



# Androgen receptor profiling predicts prostate cancer outcome

Suzan Stelloo<sup>1,†</sup>, Ekaterina Nevedomskaya<sup>1,2,†</sup>, Henk G van der Poel<sup>3</sup>, Jeroen de Jong<sup>4</sup>, Geert JLH van Leenders<sup>5</sup>, Guido Jenster<sup>6</sup>, Lodewyk FA Wessels<sup>2</sup>, Andries M Bergman<sup>7,\*</sup> & Wilbert Zwart<sup>1,\*\*</sup>

## Abstract

Prostate cancer is the second most prevalent malignancy in men. Biomarkers for outcome prediction are urgently needed, so that high-risk patients could be monitored more closely postoperatively. To identify prognostic markers and to determine causal players in prostate cancer progression, we assessed changes in chromatin state during tumor development and progression. Based on this, we assessed genomewide androgen receptor/chromatin binding and identified a distinct androgen receptor/chromatin binding profile between primary prostate cancers and tumors with an acquired resistance to therapy. These differential androgen receptor/chromatin interactions dictated expression of a distinct gene signature with strong prognostic potential. Further refinement of the signature provided us with a concise list of nine genes that hallmark prostate cancer outcome in multiple independent validation series. In this report, we identified a novel gene expression signature for prostate cancer outcome through generation of multilevel genomic data on chromatin accessibility and transcriptional regulation and integration with publically available transcriptomic and clinical datastreams. By combining existing technologies, we propose a novel pipeline for biomarker discovery that is easily implementable in other fields of oncology.

**Keywords** androgen receptor profiling; ChIP-seq; companion diagnostics for prostate cancer; FAIRE-seq; treatment prediction

**Subject Categories** Cancer; Systems Medicine; Urogenital System

**DOI** 10.15252/emmm.201505424 | Received 8 May 2015 | Revised 1 September 2015 | Accepted 3 September 2015 | Published online 27 September 2015

**EMBO Mol Med (2015) 7: 1450–1464**

## Introduction

Prostate cancer is the second most common cancer in men, with worldwide more than one million new patients diagnosed and 300,000 deaths annually (Torre *et al*, 2015). When the disease is confined to the prostate, patients can be treated with prostatectomy and/or radiotherapy with a curative intent. However, the disease recurs in 30% of patients, for which there is no cure (Amling *et al*, 2000).

Androgen receptor (AR) plays a pivotal role in prostate cancer development and progression, by mediating transcription of pro-mitotic genes, including UBE2C and cyclin D, resulting in prostate cancer cell proliferation (Xu *et al*, 2006; Wang *et al*, 2009). Upon androgen stimulation, AR dissociates from its chaperones and translocates to the nucleus (Brinkmann *et al*, 1999). Subsequently, AR binds at distinct genomic regions to mediate expression of directly responsive genes, ultimately leading to tumor cell proliferation (Itkonen & Mills, 2012). AR binding requires accessible chromatin, which is facilitated by pioneer factors, including FOXA1 and GATA2 (Bohm *et al*, 2009). Chromatin-bound AR subsequently recruits coactivators and corepressors which facilitate or repress its transcriptional activity, respectively (Shang *et al*, 2002). Differential expression levels of AR pioneer factors and coregulators correlate with clinical outcome (Bohm *et al*, 2009; Sahu *et al*, 2011), implicating deregulation of the androgen-signaling axis in prostate cancer development and progression.

Androgen deprivation therapy (ADT) abrogates androgen signaling either through diminishing androgen synthesis or through competitive binding of the receptor, both resulting in a reduction of transcriptional activity of the AR. Patients with failed salvage therapy or metastatic prostate cancer are treated with ADT as a first-line palliative treatment (Heidenreich *et al*, 2014a). Moreover, adjuvant ADT improves the chances of cure for patients treated with radiotherapy (Pilepich *et al*, 2001; Bolla *et al*, 2009). Still, not all

- 1 Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
- 2 Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands
- 3 Division of Urology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
- 4 Division of Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
- 5 Department of Pathology, Josephine Nefkens Institute, Erasmus Medical Center, Rotterdam, The Netherlands
- 6 Department of Urology, Josephine Nefkens Institute, Erasmus Medical Center, Rotterdam, The Netherlands
- 7 Division of Medical Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

\*Corresponding author. Tel: +31 20 512 2101; E-mail: a.bergman@nki.nl

\*\*Corresponding author. Tel: +31 20 512 2101; E-mail: w.zwart@nki.nl

†These authors contributed equally to this study

radiotherapy-treated patients will benefit from adjuvant ADT (Roach, 2014). After prostatectomy, there is no conclusive evidence for benefit from adjuvant ADT (Zincke *et al*, 2001; Dorff *et al*, 2011; Miocinovic *et al*, 2011; Siddiqui *et al*, 2011; Briganti *et al*, 2012; Schubert *et al*, 2012; Tsurumaki Sato *et al*, 2014), but it is not unlikely that a subgroup of patients may benefit from adjuvant ADT after prostatectomy. Early identification of high-risk patients would be of substantial clinical relevance, so that these patients could be monitored more closely.

D'Amico *et al* (1998) developed a classification system based on clinical parameters (PSA, Gleason and clinical staging) to group men in low, intermediate, and high risk of relapse after therapy with curative intent. Limitation of this classification is the lack of integration with multiple risk factors and genomic data, which could provide more personalized risk assessment. Besides clinical risk stratifications, a number of different genomic classifications have been developed that enable the identification of high-risk patients (Irshad *et al*, 2013; Lalonde *et al*, 2014; Ramos-Montoya *et al*, 2014). However, no such genomic risk assessment biomarkers are currently adopted in routine clinical practice.

ADT can keep metastatic disease under control for several years, but practically all tumors eventually develop resistance to treatment. The majority of ADT-resistant tumors maintain active AR signaling, rendering this pathway a legitimate target for a second-line endocrine therapy (Valenca *et al*, 2015). Numerous novel anti-androgens and androgen depleting agents are being introduced into the clinic, including enzalutamide (MDV3100), ARN-509, and the CYP17 inhibitor abiraterone (Potter *et al*, 1995; Tran *et al*, 2009; Clegg *et al*, 2012). Unlike older anti-androgens (bicalutamide and flutamide) (Culig *et al*, 1999; Scher & Sawyers, 2005), the new generation of anti-androgens (e.g. enzalutamide) prevent nuclear translocation of AR and do not exhibit agonistic properties. But despite clinical implementation of these improved inhibitors of AR signaling, response is partial and temporal and tumors inevitably progress into a more aggressive and typically lethal form of prostate cancer (Antonarakis *et al*, 2014). Various mechanisms underlying resistance to abiraterone and anti-androgens are known, including AR overexpression, AR splice variants that confer ligand independent AR transactivation, and alterations in expression and recruitment of AR coregulators (Lamb *et al*, 2014).

AR chromatin binding and expression of AR-responsive genes were found to deviate between androgen-sensitive and androgen-resistant cell lines (Wang *et al*, 2009). Although AR binding profiles in cell lines have been studied extensively, the genomic behavior of AR in human prostate specimens remains largely understudied. AR chromatin binding profiles found in treatment-resistant prostate tumors were also observed in prostate cancer cell lines, and highlighted genes correlated with survival (Sharma *et al*, 2013). Recently, AR binding sites were identified that differentiated normal prostate tissue from cancer, which associated with the onset and progression of prostate cancer (Chen *et al*, 2015). Still, no thorough assessment of AR binding between primary versus resistant tumor specimens has been performed to date.

By comparing chromatin accessibility (Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)-seq) and AR chromatin binding profiles (Chromatin Immunoprecipitation (ChIP)-seq) in primary versus ADT-resistant tumors, we identified a distinct gene set that enables stratification of patients with prostate cancer on

outcome. With this, our study illustrates that progressive disease yields prognostic information in primary lesions and provides a prognostic gene signature to identify patients with prostate cancer at risk of metastatic relapse after local–regional treatment.

