

# Integrative Genomic Analyses Reveal an Androgen-Driven Somatic Alteration Landscape in Early-Onset Prostate Cancer

Joachim Weischenfeldt,<sup>1,14</sup> Ronald Simon,<sup>2,14</sup> Lars Feuerbach,<sup>4,14</sup> Karin Schlangen,<sup>8,14</sup> Dieter Weichenhan,<sup>5,14</sup> Sarah Minner,<sup>2,14</sup> Daniela Wuttig,<sup>6,14</sup> Hans-Jörg Warnatz,<sup>8</sup> Henning Stehr,<sup>8</sup> Tobias Rausch,<sup>1</sup> Natalie Jäger,<sup>4</sup> Lei Gu,<sup>5</sup> Olga Bogatyrova,<sup>5</sup> Adrian M. Stütz,<sup>1</sup> Rainer Claus,<sup>5</sup> Jürgen Eils,<sup>4</sup> Roland Eils,<sup>4,12</sup> Clarissa Gerhäuser,<sup>5</sup> Po-Hsien Huang,<sup>5</sup> Barbara Hutter,<sup>4</sup> Rolf Kabbe,<sup>4</sup> Christian Lawrenz,<sup>4</sup> Sylwester Radomski,<sup>4</sup> Cynthia C. Bartholomae,<sup>7</sup> Maria Fälth,<sup>6</sup> Stephan Gade,<sup>6</sup> Manfred Schmidt,<sup>7</sup> Nina Amschler,<sup>2</sup> Thomas Haß,<sup>2</sup> Rami Galal,<sup>2</sup> Jovisa Gjoni,<sup>2</sup> Ruprecht Kuner,<sup>6</sup> Constance Baer,<sup>5</sup> Sawinee Masser,<sup>2</sup> Christof von Kalle,<sup>7</sup> Thomas Zichner,<sup>1</sup> Vladimir Benes,<sup>13</sup> Benjamin Raeder,<sup>1</sup> Malte Mader,<sup>9</sup> Vyacheslav Amstislavskiy,<sup>8</sup> Meryem Avci,<sup>7</sup> Hans Lehrach,<sup>8</sup> Dmitri Parkhomchuk,<sup>8</sup> Marc Sultan,<sup>8</sup> Lia Burkhardt,<sup>2</sup> Markus Graefen,<sup>3</sup> Hartwig Huland,<sup>3</sup> Martina Kluth,<sup>2</sup> Antje Krohn,<sup>2</sup> Hüseyin Sirma,<sup>2</sup> Laura Stumm,<sup>2</sup> Stefan Steurer,<sup>2</sup> Katharina Grupp,<sup>2</sup> Holger Sülthmann,<sup>6,15</sup> Guido Sauter,<sup>2,15</sup> Christoph Plass,<sup>5,15</sup> Benedikt Brors,<sup>4,15</sup> Marie-Laure Yaspo,<sup>8,15</sup> Jan O. Korbel,<sup>1,10,15,\*</sup> and Thorsten Schlömm<sup>3,11,15</sup>

<sup>1</sup>Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Meyerhofstr. 1, 69117 Heidelberg, Germany

<sup>2</sup>Institute of Pathology

<sup>3</sup>Martini-Clinic, Prostate Cancer Center

University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

<sup>4</sup>Division of Theoretical Bioinformatics

<sup>5</sup>Division of Epigenomics and Cancer Risk Factors

German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

<sup>6</sup>Working Group Cancer Genome Research

<sup>7</sup>Division of Translational Oncology

German Cancer Research Center and National Center of Tumor Diseases, Im Neuenheimer Feld 460, 69120 Heidelberg, Germany

<sup>8</sup>Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63, 14195 Berlin, Germany

<sup>9</sup>Hamburg University, Centre for Bioinformatics, Bundesstr. 43, 20146 Hamburg, Germany

<sup>10</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

<sup>11</sup>Department of Urology, Section for Translational Prostate Cancer Research, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

<sup>12</sup>Institute of Pharmacy and Molecular Biotechnology, and Bioquant Center, University of Heidelberg, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

<sup>13</sup>Genomics Core facility, European Molecular Biology Laboratory (EMBL), Meyerhofstr. 1, 69117 Heidelberg, Germany

<sup>14</sup>These authors contributed equally to this work

<sup>15</sup>These authors contributed equally to this work as senior authors

\*Correspondence: [jan.korbel@embl.de](mailto:jan.korbel@embl.de)

<http://dx.doi.org/10.1016/j.ccr.2013.01.002>

## SUMMARY

Early-onset prostate cancer (EO-PCA) represents the earliest clinical manifestation of prostate cancer. To compare the genomic alteration landscapes of EO-PCA with “classical” (elderly-onset) PCA, we performed deep sequencing-based genomics analyses in 11 tumors diagnosed at young age, and pursued comparative assessments with seven elderly-onset PCA genomes. Remarkable age-related differences in structural rearrangement (SR) formation became evident, suggesting distinct disease pathomechanisms. Whereas

### Significance

It is presently unknown whether genetic or mechanistic differences distinguish “classical” elderly-onset PCA from EO-PCA. Using integrative high-throughput sequencing approaches, combined with validations in a large-scale patient cohort, we show that EO-PCA formation involves a characteristic pathomechanism associated with the specific emergence of androgen-driven SRs. By comparison, elderly-onset PCAs accumulate nonandrogen-associated SRs, indicating that a different tumor formation mechanism is operating in these. This work reveals a striking age-dependent SR associated disease mechanism in a common human cancer, and presents a significant advancement in understanding age-dependencies of PCA initiation and progression, with implications for clinical management. The genomics data from our consortium further provide a valuable compendium of massively-parallel DNA sequencing data in EO-PCA.

EO-PCAs harbored a prevalence of balanced SRs, with a specific abundance of androgen-regulated ETS gene fusions including *TMPRSS2:ERG*, elderly-onset PCAs displayed primarily non-androgen-associated SRs. Data from a validation cohort of > 10,000 patients showed age-dependent androgen receptor levels and a prevalence of SRs affecting androgen-regulated genes, further substantiating the activity of a characteristic “androgen-type” pathomechanism in EO-PCA.

## INTRODUCTION

Prostate cancer (PCA) is the most common cancer in Western countries and the second most lethal cancer in men (Siegel et al., 2012). The incidence of PCA increases with age, with a median age at diagnosis of ~70 years (Grönberg, 2003). Due to relatively slow disease progression when compared to life expectancy, patients with PCA do not always require definite therapy, with active surveillance representing an established treatment option (Heidenreich et al., 2011). A relevant subset of PCA, however, is diagnosed early in life, referred to as early-onset PCA (EO-PCA), with ~2% of all tumors detected in men 50 years of age or younger (Figure 1). EO-PCAs are of substantial clinical relevance, requiring obligatory definite treatment (Heidenreich et al., 2011). This is due to the high life expectancy of these young patients, with a higher risk of dying of disease (Albertsen et al., 2005; Parker et al., 2006), although there are also studies showing a more pronounced risk of rapid death in EO-PCA (Lin et al., 2009b; Ryan et al., 2007) compared to the “classical” cases of PCA diagnosed in 60- to 80-year-old men (herein referred to as “elderly-onset PCA”).

Initial genomics surveys focused on classical, elderly-onset PCA have revealed a substantial diversity in somatic single nucleotide variants (SNVs) (Berger et al., 2011; Kumar et al., 2011), and identified recurrent SRs, including oncogene-activating gene fusion events (Berger et al., 2011; Bastus et al., 2010; Lin et al., 2009a; Mani et al., 2009; Taylor et al., 2010; Tomlins et al., 2005) in elderly-onset PCAs. Despite the recent progress in identifying molecular drivers of PCA with massively parallel sequencing of cancer genomes (Berger et al., 2011) and exomes (Kumar et al., 2011; Barbieri et al., 2012; Grasso et al., 2012), so far no study has generated whole-genome sequencing data of EO-PCAs. Hence, it has therefore remained unclear whether EO-PCA is characterized by a distinctive spectrum of driver mutations or a specific pathomechanism. A better understanding of genomic alterations associated with EO-PCA will facilitate the understanding of molecular defects leading to early disease onset and foster the development of new diagnostic, prognostic, therapeutic, and prevention strategies.

On behalf of the International Cancer Genome Consortium (ICGC) project on Early-Onset Prostate Cancer (<http://www.icgc.org>), we carried out integrated genomic analyses, including whole-genome, transcriptome, and DNA methylome sequencing (Campbell et al., 2008) in 11 patients with EO-PCA. We used these genomics data, together with a large-scale tissue microarray (Schlomm et al., 2008) (TMA)-based validation platform, to pinpoint molecular features linked with early disease occurrence.

