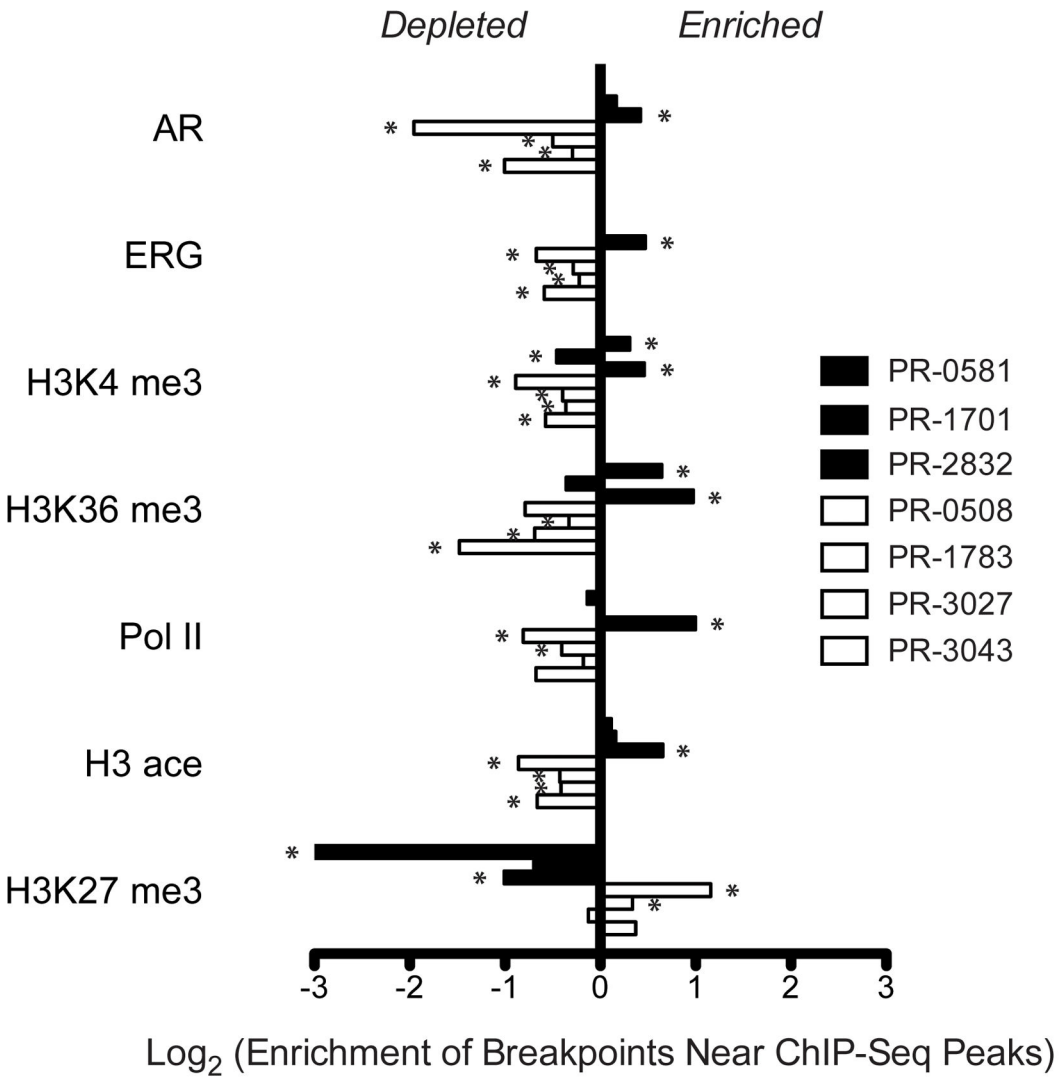

Figure 3.