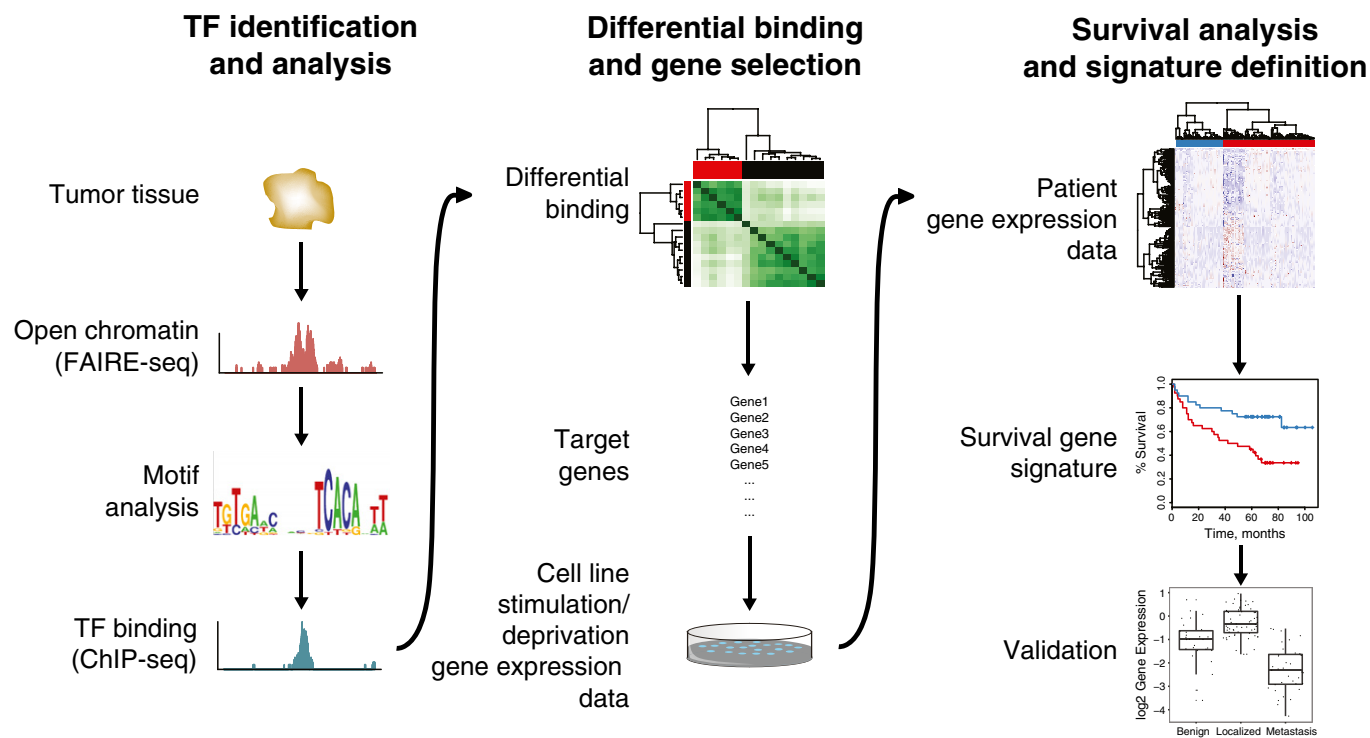
## Results

### Genomics-based pipeline for biomarker discovery and validation

By combining existing technologies, we here propose a genomics pipeline for biomarker discovery (Fig 1) and showed its application in prostate cancer, aimed at identification of prostate cancer patients with a high-risk of metastatic relapse. Firstly, transcription factor involvement was identified through motif analysis on open chromatin regions. Accessible regions were analyzed to reveal enrichment of a binding motif for a certain transcription factor involved in disease (prostate cancer). Actual transcription factor binding was mapped with ChIP-seq to identify sites that are differentially bound between two sample groups. As a proof-of-principle, we assessed AR chromatin binding profiles in this study. The target genes of the differential binding regions were subsequently coupled to gene expression data in cell lines to uncover genuine involvement of the transcription factor in expression of a distinct gene set. This gene set was subsequently tested for association with survival data of patients, and further refined into a minimal gene signature.

### Genomewide profiling of accessible chromatin regions in prostate tissues by FAIRE-seq

We assessed chromatin accessibility in multiple prostate tissue specimens as well as the changes thereof in prostate cancer development and progression. Four normal prostate tissue samples, four primary tumors, and three ADT-resistant prostate tumors were assessed, as well as three prostate cancer metastases (Fig 2A). FAIRE-seq was applied to identify accessible chromatin regions with gene-regulatory functions on a genomewide scale (Giresi & Lieb, 2009). FAIRE is based on phenol–chloroform mediated sample separation, in which accessible chromatin fragments can be separated from the condensed state, effectively enriching for regulatory genomic regions (schematically visualized in Fig 2A). Metastases and prostate adenocarcinomas contained more than 70% tumor cells with a Gleason score ranging from 7 (3 + 4) to 10 (5 + 5), while all normal prostate tissues were derived from a healthy region from prostatectomy specimens. Tumor and normal tissues were validated by our pathologists. Clinicopathological parameters are shown in Appendix Table S1. The number of FAIRE peaks identified was highly variable between the tissues, ranging from 50 peaks up to over 13,000 peaks (Appendix Table S2). Figure 2B shows four randomly selected representative coverage profiles of accessible chromatin at promoter regions. Over 50% of accessible chromatin sites in healthy and tumor specimens were found at promoter regions (Fig 2C), and average signal for each specimen showed clear enrichment of reads at transcription start sites (Appendix Fig S1). Tumor samples showed more enriched chromatin accessibility at both intron and distal intergenic regions, as opposed to normal prostate tissue where FAIRE signal was mainly found at promoters (Fig 2C).



**Figure 1. Genomics-based pipeline for biomarker discovery and validation.**

Tissue samples were processed for FAIRE-seq, and transcription factor (TF) motifs in open chromatin regions were analyzed. Selected transcription factor was mapped with ChIP-seq (in this case androgen receptor) to identify sites that are differentially bound between two sample groups. The target genes of the differential binding regions were coupled to gene expression and survival data and further refined into a minimal gene signature, which was validated in a number of gene expression datasets.

The number of sequenced reads as well as called peaks was lower in normal tissue as compared to any of the tumor samples: three out of four normal tissue samples (75%) had < 25 million reads sequenced, while nine out of ten tumor samples had above 25 million reads sequenced (90%;  $P = 0.04$ , Fisher's exact test), and three out of four normal tissue samples had < 1,000 peaks versus two out of ten tumor samples ( $P = 0.09$ , Fisher's exact test) (Appendix Table S2). This suggested presence of more condensed chromatin in normal prostate tissue compared to tumor. To confirm this, we compared normalized read count (reads per kilobase per million (RPKM)) in peaks found in at least three FAIRE-seq samples ( $N = 3,010$ ). Read counts were significantly lower in normal tissue as compared to any of the tumor stages ( $P < 2.2e-16$ , paired  $t$ -test) (Fig 2D). Normalized read count in advanced disease was lower than in primary prostate cancer ( $P < 2.2e-16$  for both resistant and metastatic tissue), but still higher than in normal tissue (Fig 2D).

To further focus on the differences in tumorigenesis, we considered peaks found in at least two out of four normal tissue samples and two out of four primary tumors. A much larger number of accessible chromatin regions are found in primary prostate cancer compared to normal tissue (Fig 2E). This corresponds to a higher raw signal in primary and other tumor stages compared to normal tissue (Fig 2F).

To identify transcription factors potentially involved in prostate cancer development, we performed motif enrichment analysis on chromatin regions selectively accessible in tumors. In addition, we analyzed the overlap of accessible FAIRE sites with ChIP-seq data

from a multitude of transcription factors in various cells using the ReMap annotation tool (Griffon *et al*, 2015). As expected, multiple motifs of transcription factors involved in prostate cancer pathobiology were found, including ERG (Gasi Tandefelt *et al*, 2014), AR (Lonergan & Tindall, 2011), and its pioneer factor FOXA1 (Robinson & Carroll, 2012) (Table EV1). The motifs are generally conserved between the individuals (Appendix Table S3). Using the ReMap tool, we found overlap of FAIRE sites with ChIP-seq signal from a number of factors from which motifs were found in FAIRE-seq data (e.g. AR, ERG, SP1, ETS1, and others) (Appendix Fig S2). Moreover, at these FAIRE regions a weak but statistically significant correlation was found between motif enrichment and ChIP-seq overlap from the ReMap tool (Spearman  $\rho = 0.30$ ,  $P = 0.02$ ). CTCF was found to be an outlier with high score of motif and low degree of overlap, while AR, MYC, and ERG showed moderate-to-high overlap and motif score.

Subsequently, mRNA expression and DNA binding for multiple of these transcription factors were assessed in prostate tumors and cell lines. For AR, CTCF, ERG, FOXA1, ETV1, and NKX3-1, mRNA expression was confirmed in four primary tumors (Fig EV1A). Using publically available ChIP-seq data from LNCaP and VCaP cells, we could illustrate these factors occupying the FAIRE-seq regions in the tumors (Fig 2G). As expected, AR binding was enriched at the enhancer-associated accessible regions, while CTCF occupied both enhancers and promoters (Taslim *et al*, 2012). For ERG and CTCF, we further validated binding at a subset of these regions in primary tumor specimens by ChIP-qPCR (Fig EV1B-E). Functional