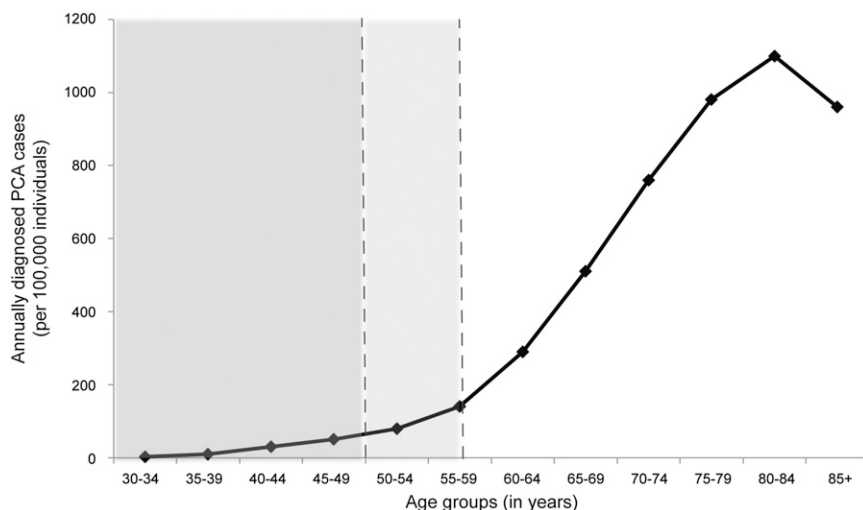
## RESULTS

### Integrative High-Throughput Sequencing and Variant Calling

Tumor and paired control tissue samples (lymphocytes) were obtained from all 11 patients with EO-PCA who provided informed consent. The patient mean age at surgery was 47 years, tumor Gleason grades ranged from 3+4 to 5+4, and tumor stages from pT2c to pT3b (Table 1). Our multitiered genomics approach (see Experimental Procedures) involved ~30- to 40-fold whole-genome sequencing (Table 2; Figure S1A available online) of tumor and control samples, with two sequencing libraries per sample displaying complementary paired read insert sizes (ie, ~200 bp paired-end, and ~4.5 kb mate-pair inserts) (Ley et al., 2008). We furthermore carried out transcriptome sequencing (RNA-seq) (Sultan et al., 2008) of each tumor in comparison to a normal prostate tissue control, using two complementary library generation approaches geared toward assessing mRNAs and MicroRNAs (miRNAs). In addition, we identified altered DNA methylation patterns by performing an enrichment of methylated DNA fragments with an Fc-coupled MBD2 protein (Gebhard et al., 2006) followed by massively parallel sequencing (Figures S1B and S1C; see Experimental Procedures).

Following the mapping of DNA reads onto the human reference assembly we applied complementary computational approaches for the detection of somatic genome alterations (see Experimental Procedures). We identified SNVs, as well as short deletions and insertions (short InDels) by directly evaluating read alignments onto the reference genome assembly (Li and Durbin, 2009; DePristo et al., 2011). Furthermore, we detected large-scale somatic structural rearrangements (SRs) by evaluating the relative depth-of-coverage of DNA reads along chromosomes, by scanning DNA reads for split or clipped read alignments, and by identifying read-pairs that abnormally mapped onto the reference (Xi et al., 2011).

Altogether, we identified 931–5,696 somatic SNVs per patient (Tables 2 and S1; Figure S1A), with the C > T nucleotide transition being the most common substitution (42.6% ± 5.6%), similar to what has been observed in elderly-onset PCAs (Berger et al., 2011; Kumar et al., 2011). On average, 16 nonsynonymous somatic SNVs fell into protein-coding genes (range 3–55), affecting altogether 175 genes. We further detected six nonsynonymous rare germline SNVs with a second somatic mutational event, leading to compound heterozygosity (Table S1; Supplemental Experimental Procedures). A low number (four) of high-confidence somatic InDels of ≤50 bp in size was identified (Table S2). By comparison, we detected a considerably larger number—20–90 per EO-PCA genome (average, 45)—of somatic SRs (Tables 2 and S2; Figures 2A and S2A). These somatic variants included copy-number alterations (average, 23), and



**Figure 1. The Age Distribution in Elderly-Onset PCA and EO-PCA**

Dark and light areas depict patients younger than 50 years (EO-PCA) and patients between 50 and 60 years old, respectively (reproduced from epidemiologic data; Grönberg, 2003).

also a relatively large number of balanced rearrangements (average, 22), most of which were translocations. PCR followed by capillary sequencing verified 85 of 95 (89%) SNVs, all four short somatic InDels, and 50 of 53 (94%) of the SRs (Table S3).

### Computational Inference and Characterization of Candidate Driver Somatic Alterations

To assess the functional relevance of these somatic alterations, we first analyzed gene-altering events in detail. Because a relevant portion of the somatic alterations identified may constitute passenger alterations (Stratton et al., 2009), we performed integrative analyses to infer candidate driver alterations in our EO-PCA samples. This involved using our somatic SNV and SR data to identify genes recurrently altered across samples, and applying a “two-hit hypothesis” approach for the identification of candidate tumor suppressor genes (Knudson, 1971), by probing for genes for which both alleles were affected by genetic

mechanisms of inactivation (see Experimental Procedures). Our search strategy resulted in the identification of 23 genes that were disrupted or mutated in at least two additional PCA patient cohorts (Barbieri et al., 2012; Berger et al., 2011; Grasso et al., 2012). We further inferred 76 genes harboring two genomic hits, a list comprising several previously described tumor suppressors including

*CDH1*, *TP53*, *PTEN*, and *NCOR2*, many of which also showed promoter hypermethylation in samples not carrying the respective genetic lesion (Table S4).

*NCOR2*, a transcriptional corepressor interacting with the androgen receptor (AR), was previously reported as mutated in PCA (Taylor et al., 2010). In addition to biallelic genetic alterations affecting *NCOR2* in one tumor (EOPC-06), we observed homozygous loss of an *NCOR2* regulator in another tumor (EOPC-04) in conjunction with *NCOR2* downregulation. We further identified high levels of DNA methylation of the *NCOR2* promoter in association with *NCOR2* downregulation in a third tumor (EOPC-01). The clinical relevance of *NCOR2* deletions was evaluated with our large-scale TMA resource (Schlomm et al., 2008) revealing *NCOR2* deletions in 3.2% (163/5100) PCAs, and showing significant association with biochemical disease recurrence ( $p = 0.039$ , Likelihood ratio test; Figure S2B).

A pronounced diversity of inactivating genetic mechanisms was also observed for *PTEN*, in agreement with a recent study of elderly-onset PCAs (Reid et al., 2012). Namely, *PTEN* was disrupted by a translocation in sample EOPC-05, also harboring a large (55 Mb) deletion removing the other *PTEN* allele, with our transcriptome data showing a pronounced *PTEN* downregulation (Figure 2B). Analysis in a large patient cohort revealed a high frequency of homozygous *PTEN* losses involving such combinations of disruptive events, which are linked with biochemical disease recurrence (Figures S2C–S2J). The additional finding of multiple upregulated *PTEN*-targeting miRNAs (Figures S2K–S2O; Tables S5, S6, and S7) indicates the involvement of different mechanisms of *PTEN* inactivation in PCA.

### An Abundance of Gene-Fusing Rearrangements, Including such Leading to ETS Fusion Genes, in EO-PCA

The striking abundance of balanced SRs in our EO-PCA samples—with approximately half of all SRs being balanced (a higher number than that recently reported [ $<30\%$ ] in elderly-onset PCAs; Berger et al., 2011)—prompted us to investigate in further detail the formation of gene-fusing rearrangements, because fusion gene formation is frequently associated with balanced SR formation (Rubin et al., 2011; Tomlins et al., 2005). Indeed, we could identify a relatively large number

**Table 1. Summary of Clinical Data for the 11 Patients with EO-PCA**

Patient ID	Age at Surgery (years)	pT Stage	pN Stage	Gleason Score	Preoperative PSA (ng/ml)	Family History of PCA <sup>a</sup>
EOPC-01	45	pT3a	pN0	3+4	30.0	no
EOPC-02	51	pT2c	pN0	3+4	23.8	NA
EOPC-03	46	pT3a	pN0	5+4	12.0	no
EOPC-04	51	pT3b	pN1	4+3	72.0	no
EOPC-05	50	pT3a	NX	3+4	5.0	no
EOPC-06	38	pT3b	pN1	3+4	18.2	NA
EOPC-07	49	pT2c	pN0	3+4	5.4	yes
EOPC-08	44	pT2c	pN0	3+4	7.8	yes
EOPC-09	48	pT2c	pN0	3+4	5.1	no
EOPC-010	46	pT2c	NX	3+4	6.0	no
EOPC-011	48	pT2c	pN0	3+4	4.8	yes

All patients were diagnosed  $<50$  years. All patients are of German ancestry. PSA, prostate-specific antigen; NA, not applicable.

<sup>a</sup>At least one first-degree relative with PCA.