Association between rearrangement breakpoints and genome-wide transcriptional/histone marks in prostate cancer. ChIP-Seq binding peaks were defined previously for the *TMPRSS2-ERG* positive (ERG+) prostate cancer cell line VCaP[35]. For each genome, enrichment of breakpoints within 50 kb of each set of binding peaks was determined relative to a coverage-matched simulated background (see Methods). ERG+ prostate tumors are in black; ETS-negative prostate tumors are in white. Enrichment is displayed as the ratio of the observed breakpoint rate to the background rate near each indicated set of ChIP-Seq peaks. Rearrangements in ETS-negative tumors are depleted near marks of active transcription (AR, ERG, H3K4me3, H3K36me3, Pol II, and H3ace) and enriched near marks of closed chromatin (H3K27me3). P-values were calculated according to the binomial distribution and are displayed in Supplementary Figure S5 and Supplementary Table 6. Significant associations passing a false discovery rate cutoff of 5% are marked with an asterisk.

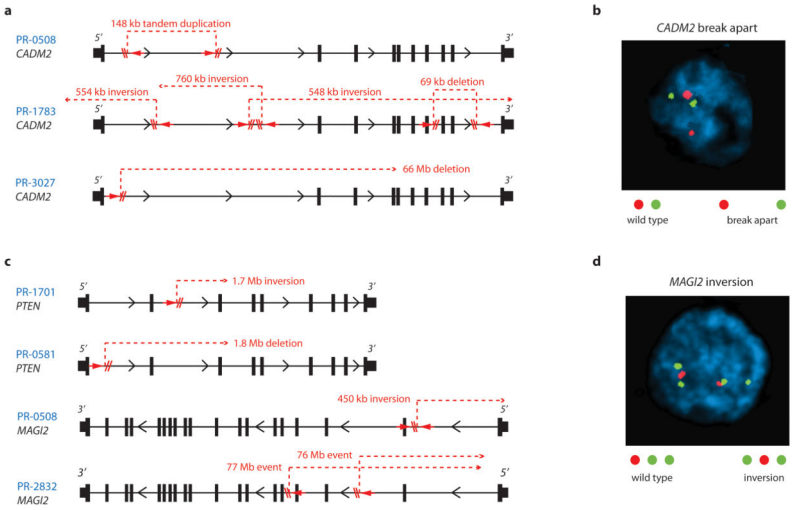

Figure 4.

Disruption of *CADM2* and the *PTEN* pathway by rearrangements. **(a)** Location of intragenic breakpoints in *CADM2*. **(b)** *CADM2* break-apart demonstrated by FISH in an independent prostate tumor. **(c)** Location of intragenic breakpoints in *PTEN* (top) and *MAGI2* (bottom). **(d)** *MAGI2* inversion demonstrated by FISH in an independent prostate tumor, using probes flanking *MAGI2* (red and green) and an external reference probe also on chromosome 7q (green). The probes and strategy for detecting novel rearrangements by FISH are diagrammed in Supplementary Figure S8.

Table 1

Landscape of Somatic Alterations in Primary Human Prostate Cancers

	PR-0508	**PR-0581***	**PR-1701***	**PR-1783**	**PR-2832***	**PR-3027**	**PR-3043**
Tumor Bases Sequenced	97.8×10^9	93.9×10^9	110×10^9	90.9×10^9	106×10^9	93.6×10^9	94.9×10^9
Normal Bases Sequenced	96.7×10^9	57.8×10^9	108×10^9	92.3×10^9	103×10^9	87.8×10^9	96.6×10^9
Tumor Haploid Coverage	31.8	30.5	35.8	29.5	34.4	30.4	30.8
Normal Haploid Coverage	31.4	18.8	34.9	30.0	33.4	28.5	31.4
Callable Fraction	0.84	0.83	0.87	0.82	0.84	0.84	0.85
Estimated Tumor Purity**	0.73	0.60	0.49	0.75	0.59	0.74	0.68
All Point Mutations (high confidence)	3898 (1447)	3829 (1430)	3866 (1936)	4503 (2227)	3465 (1831)	5865 (2452)	3192 (1713)
Nonsilent Coding Mutations (high confidence)	16 (5)	20 (3)	24 (9)	32 (20)	13 (7)	43 (16)	14 (10)
Mutation Rate	0.7 per Mb	0.7 per Mb	0.8 per Mb	1.0 per Mb	0.8 per Mb	1.2 per Mb	0.7 per Mb
Rearrangements	53	67	90	213	133	156	43

* harbors *TMPRSS2-ERG* gene fusion

** estimated from SNP array-derived allele specific copy number levels using the ABSOLUTE algorithm (see Supplementary Methods).