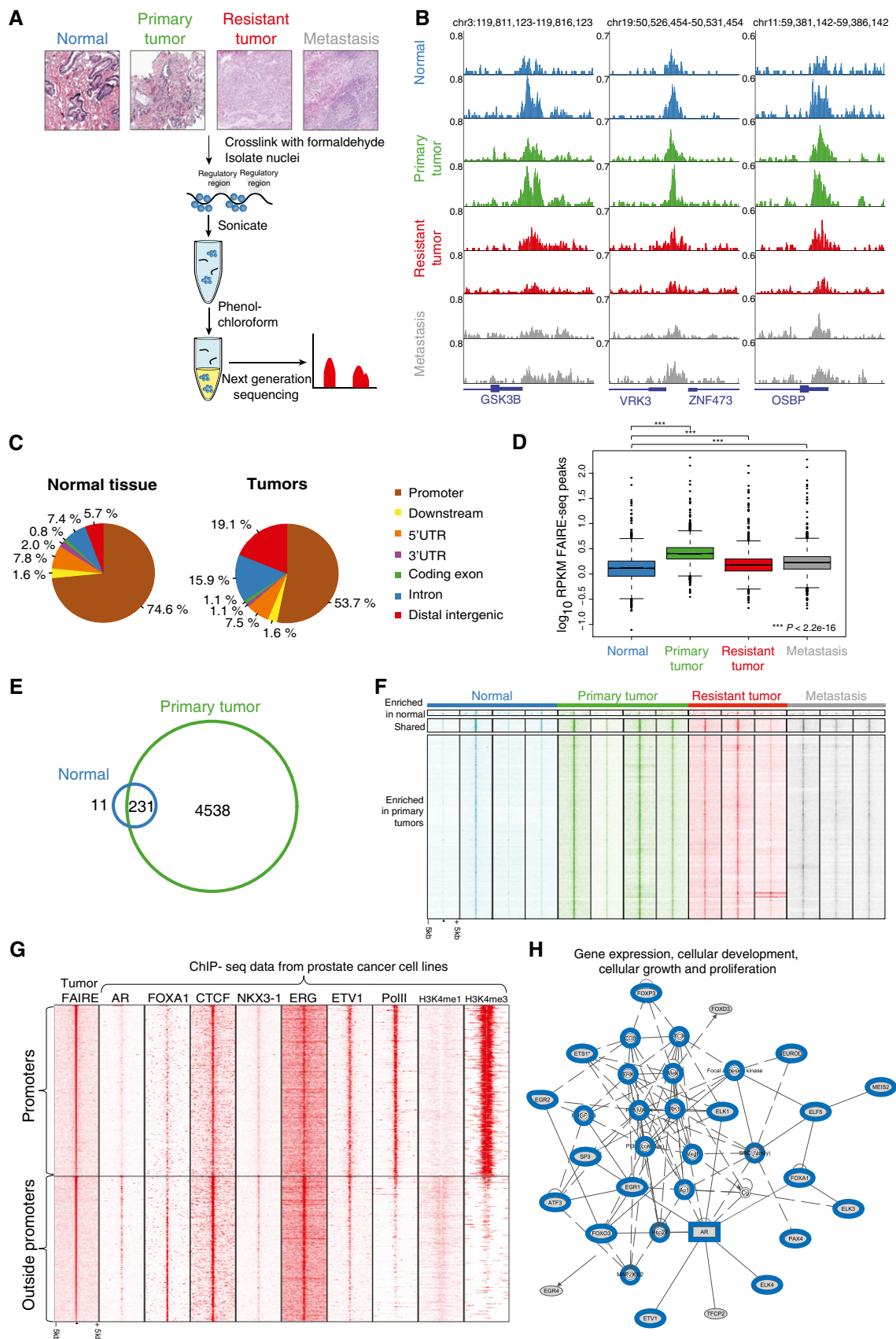


Figure 2.

**Figure 2. Genomewide profiling of chromatin accessibility in prostate cancer specimens.**

- A Overview of the experimental design. Representative examples of H&E-stained slides are shown. DNA and proteins were cross-linked using formaldehyde, followed by shearing of the chromatin and phenol–chloroform extraction. The organic phase contains compacted DNA (protein–DNA complexes), while DNA recovered from the aqueous phase represents accessible regulatory regions. DNA from the aqueous phase is further purified and sequenced.
- B Snapshots of accessible chromatin regions as assessed through FAIRE-seq in normal prostate tissue (blue), primary tumor (green), treatment-resistant tumor (red), and metastatic (gray) tissue. Reads are normalized to millions of sequenced reads per sample. Genomic coordinates are indicated.
- C Genomic distribution of FAIRE peaks in normal and tumor samples.
- D Boxplots depicting normalized FAIRE-seq read counts (RPKM) in different tissues across the peaks found in at least three samples. Read counts in benign tissue (blue) are lower than in primary tumor (green), therapy-resistant tumor (red) or metastasis (gray) ( $P < 2.2e-16$ , paired *t*-test).
- E Venn diagram, visualizing shared and unique FAIRE peaks in normal prostate samples (blue) and primary prostate tumor samples (green).
- F Heatmap showing raw read count intensity in FAIRE-seq peaks enriched in either normal or tumor samples. The window represents 5 kb around the FAIRE-seq peak.
- G Heatmap showing raw read count intensity of ChIP-seq signal from multiple cell line datasets (Appendix Table S12 for references and GEO accession numbers) at accessible regions identified in primary tumors (peaks present in at least two specimens) ranked on peak intensity. Top panel depicts promoter regions, and the bottom panel, all other regions. The window represents 5 kb around the FAIRE-seq peak.
- H Ingenuity pathway analyses, illustrating one of the networks based on motifs found in FAIRE-seq peaks that were present in at least two out of four primary tumors. Genes previously described to be involved in prostate cancer are highlighted in blue (other networks in Appendix Fig S3).

involvement of AR was further evidenced by ingenuity pathway analysis using the list of genes corresponding to the motifs identified in FAIRE-seq peaks of the primary tumors (Table EV1). This analysis yielded functional networks known to be involved in prostate cancer, with one network centered around AR (Fig 2H, Appendix Fig S3).

### Distinct genomewide AR binding pattern in primary and resistant tumors

Since AR and its interactors were enriched at FAIRE-seq regions (Fig 2G), we next performed AR ChIP-seq on five primary and three treatment-resistant tumor specimens (Fig 3A). The number of reads and AR binding events is shown in Appendix Table S4. The total number of identified AR binding events greatly varied between the tumor samples (Appendix Table S4), which is consistent with previous nuclear receptor ChIP-seq in prostate and breast tumor samples (Ross-Innes *et al*, 2012; Jansen *et al*, 2013; Sharma *et al*, 2013). Specifically, our AR binding sites varied from 238 up to 17,511 per tumor sample, which is in the same order of magnitude as in Sharma *et al*, with 300–8,500 per tumor sample (Sharma *et al*, 2013).

Unlike in FAIRE data, the majority (over 90%) of AR ChIP-seq peaks are present either in intronic or distal intergenic regions throughout the genome (Figs 2C and 3G), consistent with profiles identified in cell lines (Yu *et al*, 2010; Asangani *et al*, 2014) and tumors (Sharma *et al*, 2013). Intra-tumor heterogeneity effect was limited, as assessed on two independent sections from the same tumor specimen, with 87% overlap of AR binding sites between the biological replicates and high correlation of peak read counts ( $r = 0.76$ ) (Appendix Fig S4). Overlap of AR ChIP-seq replicates was comparable to previously described cell line data (Bolton *et al*, 2007; Jia *et al*, 2008). AR binding sites selectively enriched in resistant or primary tumors could be observed (Fig 3B), as validated by AR ChIP–qPCR analyses on additional four primary and five resistant tumors (Fig 3C). Ratios of average enrichment for “resistant enriched” over “primary tumor enriched” regions were determined, analogous to what we performed before for breast cancer samples (Jansen *et al*, 2013), showing consistent subclassification of patients in the validation set.

To assess differential AR chromatin binding on a global scale between treatment-resistant and primary tumors, differential binding analysis was performed. Peaks present in at least three tumors

( $N = 3,138$ ) were considered. In total, 339 genomic regions show differential AR binding between primary and resistant tumors (for genomic regions, see Table EV2). Tumors clustered according to their group identity (primary or resistant) based on these regions (Fig 3D). Differential AR chromatin binding at these sites is further evidenced by the difference in average read counts between the primary and resistant tumors (Fig 3E). To assess the reliability of differential AR binding events, the samples were shuffled randomly based on 56 available permutations. Few or no differential peaks were found when the samples were mislabelled, demonstrating the robustness of classification and no overfitting (Appendix Fig S5). To assess whether the identified differential peaks were also relevant in other clinical datasets, we clustered our data with AR ChIP-seq data of primary tumors from others (Chen *et al*, 2015). Three out of four primary tumors from the other dataset clustered together with the primary tumors from our study using the 339 identified AR sites (Fig EV2), suggesting a wider applicability of the differential AR binding signature.

Differential enriched AR peaks are located mostly in intronic or distal intergenic regions of the genome (Fig 3F), and primary tumor-associated AR sites are enriched on chromosomes 16 and 19 (Fig 3G). In contrast, resistance-associated AR binding sites are enriched on chromosomes 3, 7, and 8, while AR binding at chromosomes 6, 12, and 17 is diminished. The peaks remain enriched in treatment-resistant samples even when the background is subtracted (Appendix Fig S6).

Motif analysis was performed on all AR binding sites (present in at least two tumors), resistance-associated AR binding sites and primary tumor-associated AR sites. As expected, AR binding sites were enriched for AR motifs, as well as for its pioneer factor FOXA1 (Fig 3H, Appendix Fig S7, Table EV3). Motif enrichment of the AR coregulator HOXB13 was found selectively in peaks enriched in resistant tumors (Norris *et al*, 2009).

### Genes proximal to altered AR binding sites are AR-responsive in sensitive and resistant cell lines

We identified distinct AR binding regions between treatment-resistant and primary prostate tumors. Possible direct target genes were considered as those with AR binding sites within their body or 20 kb upstream from the transcription start site (Wang *et al*, 2009), yielding 158 genes (Table EV4). The limited number of target genes (158) relative to the 339 AR binding sites can be explained by the