**Table 2. Sequence Data and Somatic Genome Alterations in 11 EO-PCAs**

Patient ID	EOPC-01	EOPC-02	EOPC-03	EOPC-04	EOPC-05	EOPC-06	EOPC-07	EOPC-08	EOPC-09	EOPC-010	EOPC-011
Bp (paired) (tumor/blood)	179/165	110/135	245/99	112/230	105/102	141/135	104/100	139/102	110/134	151/148	104/102
Sequence coverage	42x/44x	30x/26x	69x/30x	32x/65x	31x/30x	41x/40x	31x/30x	29x/31x	33x/40x	46x/44x	31x/30x
PE physical coverage	30x/33x	21x/24x	75x/28x	30x/76x	29x/28x	38x/36x	29x/27x	38x/29x	27x/33x	40x/39x	27x/27x
MP physical coverage	14x/17x	15x/15x	16x/16x	16x/11x	17x/15x	16x/18x	13x/17x	17x/20x	18x/18x	17x/17x	18x/18x
mRNA-Seq reads (million reads/Gb)	18/14	9/7	10/7	10/8	13/11	14/11	13/10	12/10	16/13	14/10	13/10
miRNA-Seq in M reads (total/mapped to miRNAs)	NA	NA	NA	NA	144/ 19	66/ 16	71/ 26	82/ 12	74/9	74/ 11	92/ 24
Methylome reads in M (total/mapped)	67/21	89/23	75/20	85/24	62/16	119/22	143/30	115/24	169/32	128/27	88/21
Somatic SNVs	5,696	3,747	4,529	2,430	1,475	1,277	1,404	931	985	1,536	1,567
Mutations per Mb	2.0	1.3	1.6	0.8	0.5	0.4	0.5	0.3	0.3	0.5	0.5
Nonsilent coding SNVs	55	11	40	21	6	8	9	6	3	10	6
Nonsilent InDels	1	1	0	0	0	0	1	1	0	0	0
Somatic Dels (genes/Mb)	1452/235	442/101	671/124	598/84	1,043/194	2,283/347	961/153	524/91	481/93	771/144	566/125
Somatic Dups (genes/Mb)	138/18.8	1/0.0	4/1.0	45/6.7	0/0.0	21/2.2	1/0.6	7/1.5	15/ 2.3	1/0.0	42/ 9.7
Translocation (coding/noncoding breaks)	15/11	12/8	14/4	32/40	27/27	38/20	14/4	11/1	21/5	11/1	26/18
Inversions (coding/noncoding breaks)	3/1	2/8	1/1	11/7	2/4	10/4	2/10	2/0	21/7	3/3	10/6
Estimated tumor purity (%)	50	40	37	55	45	47	60	61	44	50	51

Production centers—sample acquisition and nucleic acid isolation: Martini Clinic, Prostate Cancer Center, and Institute of Pathology, University Clinic Hamburg-Eppendorf (UKE); paired-end (PE) whole-genome sequencing: Max Planck Institute for Molecular Genetics (MPIMG), German Cancer Research Center (DKFZ), and European Molecular Biology Laboratory (EMBL); long-range paired-end, or mate-pair (MP), genomic sequencing: EMBL; transcriptome sequencing: MPIMG; miRNA sequencing: DKFZ; methylome sequencing: DKFZ. NA, not available.

See also [Figure S1](#) and [Table S1](#).

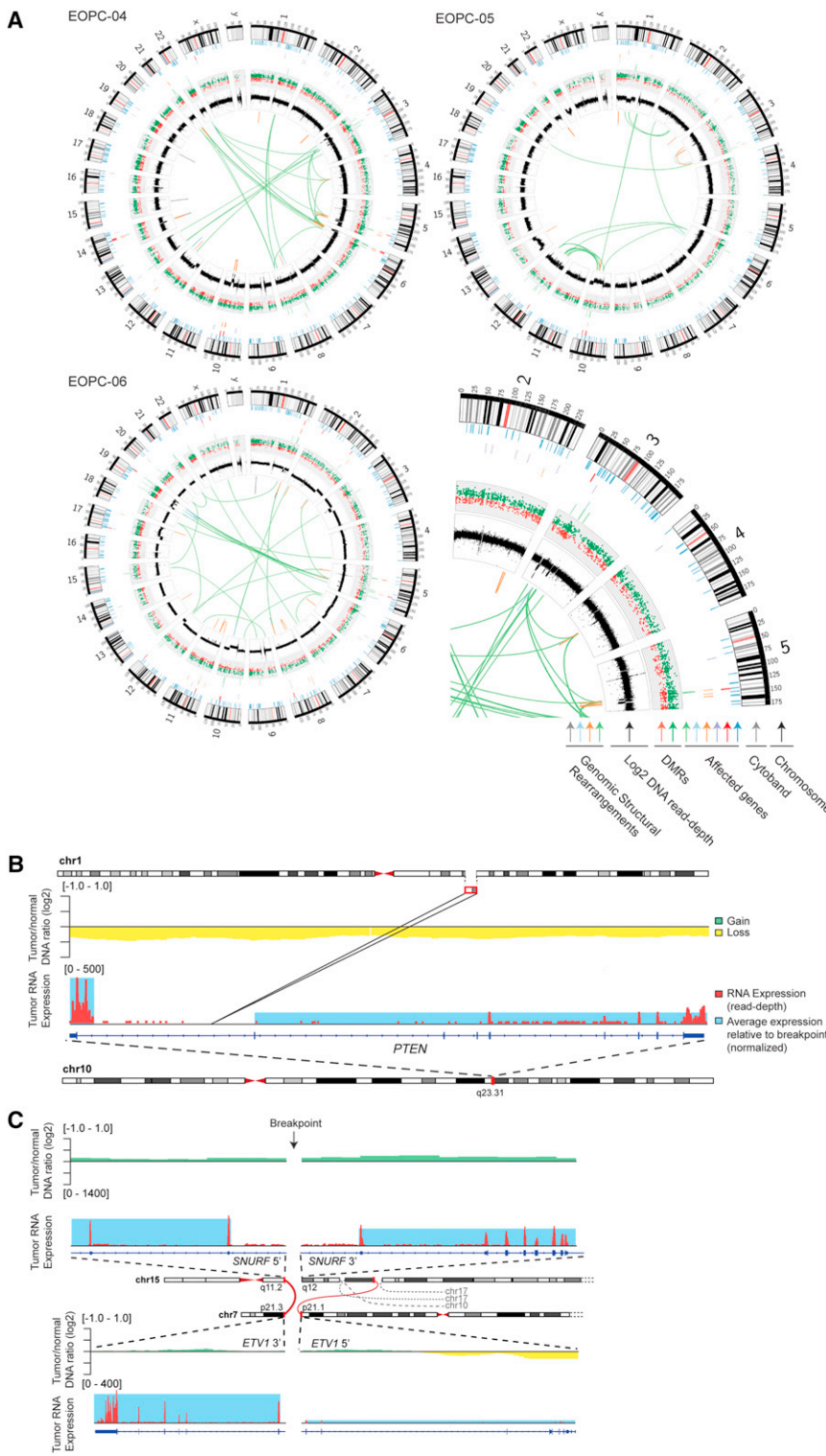
(ie, 139) of gene-fusing rearrangements (12 per sample) in the 11 EO-PCA samples, 109 of which resulted from balanced SRs (mostly translocations). While many of these gene-fusing events were out-of-frame and likely correspond to passenger events, most in-frame fusion genes received additional support from mRNA-seq reads mapping directly over the junction point of the fusion, or by the striking overexpression of the 3' fusion partner ([Table S2](#)), suggesting their possible biologic relevance.

Fusion genes supported by marked expression upregulation included the prototypical ETS gene family-associated fusion genes *TMPRSS2:ERG* and *SLC45A3:ERG* ([Tomlins et al., 2005, 2007](#)). Furthermore, we observed evidence for 17 rare (nonrecurrent) previously undescribed in-frame gene fusion events ([Table S2](#)). These included *SNURF:ETV1*, which evidently was formed in conjunction with a complex rearrangement event

(closed chain; [Berger et al., 2011](#)) involving an androgen-regulated 5'-end fusion gene partner (*SNURF*) that, based on our mRNA-Seq data, led to marked overexpression of an *ETV1* open reading frame with previously demonstrated oncogenic activity ([Tomlins et al., 2007](#)) ([Figures 2C](#) and [S2P–S2T](#)). We also identified 17 fusion gene transcripts that were exclusively discovered in our mRNA-seq data, ie, events missed by our DNA sequencing-based SR detection pipeline (see [Supplemental Experimental Procedures](#)).

### Early-Onset Prostate Cancers Harbor a Distinctive Landscape of Genetic Rearrangements

The conspicuous frequency of gene-fusing SRs in EO-PCA prompted us to perform a more detailed comparison between the genomic landscapes of young and elderly patients, by



**Figure 2. Genomes of EO-PCA Revealed by High-Throughput Sequencing**

(A) Circos plots, depicting genomic and epigenomic somatic aberrations in three EO-PCA patients (all 11 EO-PCA genomes are displayed in Figure S2A). Displayed features from outer to inner ring include chromosomes (black) with cytobands (gray) in the first ring; lines in second ring refer to genes, with colors highlighting CancerCensus genes (Futreal et al., 2004) (blue), biallelic inactivated genes (red), nonsynonymous SNVs (purple), inferred gene-disrupting deletions (orange), gene-disrupting inversions (light blue) and gene-disrupting translocations (light green); squares in third ring refer to hypermethylated (dark green) and hypomethylated DMRs (light red); the fourth ring displays somatic inferred copy-number gains and losses as read-depth plots; the innermost shows translocations (green), deletions (orange), inversions (light blue), and duplications (gray).