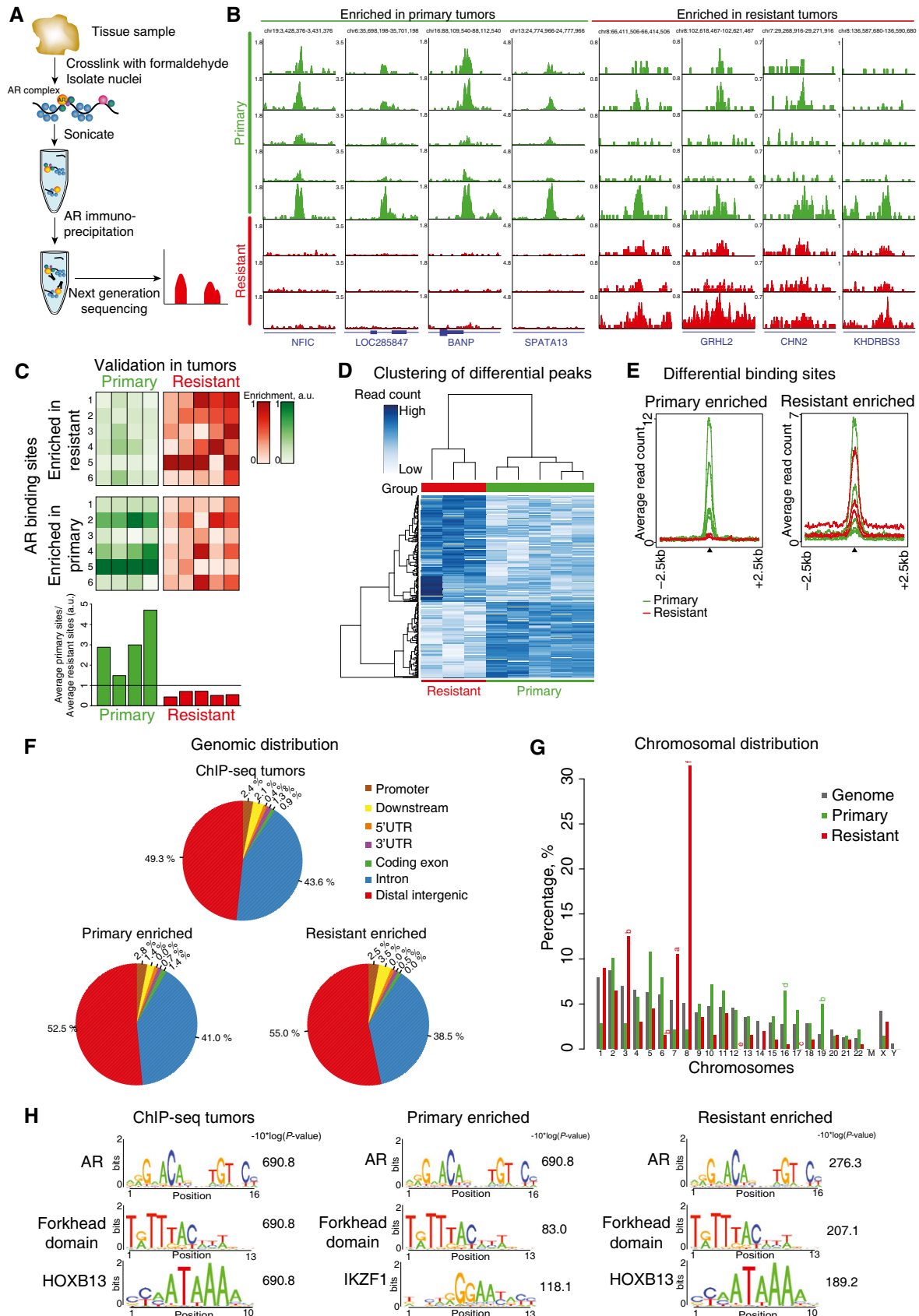


Figure 3.

**Figure 3. Distinct AR binding profiles in primary and treatment-resistant tumors.**

- A Overview of experimental design. DNA and proteins were cross-linked using formaldehyde, followed by shearing of the chromatin. Immunoprecipitation for AR was performed after which isolated DNA fragments were sequenced.
- B Snapshots of AR binding events differentially enriched in either primary (green) or resistant (red) tumors. Reads are normalized to millions of sequenced reads per sample. Genomic coordinates are shown.
- C Heatmap illustrating AR ChIP-qPCR validation in independent tumors. Intensity of the color corresponds to ChIP-qPCR enrichment (see scale bar). Binding sites identified as enriched in either primary or treatment-resistant tumors were tested. Average enrichment in primary tumors was divided over average resistant enrichment values to determine the ratio, as is visualized in a barplot (bottom panel).
- D Differential binding analysis of AR chromatin binding regions, selectively enriched in primary tumors (green) or treatment-resistant (red) tumors.
- E Average read counts for AR binding events, selectively enriched between primary (green) or resistant (red) tumors. Data are centered at AR peaks, depicting a 2.5-kb window around the peak.
- F Genomic distributions of AR binding sites shared between tumors or enriched in primary and resistant tumor tissue.
- G Distribution of AR binding sites enriched in primary (green) and resistant (red) tumors, by chromosome (%). *P*-values (one-sided binomial test) for significant enrichment relative to the entire genome (gray): (a)  $P = 0.001$ , (b)  $P = 0.002$ , (c)  $P = 0.003$ , (d)  $P = 0.005$ , (e)  $P = 9e-13$ , and (f)  $P = 2.5e-13$ .
- H Motif enrichment analysis for AR binding sites shared between tumors or enriched in primary and resistant tumor tissue. Top motifs are shown.

functional involvement of multiple AR binding sites to regulate the expression of a single gene, as illustrated at the *KLK3* locus (Appendix Fig S8).

Since genes were identified by alterations of AR binding between primary and resistant tumors, these genes could be AR-responsive in either an androgen-sensitive or androgen-resistant context. To dissect these two possibilities, expression of the 158 putative target genes was explored using a publically available dataset from LNCaP prostate cancer cells treated for up to 24 h with androgen R1881, with samples taken every few hours (Fig 4A) (Massie *et al*, 2011). A total of 102 genes out of 158 putative AR target genes were found in the androgen stimulation dataset, represented by 117 expression probes. In order to group genes according to their temporal profiles, we clustered the gene expression data using the Short Time-series Expression Miner (STEM) program. This software utilizes an algorithm that selects potential temporal profiles, assigns genes to those profiles and computes enrichment significance for each profile (Ernst *et al*, 2005; Ernst & Bar-Joseph, 2006). Three significant gene sets were found: I. upregulated in response to R1881 (41 genes); II. not affected by R1881 (49 genes); and III. downregulated in response to R1881 (12 genes) (Fig 4B and C, Table EV5). The 49 genes of gene set II were of particular interest, since genes not affected by R1881 treatment could still be associated with acquired ADT resistance. Therefore, expression of these genes was next assessed in a public gene expression dataset from LNCaP cells cultured in charcoal-treated hormone-deprived medium for up to 1 year, mimicking acquired ADT resistance (Fig 4D and E) (D'Antonio *et al*, 2008). Out of 49 genes, 48 were annotated in this dataset, corresponding to 164 probes. While not affected by R1881 treatment (Fig 4B and C), these genes exhibit differential expression during acquisition of ADT resistance (Fig 4E). Expression patterns of gene sets I and III under androgen deprivation conditions are shown in Appendix Fig S9.

Cumulatively, by integrating our AR ChIP-seq data with publically available datasets of hormonal response and acquired resistance to ADT, we identified two distinct gene sets associated with “AR response to hormonal stimuli” and “AR response in acquired resistance.”

**Functional and clinical implications of differential AR-binding-affected genes**

The AR binding landscape in tumors revealed two clearly defined gene sets, one subset of 53 androgen-responsive genes and the second subset of 49 acquired ADT resistance genes.

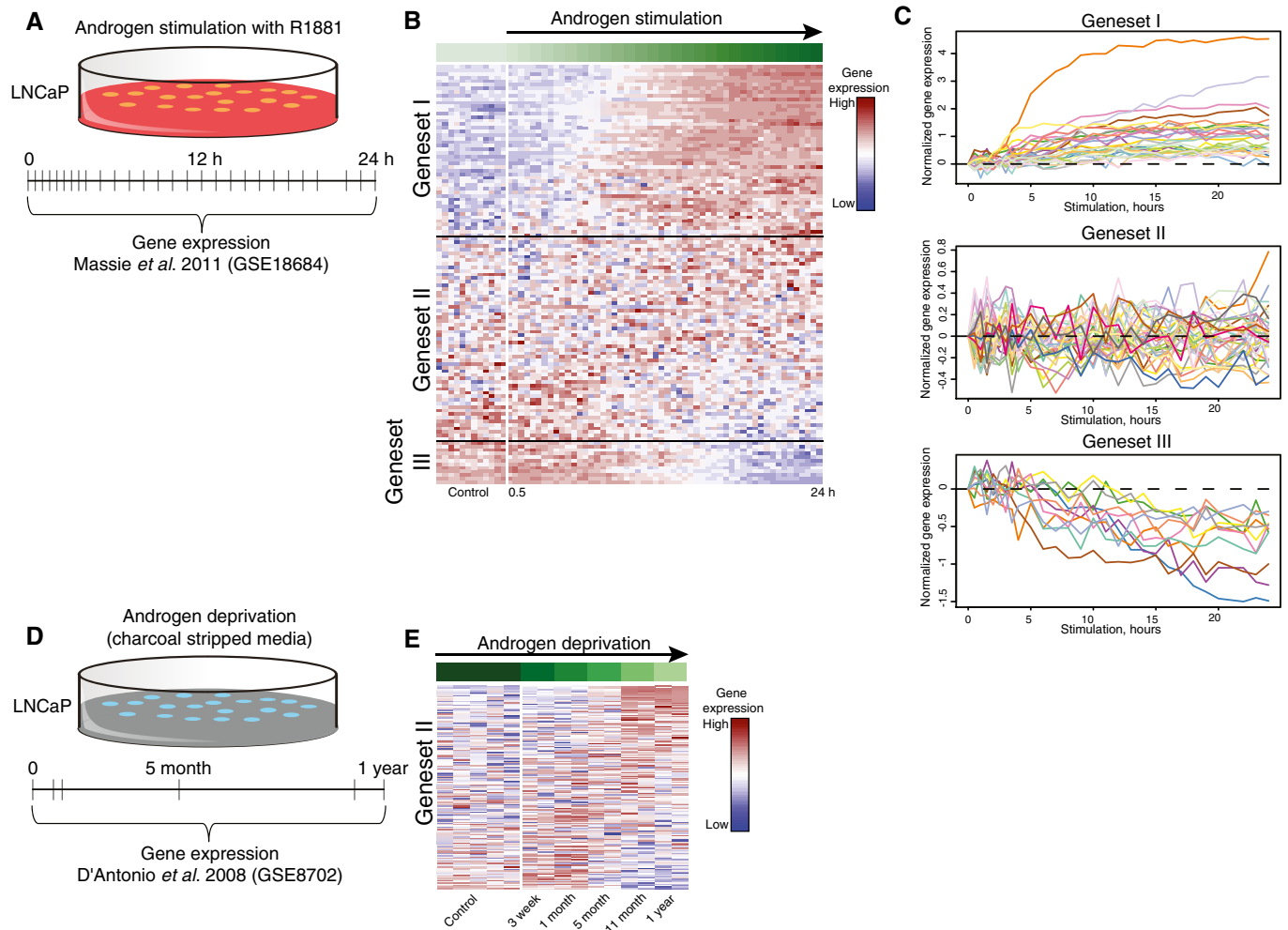
To couple “androgen-responsive” and “acquired resistance” genes with outcome, gene expression data from 131 primary prostate cancers were analyzed (Taylor *et al*, 2010). Unsupervised hierarchical clustering was performed on the 53 “androgen-responsive” and 49 “acquired resistance” genes separately. The “androgen-responsive” gene set failed to stratify patients on time to biochemical (PSA) relapse ( $P = 0.931$ , HR = 1.03; 95% CI: 0.47–2.26) (Fig 5A and B), while clustering on “acquired resistance” genes did stratify patients ( $P = 0.032$ , HR = 0.45; 95% CI: 0.21–0.95) (Fig 5C and D).