(B) Biallelic inactivation of *PTEN* by combined loss-of-heterozygosity and a disruptive t(1;10) translocation in EOPC-05. RNA expression values are depicted up- and downstream of the breakpoint, along with the inferred copy-number profile of the tumor sample relative to germline control ( $\log_2$  DNA read-depth ratio), with losses indicated in yellow and gains in green. Gene expression values are displayed in terms of the RNA-seq read-depth (red) and averaged RPKM (light blue) values upstream and downstream of the insertion event.

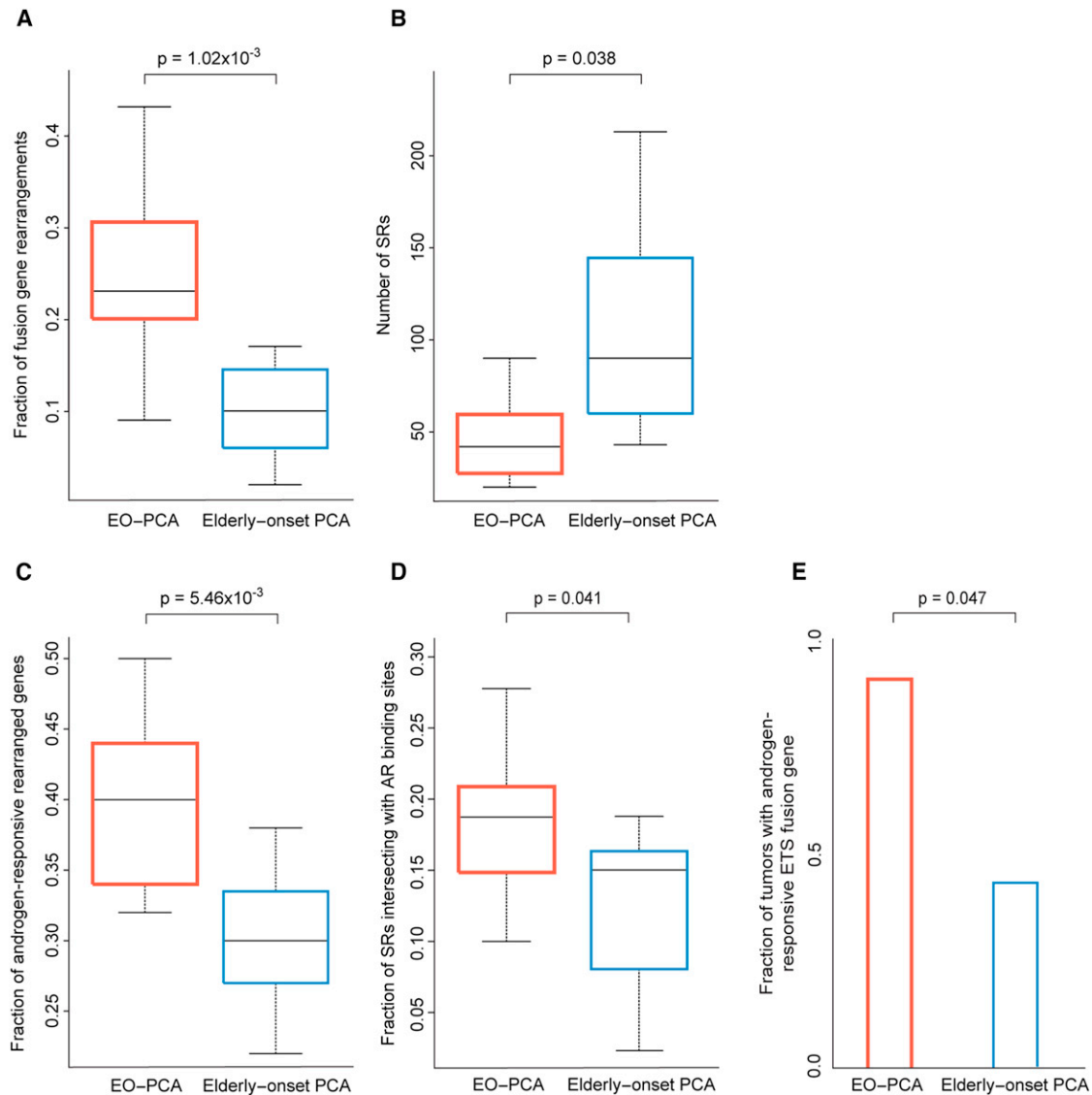
(C) Rearrangements leading to the identified *SNURF:ETV1* fusion in EOPC-03, comprising up to intron 2 of *SNURF* at the 5'-end, and continuing from intron 4 of *ETV1* at the 3'-end.

See also Figure S2 and Tables S2, S3, S4, S5, S6, and S7).

higher portion of SRs leading to gene-fusing events (25% in EO-PCA versus 10% in the elderly; ie, 2.5-fold enrichment with  $p = 0.001$ ; Welch's two-sample t test; Figure 3A). This was despite an overall lower number of SRs in EO-PCA compared to elderly-onset PCA genomes (average 45 versus 108 rearrangements, respectively;  $p = 0.038$ , Welch's two-sample t test; Figure 3B), a trend that we also confirmed upon reanalysis of the raw DNA sequence read data from Berger et al. using our computational SV detection pipeline (see Supplemental Experimental Procedures).

By comparison, when assessing somatic SNVs, we observed no significant difference between EO-PCAs and elderly-onset PCAs in terms of genome-wide (or protein-coding sequence-wide) SNV counts (Table S1). Hence, despite marked differences evident at the level of SRs, no such differences appeared to exist at the level of SNVs.

comparing genetic alterations identified in our samples with the recently published whole-genomes of seven elderly-onset ("classical") PCAs, with a mean age at diagnosis of 65 years (Berger et al., 2011). Relative to the total number of somatic rearrangements, EO-PCAs indeed displayed a significantly



**Figure 3. Whole-Genome Sequencing-Based Analyses Reveal that EO-PCAs Harbor a Markedly Different SR Landscape than Elderly-Onset PCAs**

(A) Proportion of SRs leading to gene fusions. These analyses considered all breakpoints affecting Refseq gene models.

(B) SRs identified per sample.

(C) Fraction of gene rearrangements affecting androgen-regulated genes.

(D) Fraction of SRs intersecting high-confident androgen-receptor binding sites (ARBS) using a 50-kb search window.

(E) Portion of tumors harboring ETS fusion genes.

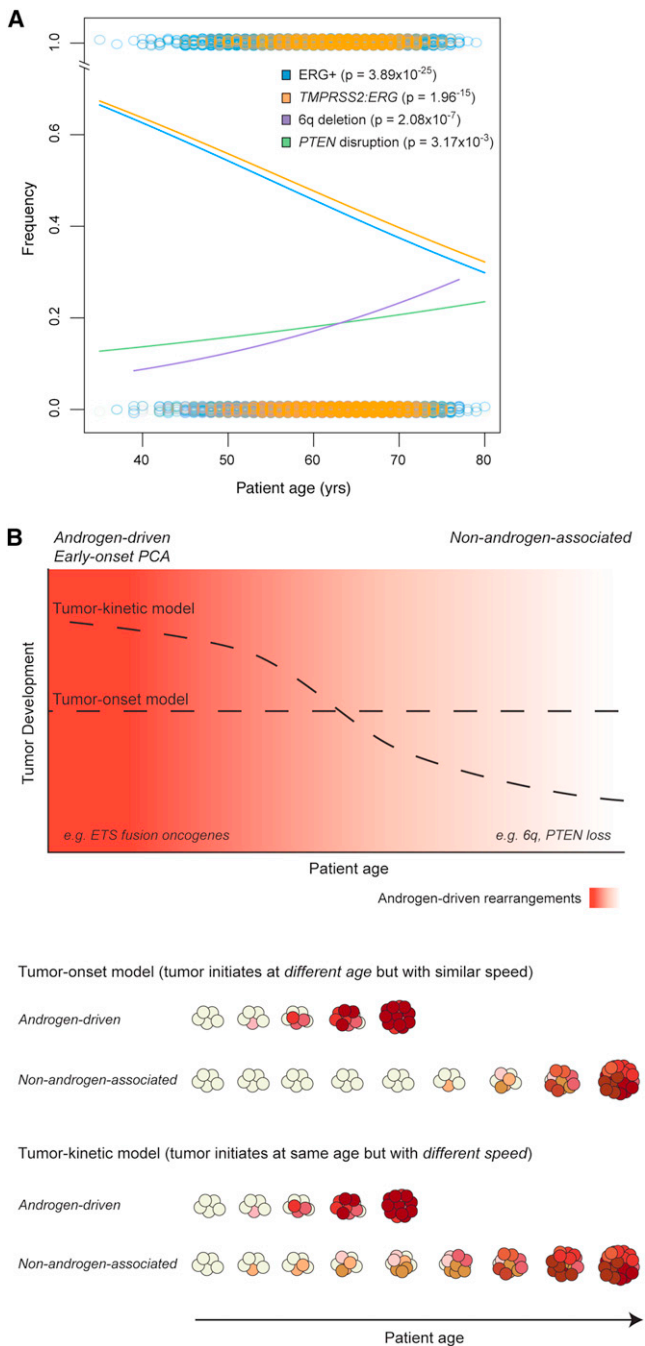
Welch's two-sample t test was used to compute p values in (A)–(D). Boxplots display the 25th to 75th percentiles (boxes), medians (lines), and 1.5 times the interquartile range (whiskers).