Ingenuity pathway analysis illustrated the “androgen-responsive” genes to be involved mainly in cellular metabolic processes and correspond to genes normally regulated by steroid hormone receptors AR and PGR (Fig 5E). The “acquired resistance” genes were additionally involved in androgen biosynthesis and contain genes related to prostatic carcinoma (Fig 5E).

**Refinement and validation of gene expression classifier**

We identified a set of 49 “acquired resistance” genes that stratify patients with prostate cancer on outcome. Since this list is likely to contain false-positive genes that may jeopardize classification, we identified a core gene set with a prognostic impact through elastic net regularization with double-loop cross-validation. Patients were assigned to two risk groups based on their cross-validated prognostic index (Fig EV3), successfully stratifying patients on time to relapse ( $P = 0.00084$ , log-rank test) (Fig 6A). The refined gene set classifier is composed of nine genes: *DNER*, *EXT2*, *AMOTL1*, *RBM33*, *ZBTB20*, *XBP1*, *PMFBP1*, *HSD17B14*, and *KLF9*. Apart from AR (which was differentially enriched proximal to these genes; Fig 3), other transcription factors may regulate expression of these nine prognostic genes as well. Therefore, we investigated the overlap of the FAIRE-seq peaks (Fig 2) in close proximity of the nine genes (< 20 kb from the transcription start site) with a large collection of public ChIP-seq data from the ReMap tool (Griffon *et al*, 2015). A strong overlap of the identified eight FAIRE-seq peaks was found with known players in prostate cancer including AR, FOXA1, ERG, bromodomains BRD2 and BRD3 and, interestingly, MYC (Appendix Table S5). The latter also occupies a central place in our functional motif analysis (Appendix Fig S3).

A prognostic index was assigned to each patient based on the nine genes, calculated as the sum of expression of the nine genes multiplied by their corresponding Cox regression coefficients (Appendix Table S6). This prognostic index is independent from other known prognostic parameters, such as Gleason score,



**Figure 4. Genes with differential AR binding site are associated with androgen response or androgen resistance.**

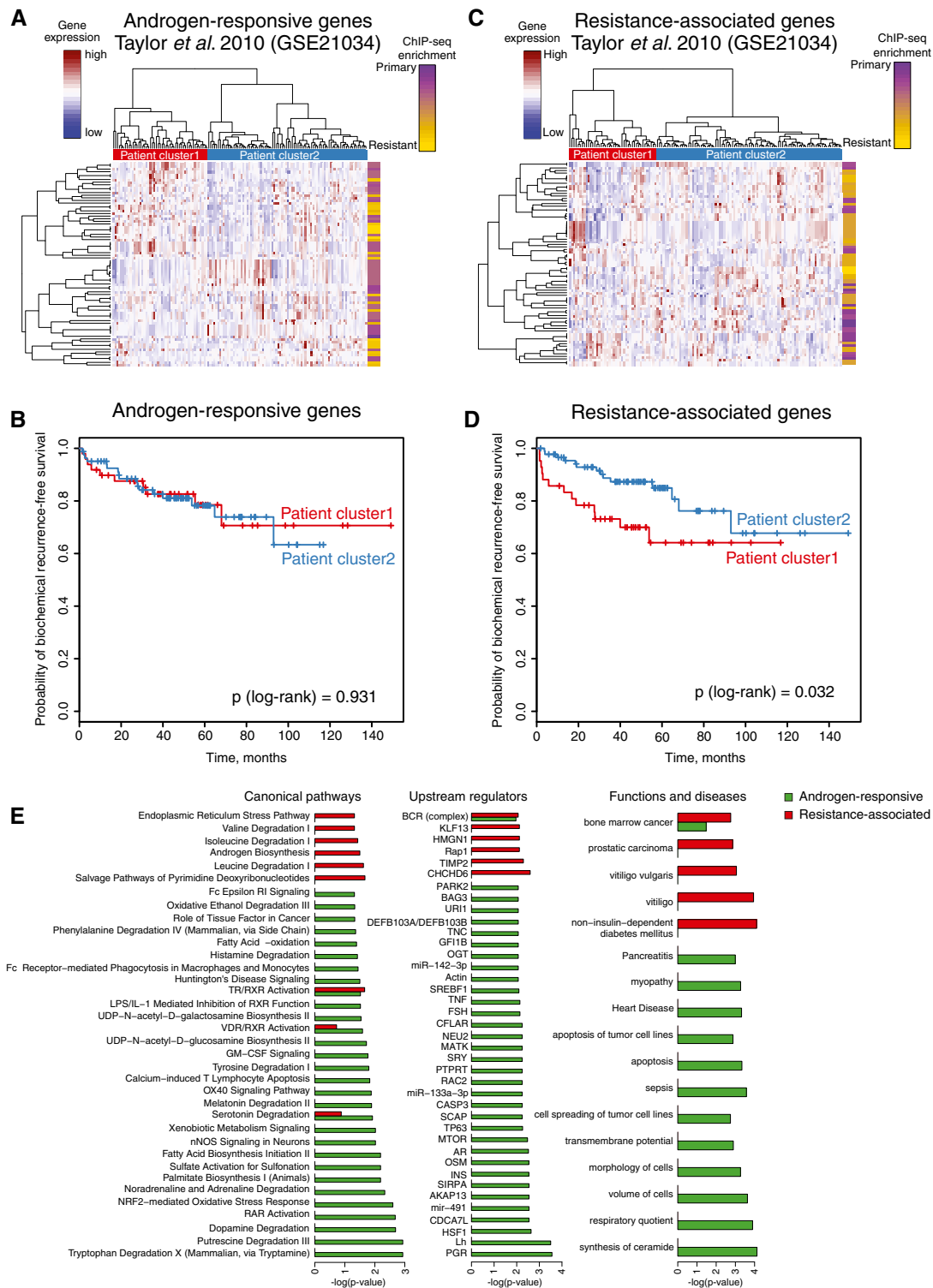
- A Visual representation of the public dataset used (GSE18684; Massie *et al*, 2011). Hormone-deprived LNCaP prostate cancer cells were treated for up to 24 h with 1 nM R1881 and processed for gene expression analyses.
- B Expression of 102 genes with a proximal AR binding site, differentially enriched between primary and resistant tumors, was analyzed. Three subclusters were identified (I, II, III), based on Short Time-series Expression Miner (STEM) clustering analysis. Color scale indicates gene expression level.
- C Line plots showing expression profiles of the 102 genes in three STEM-based regulatory modules obtained from the androgen stimulated LNCaP expression data as in B.
- D Visual representation of the public dataset used (GSE8702; D'Antonio *et al*, 2008). LNCaP prostate cancer cells were grown in the absence of androgens for up to 1 year and processed for gene expression analysis.
- E Heatmap illustrating gene expression of 48 androgen-nonresponsive genes from gene set II in hormone-deprived LNCaP cells (GSE8702).

pathologic T stage, lymph node status, and initial PSA (see Appendix Table S7 for patient characteristics), and is significantly associated with time to biochemical recurrence when adjusted for those known prognostic factors (Appendix Table S8). Furthermore, we tested the ability of the classifier to identify patients that develop biochemical recurrence within 5 years after prostatectomy using receiver operating characteristic (ROC) curves. The area under the curve (AUC) value for the 9-gene classifier was slightly higher as compared to an AUC value based on clinical parameters only (0.86 versus 0.83), while combining the genomic classifier with clinical parameters resulted in an AUC value of 0.9 (Appendix Table S9).