See also [Table S2](#).

### Early-Onset Prostate Cancers Harbor an Androgen-Driven Somatic Genome Alteration Landscape

Because gene-fusing events in PCA can be mediated directly upon activation of the AR (Mani et al., 2009), facilitating gene fusion events that involve androgen-regulated genes bound by the AR (Rubin et al., 2011), we hypothesized that the observed differences in the somatic SR spectrum of EO-PCA versus elderly-onset PCA may have resulted from the differential usage of an AR-dependent SR formation mechanism driving

specific DNA rearrangements. To test this hypothesis, we first generated a catalog of genes under androgen-regulation, by measuring differential expression before and after dihydrotestosterone-stimulation of LNCaP cells using gene expression microarrays (see [Supplemental Experimental Procedures](#) and [Table S2](#)). Strikingly, we observed a significantly higher fraction of gene rearrangements affecting androgen-driven genes—in which a break occurred in a gene differentially expressed upon dihydrotestosterone-stimulation—in EO-PCA compared to



**Figure 4. TMA Verifies Distinct Age Spectra of Androgen-Dependent and -Independent Progression Types**

(A) Large-scale TMA analyses depict age relationships of the androgen-driven ETS-fusion protein *TMPRSS2:ERG* in EO-PCA, and of nonandrogen-associated rearrangements in elderly-onset PCA. Frequency of assessed protein (rings) and binomial logistic regressions (lines) depict *ERG* overexpression (blue;  $n = 9,567$ ), *TMPRSS2:ERG* fusion gene presence (orange;  $n = 6,071$ ), chromosome 6q deletion (purple;  $n = 3,493$ ), and *PTEN* deletion/disruption (green;  $n = 5,374$ ) as a function of patient age (age at diagnosis ranging between 36 and 80 years). *ERG* overexpression was detected by immunohistochemistry, and genomic rearrangements by FISH. P values are based on binomial logistic regression.

(B) Proposed tumor progression models for androgen-type PCAs. See also Figure S3.

elderly-onset PCA (Figure 3C;  $p < 0.01$ ; Welch's two-sample t test). In further support of these findings, we observed a statistical enrichment of AR signaling and WNT signaling (which interacts with AR signaling in PCA; Yang et al., 2006) pathways among genes involved in gene fusion events in EO-PCAs, but not in elderly-onset PCAs (see Experimental Procedures and Table S2). These findings indicate a prevalence of EO-PCA to acquire specific, androgen-driven, somatic genome alterations.

AR binding to nuclear DNA can facilitate genomic rearrangements through recruitment of topoisomerase 2B (TOP2B), leading to DNA double strand break formation within or nearby transcriptional hubs, joining different sections of the genome that upon breakage may reconnect to lead to intra- or inter-chromosomal SRs (Haffner et al., 2010; Lin et al., 2009a). Hence, we followed the assumption that if androgen drives the formation of SRs characteristically arising in EO-PCA, we would observe a relative enrichment of genomic AR binding sites near the breakpoints of SRs in EO-PCA relative to elderly-onset PCA. High-confidence binding sites of the AR were recently mapped in LNCaP cells by chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq), upon androgen stimulation (Urbanucci et al., 2012). Indeed, when relating these ChIP-Seq data to our SR data, we observed a marked enrichment of SRs intersecting with AR binding sites (Urbanucci et al., 2012) in EO-PCA compared to elderly-onset PCA (Figure 3D;  $p < 0.05$ , Welch's two sample t test). These results are consistent with the chromosomal looping by the AR bound to its cognate binding sites specifically facilitating SRs in EO-PCA in an androgen-driven manner.

### ETS Fusion Genes Are a Hallmark of Early-Onset Prostate Cancer

We next assessed the status of ETS family oncogene containing fusion genes, with ETS fusions representing prototypic PCA driver gene rearrangements mediated by androgen stimulation and the AR (reviewed in Rubin et al., 2011). Even though the number of sequenced PCA genomes is presently small, comparison of the ETS rearrangement status in 11 EO-PCAs versus the seven elderly-onset PCAs published by Berger and coworkers already revealed statistically significant differences, with 10 of 11 (~90%) EO-PCAs, but only three of seven (~40%) elderly-onset PCAs harboring such androgen-driven ETS rearrangements (Figure 3E;  $p = 0.047$ ; two-sided Fisher's exact test). Notably, the abundance of ETS fusions in EO-PCA was further supported by our reanalysis of data published by Barbieri et al., a study in which four of five patients  $\leq 50$  years harbored *TMPRSS2:ERG*-fusions, compared to 54 of 107 patients  $> 50$  years (Barbieri et al., 2012).

To verify the specific abundance of androgen-driven SRs in EO-PCA in a larger patient cohort, we made use of our TMA resource, covering thousands of patients (Figure 4A). We used a surrogate for assessing androgen-driven ETS gene fusion events, by applying a break-apart FISH probe for evaluating the presence of the prototypical androgen-driven *TMPRSS2:ERG* fusion gene (Lin et al., 2009a; see Experimental Procedures) in patients with a disease age-of-onset of 36–80 years. Indeed, the TMA analyses verified the continuous and significant relative increase of the ETS gene fusion *TMPRSS2:ERG* ( $p < 2 \times 10^{-15}$ , logistic regression) in patients with early disease onset, hence

further strengthening our finding based on whole-genome sequencing of an abundance of androgen-driven alterations in EO-PCA (Figures 4A and S3). Because *ERG* gene fusions may involve androgen-regulated genes other than *TMPRSS2*, we additionally measured *ERG* protein expression by immunohistochemistry (IHC), an analysis that further substantiated the marked overexpression of *ERG* ( $p < 4 \times 10^{-25}$ , logistic regression) in young versus elderly patients (Figures 4A and S3).

### Nonandrogen-Associated Structural Rearrangements Frequently Accumulate in Elderly-Onset Prostate Cancer

Because SRs affecting androgen-regulated genes are markedly enriched in EO-PCAs, whereas elderly-onset PCAs show an overall higher load of SRs (Figure 3), we screened several additional recurrent SRs (Sun et al., 2007) for a relationship with patient age at diagnosis, including those that do not involve androgen-regulated genes. To this end, we performed extensive FISH analyses, targeting the chromosome 6q15 region (Liu et al., 2007b), the *PTEN* locus (Krohn et al., 2012), the *CHD1* locus (Huang et al., 2012), and *NCOR2*. Our data showed a significant increase of 6q15, *PTEN*, and *CHD1* genomic breaks, all of which are not considered to be androgen-regulated, in elderly patients (Figures 4A and S3)—findings that were independent of tumor stage or Gleason grade (see also Table S2, which presents data from detailed assessments of *PTEN* and *ERG*). These results show that nonandrogen-associated SRs accumulate in elderly-onset PCA. As a consequence, elderly-onset PCAs acquire high loads of SRs, most of which correspond to copy-number unbalanced alterations.

By comparison, deletions of *NCOR2* were significantly associated with young age (Figure S3). Because *NCOR2* is an AR corepressor (Hodgson et al., 2005; Godoy et al., 2012), deletions of this gene are predicted to lead to increased AR levels and in turn may contribute to androgen-driven rearrangements in young patients.

### Increased Levels of Androgen or Its Receptor May Explain the Early Disease Onset in Androgen-Driven PCAs

The high abundance of androgen-driven genetic alterations in EO-PCA suggests a relevance of those alterations for the timing and initiation of tumorigenesis in the cells of the prostate. Specifically, the preponderance of androgen-driven SRs in young patients may intuitively be explained by a particularly early development of tumors occurring in young patients (tumor-onset model; Figure 4B). Alternatively, particularly rapid growth kinetics of androgen-driven tumors may also explain the observed age relationship (tumor-kinetic model; Figure 4B).

Because the initiation of tumorigenesis in the prostate cannot be confidently ascertained in humans—with PCAs initiating years or decades before being diagnosed (Knudsen and Vasioukhin, 2010)—we sought to evaluate the possible role of tumor kinetics in the abundance of androgen-driven genetic alterations in EO-PCA. We therefore employed IHC to measure cancer cell proliferation using the Ki67 labeling index (LI) as a marker (Bubendorf et al., 1996). Importantly, while growth kinetics differed markedly between tumor grades (Figure 5A), there were no observable differences between the growth kinetics of

androgen-driven (using *ERG* positivity as a proxy) versus nonandrogen-associated (*ERG*-negative) tumors (Figure 5A). Furthermore, there was no correlation between cancer cell proliferation and age at diagnosis ( $R^2 \sim 0$ ;  $p = 0.44$ , Spearman's rank correlation). Hence, we can practically exclude a role of growth kinetics in facilitating androgen-driven cancer occurrence in young patients. Furthermore, the overall similar aggressive potential of androgen-driven (*ERG*-positive) tumors versus nonandrogen-associated (*ERG*-negative) tumors is consistent with tumor growth kinetics playing no marked role in generating the herein observed age-associated SR spectra (Figure S4A). Hence, we conclude that differences in the age-of-initiation of tumors harboring androgen-driven SRs versus those not showing an abundance of such SRs must have resulted in the characteristic somatic DNA alteration landscapes of EO-PCA versus elderly-onset PCA (Figure 4B).

### Androgen Receptor Levels Are Elevated in Early-Onset Prostate Cancer

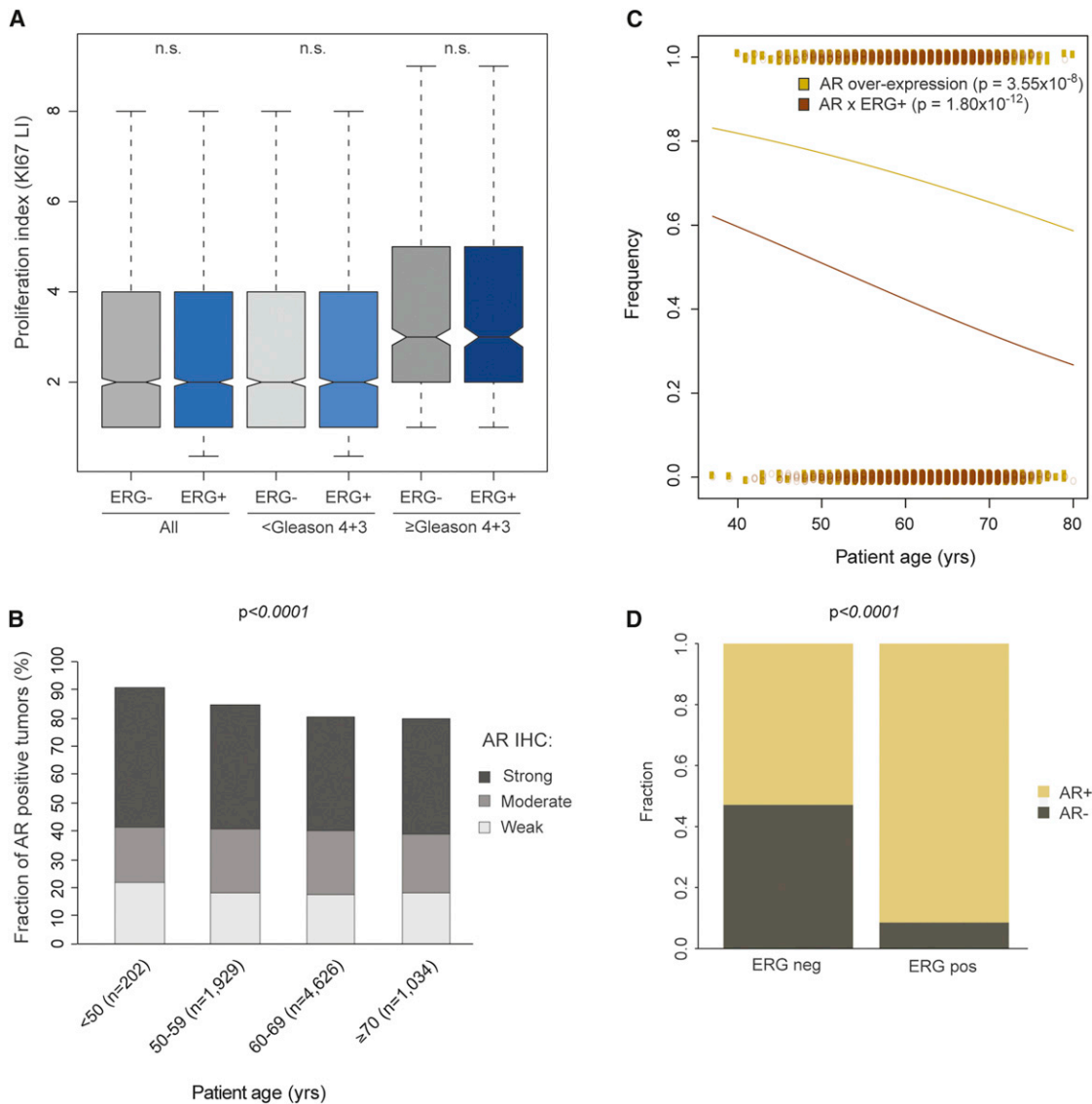
Our findings hence point toward a crucial role of androgen and AR activation in shaping the genomic landscape of PCAs, and driving tumor initiation, motivating further *in vivo* analyses. While it is impossible to estimate testosterone levels retrospectively in patients during cancer initiation and evolution, patterns of AR activity can be studied in tumor samples by measuring AR protein expression (as a surrogate) in tumor samples at diagnosis (Lee and Chang, 2003). We pursued AR expression analysis using IHC, relating AR levels to patient age and *ERG* status. Importantly, our data revealed both a significantly increased AR level in young patients (Figures 5B and 5C) and a significant positive correlation of AR levels with *ERG* rearrangements ( $p < 4 \times 10^{-8}$  and  $p < 2 \times 10^{-12}$ , respectively; Figure 5C; see also Figures 5B–5D, S4B, and S4C). Hence, high AR expression levels are associated with *ERG* rearrangements and young age, indicating an age-dependent activity of AR in PCA. *ERG*, when highly expressed, was shown to lead to an inhibition of AR expression in PCA cell lines (Yu et al., 2010), which makes an inverse scenario in which *ERG* rearrangements result in high AR levels exceedingly unlikely. Hence, our data indicate a key role of androgen, and the AR, in mediating age-dependent genomic alterations characteristically observed in EO-PCA.

## DISCUSSION

PCA is typically characterized by a prolonged clinical course with an estimated disease onset one to two decades prior to clinical diagnosis (Lilja et al., 2007). Our genomics data based on massively parallel DNA sequencing, comprising the largest number of PCA whole-genomes sequenced to date, show that most EO-PCAs involve an androgen-driven pathomechanism characterized by a marked absolute and relative abundance of DNA structural alterations involving androgen-regulated genes (“androgen-type prostate cancer”). Our data are consistent with a role of tumor onset, rather than tumor growth kinetics, in driving the early occurrence of androgen-type PCA.

It is tempting to speculate that this early development of androgen-type PCA may be driven by a particularly strong AR activation, eg, by high testosterone levels, associated with young patient age. It has previously been established that





**Figure 5. EO-PCA Is Associated with High AR levels**

(A) Proliferation rates, measured by KI67 IHC staining, for high- and low-grade PCAs, and specifically for ERG positive (proxy for androgen-driven; n = 2,888) and ERG-negative tumors (proxy for nonandrogen-associated; n = 3,095). Low, Gleason grade < 4+3; High, Gleason grade ≥ 4+3; KI67 Labeling Index, LI.

(B) Age-relationship of AR overexpression, detected as strong, moderate, or weak expression based on judgment of IHC staining intensity by an experienced pathologist (see Supplemental Experimental Procedures).

(C) Age-related interaction between AR overexpression and ERG rearranged tumors. Logistic regression of tumors with AR overexpression (yellow line, n = 4,668; p = 3.55 × 10<sup>-8</sup>), and tumors with both AR and ERG overexpression (AR × ERG; frequency of tumors displaying both ERG and AR overexpression, red line, n = 4,172; p = 1.80 × 10<sup>-12</sup>; logistic regression).

(D) AR overexpression in ERG-positive and ERG-negative tumors (n = 4,172; p values based on Fisher’s exact test). Boxplots in (A) display the 25th to 75th percentiles (boxes), medians (lines), and 1.5 times the interquartile range (whiskers). Notches show the 95% confidence interval of the median. n.s., not significant.

See also Figure S4.

average serum testosterone levels in healthy men continuously decrease with age (Liu et al., 2007a; Mohr et al., 2005; Figure S4B) in a manner that remarkably parallels our observed age-associated decrease of ERG positivity in PCA. Furthermore, in vitro-data have demonstrated that the formation of ERG rearrangements can be induced by serum androgen levels in prostate epithelial cells (Lin et al., 2009a; Mani et al., 2009)

through a mechanism involving inter- or intra-chromosomal transcriptional hubs with co-recruitment of *TOP2B*, resulting in SR-inducing DNA double-strand breaks (Haffner et al., 2010; Lin et al., 2009a). Increased AR levels in young patients, which we established by in vivo measurements at diagnosis, provide further support to a crucial role of androgen in EO-PCA. In addition, the strong link seen between AR expression and

positive ERG status across all age groups, including patients older than 70 years suggests that high AR expression, and hence activity, may result in androgen-type PCA, with androgen-driven SRs irrespective of patient age.

Importantly, our finding of a specific pathomechanism driving EO-PCA may be of relevance for a tailored clinical management. Androgen-mediated rearrangements cause a complex androgen-associated modulation of transcriptional patterns and cellular pathways (Brase et al., 2011), with potential consequences for disease progression and response to androgen-ablative treatment in EO-PCA. Furthermore, cancer screening by detection of androgen-mediated rearrangements, eg, from biopsies, circulating tumor cells, or free circulating DNA, may be particularly effective in young men.