One of the most widely used prostate cancer risk assessment classification systems was proposed by D'Amico and colleagues

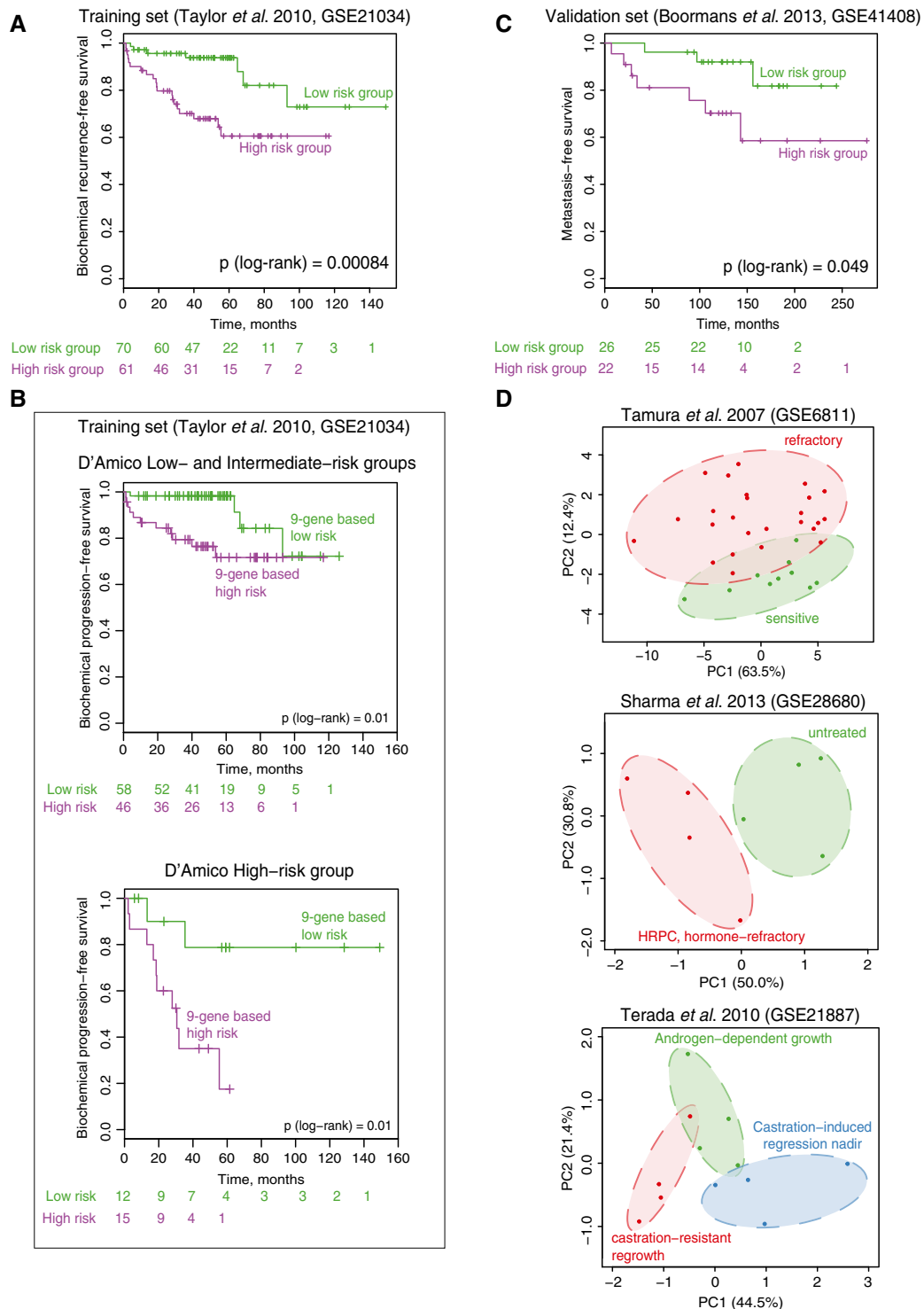
(D'Amico *et al*, 1998). It utilizes clinical TNM stage, preoperative PSA level, and biopsy Gleason score to stratify patients in three risk groups (low-, intermediate-, and high-risk groups). To assess whether our 9-gene classifier is independent from the D'Amico classification, we directly compared the output of D'Amico classification and our 9-gene signature. Low- and high-risk patients as defined by the genomic signature were found in all three D'Amico risk groups (Appendix Table S10), and the prognostic potential of our risk group stratification was independent of D'Amico classification (Appendix Table S11). Since there was no difference in survival between D'Amico low- and intermediate-risk groups in patients from Taylor *et al*'s cohort (Fig EV4), we combined low- and intermediate-risk groups together. The 9-gene expression signature was able to further





**Figure 5. Unsupervised hierarchical clustering and survival analysis based on expression of the selected groups of genes in 131 patients with primary prostate cancer.**

- A Heatmap of unsupervised hierarchical clustering of 131 patients based on the expression of androgen-responsive gene set.
- B Kaplan–Meier curves of biochemical recurrence-free survival of the two groups of patients identified based on expression of androgen-responsive genes.
- C Heatmap of unsupervised hierarchical clustering of patients based on the expression of resistance-associated gene set.
- D Kaplan–Meier curves of biochemical recurrence-free survival of the two groups of patients identified based on expression of resistance-associated genes.
- E Ingenuity pathway analysis of the two groups of genes (androgen-responsive and resistance-associated). Barplots show significantly enriched pathways, functions, and potential upstream regulators.



**Figure 6. Refinement and validation of gene expression classifier for the survival of patients with prostate cancer.**

- A Kaplan–Meier survival curves of the patients stratified into two risk groups based on the prognostic index assigned in cross-validated training procedure in the training cohort (Taylor *et al.*, 2010).
- B Kaplan–Meier survival curves for patients stratified based on the 9-gene classifier within low/intermediate- and high-risk groups based on clinical parameters (D'Amico classification).
- C Validation of the prognostic signature in an independent patient set (Boormans *et al.*, 2013). Kaplan–Meier survival curves for patients stratified based on the 9-gene prognostic index.
- D Principal component analysis score plot based on the expression of the 9-gene signature in three independent expression datasets with hormone-sensitive and hormone-resistant tumors and xenografts.

stratify patients in both low/intermediate- and high-risk groups (Fig 6B), which further highlights the utility of our 9-gene signature on top of existing classification tools.

For validation of the prognostic signature, we used an independent gene expression dataset from a cohort of 48 patients with prostate cancer (Boormans *et al*, 2013). We applied the coefficients from the trained Cox regression to calculate the prognostic index in the validation set and successfully stratified patients into two risk groups with significant difference in time to metastasis ( $P$  (log-rank) = 0.049) (Fig 6C). Furthermore, we examined two other independent gene expression clinical studies as well as a set of prostate cancer xenografts. We used principal component analysis (PCA) to explore whether expression of the nine genes can separate clinical groups in these datasets, providing new dimensions (principal components) that summarize expression of the nine selected genes. Hormone-sensitive and hormone-refractory tumors and xenografts are separated along either the first or second principal components (Fig 6D). With this, we provide evidence that our classifier does not only stratify patients on biochemical relapse (Fig 6A and B) and metastasis (Fig 6C), but also successfully makes a distinction between hormone-refractory and hormone-sensitive tumors (Fig 6D).

Since expression of selected genes is associated with outcome, similar expression differences may be found in metastatic samples. Therefore, we examined expression levels of the nine individual genes in six additional cohorts (Appendix Fig S10, Appendix Table S12) that include benign (normal), primary, and/or metastatic prostate cancer samples. Directionality of gene expression in poor-outcome patients in our discovery set was identical as found in metastatic samples as compared to primary lesions for all six independent cohorts, indicating robustness and consistency (Appendix Fig S10). These data indicate that potential drivers of poor outcome, as identified in primary prostate cancers, are possibly preserved in the metastatic setting.

## Discussion

According to current clinical guidelines, endocrine therapy for prostate cancer is prescribed as an adjuvant treatment after radiotherapy and in the metastatic setting with a palliative intent (Heidenreich *et al*, 2014b). The rationale for not applying endocrine therapies in the adjuvant management of prostate cancer after prostatectomy is provided by multiple clinical trials, which illustrated limited to no clinical benefit of blocking AR function on disease-free survival of the entire population (Zincke *et al*, 2001; Dorff *et al*, 2011; Miocinovic *et al*, 2011; Siddiqui *et al*, 2011; Briganti *et al*, 2012; Schubert *et al*, 2012; Tsurumaki Sato *et al*, 2014). Yet, since ~30% of patients with prostate cancer do develop a relapse later in life, it is believed that a distinct subgroup of patients may derive benefit from ADT or other subsequent treatment in the adjuvant setting. Above all, patients with high risk of relapse are in need of closer monitoring postoperatively. To date, no genomic biomarkers are clinically applied that may aid in identifying high-risk patients.

Here, we combined existing approaches and technologies (FAIRE-seq, ChIP-seq, expression analyses, and survival data) as a potential “genomics pipeline” for biomarker discovery. FAIRE-seq analyses in prostate cancer specimens led to the identification of a

large set of transcription factor motifs, including AR and multiple of its interaction partners. As a proof-of-principle, we studied the genomic behavior of AR. General applicability of such a genomic pipeline remains to be determined, and future studies should be aimed at assessing genomic features of other hits from the motif analyses to illustrate whether this approach is also applicable for other factors apart from AR.

We illustrate that prognostic biomarkers for the survival of patients with prostate cancer can be identified through the assessment of AR/chromatin interaction landscapes in tumor samples. By comparing AR binding patterns in primary tumor tissue specimens with those found in tumors that have an acquired resistance to treatment, a reprogramming of the AR interactome was observed. This reprogramming has far-reaching consequences, in which a distinct and unique gene set with acquired AR-responsive features provides prognostic potential for the survival of patients with prostate cancer, independent of classical prognostic parameters and clinical risk stratification system (D’Amico). Differential AR binding sites were clearly enriched for specific chromosomes, and a strong enrichment of resistance-associated AR binding sites was found at chromosome 8. Chromosome 8 has previously been implicated in prostate cancer progression, containing multiple oncogenes, including Myc (El Gammal *et al*, 2010). A selective enrichment of AR binding sites at distinct chromosomal regions could yield direct biologically relevant information on prostate cancer progression and may uncover drivers in ADT resistance.

Since the gene set identified in tumors with an acquired resistance yields prognostic potential in both cohorts of patients with primary prostate cancer, these data indicate that the AR-driven processes that were identified in the samples with an acquired resistance may already be active in the primary tumor, driving prostate tumor progression and metastasis formation. The expression of the acquired-resistance classifier in primary lesions may be used to identify patients with high risk of relapse after radical prostatectomy. Identification of high-risk patients is of clear clinical benefit, since it would enable closer monitoring of these patients. Since our classifier is based on differences between primary and resistant tumors, it remains to be determined whether these high-risk patients would be likely to respond to ADT or would be better treated with nonhormonal therapeutics.

AR is generally considered as the main driver in prostate cancer tumorigenesis and tumor progression. This notion is also further emphasized by the fact that all targeted therapies in prostate cancer treatment are directly aimed in a functional inhibition of AR activity. Nonetheless, analogous to recent finding in breast cancer (Robinson *et al*, 2011; Mohammed *et al*, 2015), other transcriptional regulators could potentially play a role in prostate cancer development and progression. AR ChIP-seq differential binding analysis yielded HOXB13 motif as selectively enriched in resistant tissue, suggesting a role of HOXB13 in resistant prostate cancer as found by others (Jeong *et al*, 2012). In addition, motif analyses on accessible chromatin regions by FAIRE-seq identified motifs for a large range of other transcription factors previously linked to prostate cancer development and progression, including CTCF, SP1, FOS, and ETS domain family of transcription factors (Shemshedini *et al*, 1991; Lu *et al*, 2000; Taslim *et al*, 2012; Chen *et al*, 2013). A subset of these transcription factors were previously defined as “druggable,” including ETS family members

ETV1 (Rahim *et al*, 2014) and ERG (Nhili *et al*, 2013) as well as specific protein (SP) transcription factors (Safe *et al*, 2014). Out of the entire nine-gene signature, the three transcription factors (XBP1, ZBTB20, and KLF9) are previously described in relation to prostate cancer, while the other genes are not. KLF9 has been suggested to play a role in the progression of prostate cancer to a castration-resistant stage (Shen *et al*, 2014). XBP1 on the other hand is highly expressed in primary tumor, while hormone-refractory tumors show weak XBP1 expression (Takahashi *et al*, 2002; Cuperlovic-Culf *et al*, 2010). Furthermore, XBP1 is represented in the 150-core gene set of Sharma *et al*, while ZBTB20 is found in the gene set of Chen *et al*, and both signatures were based on AR binding in prostate specimens (Sharma *et al*, 2013; Chen *et al*, 2015). All these proteins may be of particular interest for new treatment options in prostate cancer, and future studies are aimed to further elucidate the potential roles of these transcription factors in prostate carcinogenesis as well as their therapeutic potential in this setting.

In summary, using integrative genomics of FAIRE-seq, AR ChIP-seq, publically available transcriptomic data and patient survival data, we successfully determined a minimal gene signature for the outcome of patients with prostate cancer. Even though the prognostic potential of these genes is apparent, future clinical trials should determine whether these genes could be informative for selective response to treatment of prostate cancer.

## Materials and Methods

### Clinical samples

Fresh frozen postoperative prostate needle biopsies from normal (pathologically validated tumor-free region of the peripheral zone) and tumor samples were obtained from prostatectomy specimens at the Netherlands Cancer Institute (Amsterdam, The Netherlands). The androgen-blockade-resistant tumor samples [transurethral resection of the prostate (TURP)] and lymph node metastases were obtained from the Erasmus University Medical Center (Rotterdam, The Netherlands). The hematoxylin and eosin (H&E) slides of all these cases were reviewed by our pathologists. See Appendix Table S1 for clinicopathological parameters. This study was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands and has been approved by the local medical ethics committees.

### Formaldehyde-assisted isolation of regulatory elements

FAIRE experiments were performed as previously described (Giresi & Lieb, 2009). Briefly, fresh frozen tissues were cross-linked with 1% formaldehyde for 20 min. Cells were washed, and nuclei were isolated as described (Zwart *et al*, 2013). Subsequently, chromatin was sonicated and cleared by centrifugation. The soluble chromatin was subjected to three consecutive phenol–chloroform–isoamyl alcohol (25:24:1) extractions. The samples were reverse-cross-linked overnight at 65°C and purified by ethanol precipitation. Purified samples were treated with RNase A and proteinase K and repurified by PCR purification kit (Roche).

### RNA isolation and mRNA expression

Total RNA was isolated after treatment with RNA-Bee reagent (Tel-Test, Inc.), and cDNA was synthesized from 500 ng RNA using SuperScript III Reverse Transcriptase (Invitrogen) with random hexamer primers. qPCR was performed with SYBR Green (GC Biotech) on a Roche LightCycler. Primer sequences for mRNA expression analysis are listed in Appendix Table S13.

### Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitations were performed as described before (Jansen *et al*, 2013; Zwart *et al*, 2013). A total of 10 µg (ChIP-seq) or 5 µg (ChIP-qPCR) of antibody was used, with 100 µl (ChIP-seq) or 50 µl (ChIP-qPCR) of Protein A magnetic beads (Invitrogen). Antibodies used were AR-N20 (sc-618; Santa Cruz), CTCF (07-729; Millipore), and ERG (sc-353; Santa Cruz). Primer sequences for qPCR analysis are in Appendix Table S13.

### Solexa sequencing and enrichment analysis

DNA was amplified as described (Jansen *et al*, 2013). Sequences were generated by the Illumina HiSeq 2000 Genome Analyzer (using 50-bp reads) and aligned to the Human Reference Genome (assembly hg19, February 2009). Reads were filtered based on MAPQ quality: Only reads with MAPQ above 20 were considered to eliminate reads from repetitive elements. Enrichment over input control was determined using DFilter (Kumar *et al*, 2013) and MACS peak caller version 1.4 (Zhang *et al*, 2008). Only peaks called by both algorithms were used for the analysis. MACS was run with the default parameters, except  $P = 10^{-7}$  for ChIP-seq data. DFilter was run with  $bs = 100$ ,  $ks = 50$  for FAIRE data and  $bs = 50$ ,  $ks = 30$ , refine, nonzero for ChIP data. Read counts, number of aligned reads, and number of peaks are shown in Appendix Tables S2 (FAIRE) and S4 (ChIP).

The raw and processed data are deposited in the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>; accession No. GSE65478).

### Data analysis

Motif analyses were performed using Cistrome (cistrome.org), applying the SeqPos motif tool with default settings (region width: 600 bp;  $P$ -value cutoff: 0.001). Genomic distributions of binding sites were analyzed using the *cis*-regulatory element annotation system (CEAS). Differential binding analysis (DBA) was performed as described (Ross-Innes *et al*, 2012), considering all peaks that were found in at least three specimens, without control read subtraction and using a false discovery rate (FDR) below  $< 0.10$ . For integration with gene expression data, binding events were considered proximal to the gene when identified in a gene body or within 20 kb upstream from the transcription start site. Publically available ChIP-seq data for a number of factors in prostate cancer cell lines were used, and details are summarized in Appendix Table S12.

### Expression analyses and survival data

Time-series gene expression was analyzed using Short Time-series Expression Miner (STEM) (Ernst & Bar-Joseph, 2006) to identify

gene sets that show significant temporal expression profiles. Default parameters were used, except for the minimum expression change set to 0.5 (based on the difference from 0). Two datasets of gene expression in LNCaP prostate cancer cells were used: GSE18684 and GSE8702. Details of the publically available gene expression datasets that were used in this study (GSE21034, GSE3933, GSE35988, GSE3325, GSE32269, GSE29079, GSE41408, GSE6811, GSE28680, GSE21887) are summarized in Appendix Table S12. These datasets were selected based on the availability of gene expression data from frozen tissue, relatively large sample sizes, measured on comprehensive gene expression platforms and clearly defined clinical groups. Clinical parameters (where available) of the gene expression datasets are summarized in Appendix Table S7. Cox regression regularized by an elastic net penalty from the glmnet package in R (Friedman *et al*, 2010) was used in order to select the optimal set of genes that are significantly associated with survival. Average centered probe levels were used. Double cross-validation was performed to assess the overall predictive ability of the procedure. In the outer loop, leave-one-out cross-validation was used. In the inner loop, regularization parameter lambda of the elastic net and the set of active covariates were determined with 10-fold cross-validation. Then, Cox regression was fitted with the selected variables and prognostic index estimated for the left-out sample. The final set of covariates for constructing Cox model was selected based on the average lambda value. In the validation cohort, prognostic index for each patient was defined as a sum of gene expression values multiplied by their corresponding coefficients derived from the Cox regression. Patients with prognostic index below or equal to zero were assigned to the low-risk group, while patients with the prognostic index above zero were assigned to the high-risk group. Differences in survival were assessed using log-rank test. Prognostic index was also used as a covariate in multivariate Cox regression analysis. The clinical parameters used in the multivariate analyses included pathologic Gleason score, T stage, lymph node status, and pretreatment PSA.

**Expanded View** for this article is available online:  
<http://emmm.embopress.org>

### Acknowledgements

The authors thank Movember, grant number NKI01, and KWF/Alpe d'HuZes for financial support. We thank the NKI Genomics Core Facility for sequencing analyses and the Core Facility Molecular Pathology and Biobanking for help with tissue access.

### Author contributions

This study was designed by SS, EN, AMB, and WZ. Experiments were performed by SS, and data analyses were performed by SS, EN, and LFAW. HGvdP, JdJ, GJLHVl, AMB, and GJ provided samples and clinical information. Project supervision was performed by WZ, AMB, and LFAW. The manuscript was written by SS, EN, AMB, and WZ with the help of all other authors.

### Conflict of interest

The authors declare that they have no conflict of interest.