With over 200 PCA samples having already been subjected to exome sequencing (Barbieri et al., 2012; Grasso et al., 2012; Kumar et al., 2011), the identification of previously uncharacterized recurrent somatic SNVs affecting protein-coding regions is becoming exceedingly challenging. At the same time, approaches for clinical validation are becoming increasingly relevant. In this regard, our study shows the utility of evaluating genetic findings on large-scale TMAs containing thousands of samples. Moreover, our study emphasizes the power of whole-genomic sequencing, revealing information on mechanisms of SR formation by ascertaining genetic variants beyond those detectable by exome sequencing. Our genome-based findings on SR formation in conjunction with large-scale FISH and IHC-based assessment of ERG and AR enabled us to find a remarkable missing link in the relationship of the AR, somatic genome alterations, and PCA subtypes. In summary, our findings demonstrate striking age-dependent differences in the mechanistic landscapes of structural genomic alterations in a common cancer.

## EXPERIMENTAL PROCEDURES

### Patients

Informed consent and an ethical vote (institutional reviewing board) were obtained according to the current ICGC guidelines (see <http://www.icgc.org>). The patients did not receive any neo-adjuvant radiotherapy, androgen deprivation therapy, or chemotherapy prior to the surgical removal of tumor tissue. Tumor samples and one normal prostate control were frozen at  $-20^{\circ}\text{C}$  and subsequently stored at  $-80^{\circ}\text{C}$ .

### DNA Library Preparation and Sequencing

DNA library preparation and whole-genome sequencing was performed on Illumina sequencers as described recently (Rausch et al., 2012a) with the raw length of the reads displaying a median of 101 bp for short paired-end insert-size libraries and 36 bp for large insert-size mate-pair libraries, and a median insert-size of the sequenced libraries of 155–206 bp (short insert-size paired-end), and 4,265–5,350 bp (large insert-size mate-pairs).

### DNA Read Mapping and Sequence Variant Calling

Reads were aligned to the hg19 assembly of the human reference genome using the proprietary ELAND2 tool from Illumina, as well as the BWA tool (Li and Durbin, 2009). Post-processing of the aligned reads included merging of lane-level data and removal of duplicate read-pairs using Picard tools (<http://picard.sourceforge.net>). Only uniquely aligned reads were considered for downstream mutation analysis. Aligned reads were converted to the SAM/BAM format using SAMtools (Li et al., 2009), before initiating a set of complementary variant calling algorithms. We applied three distinct computational pipelines for SNV discovery, and subsequently kept SNV calls if

they were identified by at least two out of the three pipelines (Details and parameters are in the Supplemental Experimental Procedures). InDels were called using the Genome Analysis Toolkit (GATK) (DePristo et al., 2011) and SAMtools (Li et al., 2009) mpileup. Consensus InDels were manually inspected to identify potential misalignments. SRs  $\sim 200$  bp to megabases in size were detected using the DELLY tool (Rausch et al., 2012b), as previously described (Rausch et al., 2012a), employing discordantly mapping read-pairs, split-read analysis (using an approach comparable to the soft-clipping functionality of CREST; Wang et al., 2011), as well as sequencing depth-of-coverage (for further details, see Supplemental Experimental Procedures). PCR verification and Sanger sequencing of sequence variants was performed as previously described (Rausch et al., 2012a). The False Discovery Rate (FDR) was obtained for randomly picked SNVs with a minor allele frequency (MAF) greater than 0.15 (to enable capillary sequencing-based data interpretation). We further validated a number of SNVs below 0.15 MAF (marked in gray in Table S3; see also Figure S1A).

### mRNA Library Sequencing and Analysis

Strand-specific mRNA sequencing libraries were prepared from 10  $\mu\text{g}$  of total RNA as recently described (Parkhomchuk et al., 2009). However, instead of shearing the cDNA, we fragmented the RNA PolyA+ fraction. The following modifications were implemented: the purified polyA+ RNA fraction was fragmented at  $70^{\circ}\text{C}$  for 5 min using RNA fragmentation reagents (Ambion, Cat. no. AM8740) and following the manufacturer instructions; the first strand synthesis was performed with random hexamers (dN)6 primers. Sequencing was carried out with  $2 \times 51$  cycles on the Illumina HiSeq2000 instrument (further details are in the Supplemental Experimental Procedures). Reads were aligned to the hg19 genome assembly using BWA (0.5.9-r16) (Li and Durbin, 2009). For calculation of fusion gene expression values, RNA-seq-derived average exon RPKM values were extracted 5' and 3' to the breakpoint and a relative expression ratio was calculated after normalizing to the respective exon RPKM of the normal control. The minimum exon expression value was set to 1 RPKM. Further details are in the Supplemental Experimental Procedures.

### Small RNA Sequencing and Analysis

Small RNAs (mainly miRNAs) up to 40 nt were size-fractionated on a polyacrylamide gel from up to 5  $\mu\text{g}$  mRNA-depleted, DNase-treated (RNase-free DNase I, QIAGEN, Hilden, Germany) RNA. Small RNA libraries were prepared using the NEBNext Small RNA Sample Prep Set (NEB, Frankfurt/M., Germany) as described by the manufacturer, with some modifications (see Supplemental Experimental Procedures for details). Amplicons of  $\sim 90$ –100 bp were sequenced on an Illumina HiSeq2000 instrument (read length: 50 bp) and mapped to the reference using the miRDeep2 package (Friedländer et al., 2012).

### DNA Methylation Sequencing and Analysis

MCIp for enrichment of highly methylated DNA was performed as described previously (Gebhard et al., 2006) (see Supplemental Experimental Procedures for details). For deep-sequencing library preparation with highly methylated DNA, the NebNext chemistry (New England Biolabs, Ipswich, MA, USA) and barcoded adaptors compatible with the SOLiD sequencing platform were used. Single-end 50 bp reads were generated using the SOLiD 4 next-generation sequencing platform (Applied Biosystems, Life Technologies Corporation, Carlsbad, CA, USA). Mapped (software BFAS; Homer et al., 2009) and quality controlled reads were used to identify DMRs between tumor and normal (further details in the Supplemental Experimental Procedures).

### Biallelic Inactivation Inference and Pathway Analysis

We scanned for genes inactivated in a biallelic fashion by extracting loci with two overlapping events, ie, germline or somatic homozygous nonsynonymous SNVs or an exon-disrupting deletion overlapping a second gene-disrupting aberration such as an SNV, deletion, inversion, or translocation. Genes within deletions were additionally required to have read-depth support. With estimated raw SNV allele frequencies ranging from 0.15 to 0.4, raw SV allele frequencies estimated to range from 0.20 to 0.45 and estimated tumor purities of  $\sim 0.5$ , the vast majority of intersecting events (eg, a deletion overlapping a mutation, or gene disruption) occurred on different alleles in the same cell

population (rather than representing subclonal events affecting different cell subpopulations). For carrying out the pathway enrichment analysis, fusion genes were extracted and analyzed by Genomatix Genome Analyzer with standard parameters. The *p* values were adjusted for multiple-testing by controlling the FDR according to Benjamini and Hochberg.

#### PCA Tissue Microarray Resource

Details on the PCA prognosis TMA (earlier described in Schlomm et al., 2008) used for FISH and IHC analyses are in the [Supplemental Experimental Procedures](#).

#### ERG Immunohistochemistry Analysis

ERG IHC using antibody ERG (clone EPR3864, dilution 1:450, Epitomics) was performed as previously described (Minner et al., 2011). See [Supplemental Experimental Procedures](#) for further details.

#### FISH Analysis

FISH analysis was performed as previously described before (Minner et al., 2011). Details on probes are in the [Supplemental Experimental Procedures](#).

#### TMA Statistical Analysis

Logistic regression analyses were performed using the generalized linear model glm package in R. Statistical significance and fitness of the models were verified with the Wald test and likelihood ratio test.

#### ACCESSION NUMBERS

The European Genome-phenome archive database (hosted at the EBI) accession number for the short-read sequencing data reported in this paper is EGAS00001000258.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, seven tables, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.ccr.2013.01.002>.

#### ACKNOWLEDGMENTS

We thank Anke Renter, Marion Bähr, Oliver Mücke, Andreas Schlattl, Anne Stranghöner, Christina Koop, Bettina Haase, and Dinko Pavlinic for assistance with experiments and Megumi Ohnishi-Seebacher for helpful comments. This project was supported by the German Federal Ministry of Education and Science in the program for medical genome research (FKZ: 01KU1001A, -B, -C, and -D). P.H. was supported by a stipend from the Alexander von Humboldt Foundation, and J.K. was additionally supported by the European Commission (Health-F2-2010-260791). The authors acknowledge assistance provided by sequencing and high-performance computational analysis centers at the Max Planck Institute for Molecular Genetics, German Cancer Research Center, and European Molecular Biology Laboratory.

Received: May 10, 2012

Revised: August 16, 2012

Accepted: January 3, 2013

Published: February 11, 2013

#### REFERENCES

Albertsen, P.C., Hanley, J.A., and Fine, J. (2005). 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* 293, 2095–2101.

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.

Bastus, N.C., Boyd, L.K., Mao, X., Stankiewicz, E., Kudahetti, S.C., Oliver, R.T.D., Berney, D.M., and Lu, Y.J. (2010). Androgen-induced TMPRSS2: ERG fusion in nonmalignant prostate epithelial cells. *Cancer Res.* 70, 9544–9548.

Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., et al. (2011). The genomic complexity of primary human prostate cancer. *Nature* 470, 214–220.

Brase, J.C., Johannes, M., Mannsperger, H., Fälth, M., Metzger, J., Kacprzyk, L.A., Andrasiuk, T., Gade, S., Meister, M., Sirma, H., et al. (2011). TMPRSS2-ERG -specific transcriptional modulation is associated with prostate cancer biomarkers and TGF- $\beta$  signaling. *BMC Cancer* 11, 507.

Bubendorf, L., Sauter, G., Moch, H., Schmid, H.P., Gasser, T.C., Jordan, P., and Mihatsch, M.J. (1996). Ki67 labelling index: an independent predictor of progression in prostate cancer treated by radical prostatectomy. *J. Pathol.* 178, 437–441.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* 40, 722–729.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52.

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.

Gebhard, C., Schwarzfischer, L., Pham, T.H., Schilling, E., Klug, M., Andreesen, R., and Rehli, M. (2006). Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* 66, 6118–6128.

Godoy, A.S., Sotomayor, P.C., Villagran, M., Yacoub, R., Montecinos, V.P., McNerney, E.M., Moser, M., Foster, B.A., and Onate, S.A. (2012). Altered corepressor SMRT expression and recruitment to target genes as a mechanism that change the response to androgens in prostate cancer progression. *Biochem. Biophys. Res. Commun.* 423, 564–570.

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.

Grönberg, H. (2003). Prostate cancer epidemiology. *Lancet* 361, 859–864.

Haffner, M.C., Aryee, M.J., Toubaji, A., Esopi, D.M., Albadine, R., Gurel, B., Isaacs, W.B., Bova, G.S., Liu, W., Xu, J., et al. (2010). Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nat. Genet.* 42, 668–675.

Heidenreich, A., Bellmunt, J., Bolla, M., Joniau, S., Mason, M., Matveev, V., Mottet, N., Schmid, H.P., van der Kwast, T., Wiegel, T., and Zattoni, F.; European Association of Urology. (2011). EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. *Eur. Urol.* 59, 61–71.

Hodgson, M.C., Astapova, I., Cheng, S., Lee, L.J., Verhoeven, M.C., Choi, E., Balk, S.P., and Hollenberg, A.N. (2005). The androgen receptor recruits nuclear receptor CoRepressor (N-CoR) in the presence of mifepristone via its N and C termini revealing a novel molecular mechanism for androgen receptor antagonists. *J. Biol. Chem.* 280, 6511–6519.

Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE* 4, e7767.

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.

- Knudsen, B.S., and Vasioukhin, V. (2010). Mechanisms of prostate cancer initiation and progression. *Adv. Cancer Res.* 109, 1–50.
- Knudson, A.G.J., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* 68, 820–823.
- Krohn, A., Diedler, T., Burkhardt, L., Mayer, P.S., De Silva, C., Meyer-Kornblum, M., Kötschau, D., Tennstedt, P., Huang, J., Gerhäuser, C., et al. (2012). Genomic deletion of PTEN is associated with tumor progression and early PSA recurrence in ERG fusion-positive and fusion-negative prostate cancer. *Am. J. Pathol.* 181, 401–412.
- Kumar, A., White, T.A., MacKenzie, A.P., Clegg, N., Lee, C., Dumpit, R.F., Coleman, I., Ng, S.B., Salipante, S.J., Rieder, M.J., et al. (2011). Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci. USA*. 108, 17087–17092.
- Lee, D.K., and Chang, C. (2003). Endocrine mechanisms of disease: Expression and degradation of androgen receptor: mechanism and clinical implication. *J. Clin. Endocrinol. Metab.* 88, 4043–4054.
- Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., Hickenbotham, M., et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Lilja, H., Ulmert, D., Björk, T., Becker, C., Serio, A.M., Nilsson, J.A., Abrahamsson, P.A., Vickers, A.J., and Berglund, G. (2007). Long-term prediction of prostate cancer up to 25 years before diagnosis of prostate cancer using prostate kallikreins measured at age 44 to 50 years. *J. Clin. Oncol.* 25, 431–436.
- Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., and Rosenfeld, M.G. (2009a). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139, 1069–1083.
- Lin, D.W., Porter, M., and Montgomery, B. (2009b). Treatment and survival outcomes in young men diagnosed with prostate cancer: a Population-based Cohort Study. *Cancer* 115, 2863–2871.
- Liu, P.Y., Beilin, J., Meier, C., Nguyen, T.V., Center, J.R., Leedman, P.J., Seibel, M.J., Eisman, J.A., and Handelsman, D.J. (2007a). Age-related changes in serum testosterone and sex hormone binding globulin in Australian men: longitudinal analyses of two geographically separate regional cohorts. *J. Clin. Endocrinol. Metab.* 92, 3599–3603.
- Liu, W., Chang, B.L., Cramer, S., Koty, P.P., Li, T., Sun, J., Turner, A.R., Von Kap-Herr, C., Bobby, P., Rao, J., et al. (2007b). Deletion of a small consensus region at 6q15, including the MAP3K7 gene, is significantly associated with high-grade prostate cancers. *Clin. Cancer Res.* 13, 5028–5033.
- Mani, R.S., Tomlins, S.A., Callahan, K., Ghosh, A., Nyati, M.K., Varambally, S., Palanisamy, N., and Chinnaiyan, A.M. (2009). Induced chromosomal proximity and gene fusions in prostate cancer. *Science* 326, 1230.
- Minner, S., Enodien, M., Sirma, H., Luebke, A.M., Krohn, A., Mayer, P.S., Simon, R., Tennstedt, P., Müller, J., Scholz, L., et al. (2011). ERG status is unrelated to PSA recurrence in radically operated prostate cancer in the absence of antihormonal therapy. *Clin. Cancer Res.* 17, 5878–5888.
- Mohr, B.A., Guay, A.T., O'Donnell, A.B., and McKinlay, J.B. (2005). Normal, bound and nonbound testosterone levels in normally ageing men: results from the Massachusetts Male Ageing Study. *Clin. Endocrinol. (Oxf.)* 62, 64–73.
- Parker, C., Muston, D., Melia, J., Moss, S., and Dearnaley, D. (2006). A model of the natural history of screen-detected prostate cancer, and the effect of radical treatment on overall survival. *Br. J. Cancer* 94, 1361–1368.
- Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 37, e123.
- 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. (2012a). Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* 148, 59–71.
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012b). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.
- Reid, A.H., Attard, G., Brewer, D., Miranda, S., Riisnaes, R., Clark, J., Hylands, L., Merson, S., Vergis, R., Jameson, C., et al. (2012). Novel, gross chromosomal alterations involving PTEN cooperate with allelic loss in prostate cancer. *Mod. Pathol.* 25, 902–910.
- Rubin, M.A., Maher, C.A., and Chinnaiyan, A.M. (2011). Common gene rearrangements in prostate cancer. *J. Clin. Oncol.* 29, 3659–3668.
- Ryan, C.J., Elkin, E.P., Cowan, J., and Carroll, P.R. (2007). Initial treatment patterns and outcome of contemporary prostate cancer patients with bone metastases at initial presentation: data from CaPSURE. *Cancer* 110, 81–86.
- Schlomm, T., Iwers, L., Kirstein, P., Jessen, B., Köllermann, J., Minner, S., Passow-Drolet, A., Mirlacher, M., Milde-Langosch, K., Graefen, M., et al. (2008). Clinical significance of p53 alterations in surgically treated prostate cancers. *Mod. Pathol.* 21, 1371–1378.
- Siegel, R., Naishadham, D., and Jemal, A. (2012). Cancer statistics, 2012. *CA Cancer J. Clin.* 62, 10–29.
- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. *Nature* 458, 719–724.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960.
- Sun, J., Liu, W., Adams, T.S., Sun, J., Li, X., Turner, A.R., Chang, B., Kim, J.W., Zheng, S.L., Isaacs, W.B., et al. (2007). DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. *Prostate* 67, 692–700.
- 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.
- Urbanucci, A., Sahu, B., Seppälä, J., Larjo, A., Latonen, L.M., Waltering, K.K., Tammela, T.L., Vessella, R.L., Lähdesmäki, H., Jänne, O.A., and Visakorpi, T. (2012). Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer. *Oncogene* 31, 2153–2163.
- Wang, J., Mullighan, C.G., Easton, J., Roberts, S., Heatley, S.L., Ma, J., Rusch, M.C., Chen, K., Harris, C.C., Ding, L., et al. (2011). CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* 8, 652–654.
- Xi, R., Hadjipanayis, A.G., Luquette, L.J., Kim, T.M., Lee, E., Zhang, J., Johnson, M.D., Muzny, D.M., Wheeler, D.A., Gibbs, R.A., et al. (2011). Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proc. Natl. Acad. Sci. USA* 108, E1128–E1136.
- Yang, X., Chen, M.W., Terry, S., Vacherot, F., Bemis, D.L., Capodice, J., Kitajewski, J., de la Taille, A., Benson, M.C., Guo, Y., and Buttyan, R. (2006). Complex regulation of human androgen receptor expression by Wnt signaling in prostate cancer cells. *Oncogene* 25, 3436–3444.
- Yu, J., Yu, J., Mani, R.S., Cao, Q., Brenner, C.J., Cao, X., Wang, X., Wu, L., Li, J., Hu, M., et al. (2010). An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 17, 443–454.