### For more information

<http://www.nki.nl/divisions/molecular-pathology/zwart-w-group/>

### The paper explained

#### Problem

Prostate cancer is a common disease, with over 1 million newly diagnosed cases and about 250,000 deaths per year. Currently, no optimal prognostic biomarkers are available for application in daily clinical practice. Such biomarkers would be of profound clinical benefit, since they would enable high-risk patients to be monitored more closely.

#### Results

In search for novel prognostic biomarkers in prostate cancer, we started with an unbiased assessment of accessible chromatin in prostate specimens. Healthy human prostates were compared with prostate cancer specimens, and differences in chromatin accessibility were studied. We found a clear enrichment of DNA binding motifs of multiple transcription factors that were previously described to be involved in prostate cancer, including the androgen receptor (AR). This motivated us to study the chromatin binding features of AR in prostate tumors, in search of different patterns of DNA binding between tumors that were sensitive and resistant to therapeutics. AR binding sites that were differentially enriched between sensitive and resistant tumors were coupled to the most proximal genes, after which expression of these genes was tested for correlations with outcome in multiple prostate cancer datasets. This resulted in a refined and optimized 9-gene set that could stratify patients with prostate cancer on outcome.

#### Impact

Yearly, 250,000 patients with prostate cancer die from the consequences of the disease. An early identification of patients who are at high risk for relapse would enable a closer monitoring of these patients. Our results have led to the discovery of such a biomarker, which is independent of known prognostic features, and can identify in synergy with these clinical parameters a subpopulation of patients with prostate cancer who are at high risk for disease relapse after prostatectomy. With this study, the concept of a "genomics pipeline" of biomarker discovery is proposed that may prove useful in other fields of oncology as well.

## References

- Amling CL, Blute ML, Bergstralh EJ, Seay TM, Slezak J, Zincke H (2000) Long-term hazard of progression after radical prostatectomy for clinically localized prostate cancer: continued risk of biochemical failure after 5 years. *J Urol* 164: 101–105
- Antonarakis ES, Lu C, Wang H, Lubner B, Nakazawa M, Roeser JC, Chen Y, Mohammad TA, Chen Y, Fedor HL *et al* (2014) AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* 371: 1028–1038
- Asangani IA, Dommetti VL, Wang X, Malik R, Cieslik M, Yang R, Escara-Wilke J, Wilder-Romans K, Dhanireddy S, Engelke C *et al* (2014) Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* 510: 278–282
- Bohm M, Locke WJ, Sutherland RL, Kench JG, Henshall SM (2009) A role for GATA-2 in transition to an aggressive phenotype in prostate cancer through modulation of key androgen-regulated genes. *Oncogene* 28: 3847–3856
- Bolla M, de Reijke TM, Van Tienhoven G, Van den Bergh AC, Oddens J, Poortmans PM, Gez E, Kil P, Akdas A, Soete G *et al* (2009) Duration of androgen suppression in the treatment of prostate cancer. *N Engl J Med* 360: 2516–2527

- Bolton EC, So AY, Chaivorapol C, Haqq CM, Li H, Yamamoto KR (2007) Cell- and gene-specific regulation of primary target genes by the androgen receptor. *Genes Dev* 21: 2005–2017
- Boormans JL, Korsten H, Ziel-van der Made AJ, van Leenders GJ, de Vos CV, Jenster G, Trapman J (2013) Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer. *Int J Cancer* 133: 335–345
- Briganti A, Joniau S, Gontero P, Abdollah F, Passoni NM, Tombal B, Marchioro G, Kneitz B, Walz J, Frohneberg D et al (2012) Identifying the best candidate for radical prostatectomy among patients with high-risk prostate cancer. *Eur Urol* 61: 584–592
- Brinkmann AO, Blok LJ, de Ruiter PE, Doesburg P, Steketee K, Berrevoets CA, Trapman J (1999) Mechanisms of androgen receptor activation and function. *J Steroid Biochem Mol Biol* 69: 307–313
- Chen Y, Chi P, Rockowitz S, Iaquina PJ, Shamu T, Shukla S, Gao D, Sirota I, Carver BS, Wongvipat J et al (2013) ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* 19: 1023–1029
- Chen Z, Lan X, Thomas-Ahner JM, Wu D, Liu X, Ye Z, Wang L, Sunkel B, Grenade C, Chen J et al (2015) Agonist and antagonist switch DNA motifs recognized by human androgen receptor in prostate cancer. *EMBO J* 34: 502–516
- Clegg NJ, Wongvipat J, Joseph JD, Tran C, Ouk S, Dilhas A, Chen Y, Grillot K, Bischoff ED, Cai L et al (2012) ARN-509: a novel antiandrogen for prostate cancer treatment. *Cancer Res* 72: 1494–1503
- Culig Z, Hoffmann J, Erdel M, Eder IE, Hobisch A, Hittmair A, Bartsch G, Utermann G, Schneider MR, Parczyk K et al (1999) Switch from antagonist to agonist of the androgen receptor bicalutamide is associated with prostate tumour progression in a new model system. *Br J Cancer* 81: 242–251
- Cuperlovic-Culf M, Belacel N, Davey M, Ouellette RJ (2010) Multi-gene biomarker panel for reference free prostate cancer diagnosis: determination and independent validation. *Biomarkers* 15: 693–706
- D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ et al (1998) Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 280: 969–974
- D'Antonio JM, Ma C, Monzon FA, Pflug BR (2008) Longitudinal analysis of androgen deprivation of prostate cancer cells identifies pathways to androgen independence. *Prostate* 68: 698–714
- Dorff TB, Flaig TW, Tangen CM, Hussain MH, Swanson GP, Wood DP Jr, Sakr WA, Dawson NA, Haas NB, Crawford ED et al (2011) Adjuvant androgen deprivation for high-risk prostate cancer after radical prostatectomy: SWOG S9921 study. *J Clin Oncol* 29: 2040–2045
- El Gammal AT, Bruchmann M, Zustin J, Isbarn H, Hellwinkel OJ, Kollermann J, Sauter G, Simon R, Wilczak W, Schwarz J et al (2010) Chromosome 8p deletions and 8q gains are associated with tumor progression and poor prognosis in prostate cancer. *Clin Cancer Res* 16: 56–64
- Ernst J, Nau GJ, Bar-Joseph Z (2005) Clustering short time series gene expression data. *Bioinformatics* 21(Suppl 1): i159–i168
- Ernst J, Bar-Joseph Z (2006) STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7: 191
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33: 1–22
- Gasi Tandefelt D, Boormans J, Hermans K, Trapman J (2014) ETS fusion genes in prostate cancer. *Endocr Relat Cancer* 21: R143–R152
- Giresi PG, Lieb JD (2009) Isolation of active regulatory elements from eukaryotic chromatin using Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE). *Methods* 48: 233–239
- Griffon A, Barbier Q, Dalino J, van Helden J, Spicuglia S, Ballester B (2015) Integrative analysis of public CHIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res* 43: e27
- Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, Mason M, Matveev V, Wiegel T, Zattoni F et al (2014a) EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent-update 2013. *Eur Urol* 65: 124–137
- Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, Mason M, Matveev V, Wiegel T, Zattoni F et al (2014b) EAU guidelines on prostate cancer. Part II: treatment of advanced, relapsing, and castration-resistant prostate cancer. *Eur Urol* 65: 467–479
- Irshad S, Bansal M, Castillo-Martin M, Zheng T, Aytas A, Wenske S, Le Magnen C, Guarnieri P, Sumazin P, Benson MC et al (2013) A molecular signature predictive of indolent prostate cancer. *Sci Transl Med* 5: 202ra122
- Itkonen H, Mills IG (2012) Chromatin binding by the androgen receptor in prostate cancer. *Mol Cell Endocrinol* 360: 44–51
- Jansen MP, Knijnenburg T, Reijm EA, Simon I, Kerkhoven R, Droog M, Velds A, van Laere S, Dirix L, Alexi X et al (2013) Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res* 73: 6632–6641
- Jeong TO, Oh KJ, Xuan Nguyen NT, Kim YR, Kim MS, Lee SD, Ryu SB, Jung C (2012) Evaluation of HOXB13 as a molecular marker of recurrent prostate cancer. *Mol Med Rep* 5: 901–904
- Jia L, Berman BP, Jariwala U, Yan X, Cogan JP, Walters A, Chen T, Buchanan G, Frenkel B, Coetzee GA (2008) Genomic androgen receptor-occupied regions with different functions, defined by histone acetylation, coregulators and transcriptional capacity. *PLoS ONE* 3: e3645
- Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, Prabhakar S (2013) Uniform, optimal signal processing of mapped deep-sequencing data. *Nat Biotechnol* 31: 615–622
- Lalonde E, Ishkanian AS, Sykes J, Fraser M, Ross-Adams H, Erho N, Dunning MJ, Halim S, Lamb AD, Moon NC et al (2014) Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet Oncol* 15: 1521–1532
- Lamb AD, Massie CE, Neal DE (2014) The transcriptional programme of the androgen receptor (AR) in prostate cancer. *BJU Int* 113: 358–366
- Lonergan PE, Tindall DJ (2011) Androgen receptor signaling in prostate cancer development and progression. *J Carcinog* 10: 20
- Lu S, Jenster G, Epner DE (2000) Androgen induction of cyclin-dependent kinase inhibitor p21 gene: role of androgen receptor and transcription factor Sp1 complex. *Mol Endocrinol* 14: 753–760
- Massie CE, Lynch A, Ramos-Montoya A, Boren J, Stark R, Fazli L, Warren A, Scott H, Madhu B, Sharma N et al (2011) The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *EMBO J* 30: 2719–2733
- Miocinovic R, Berglund RK, Stephenson AJ, Jones JS, Fergany A, Kaouk J, Klein EA (2011) Avoiding androgen deprivation therapy in men with high-risk prostate cancer: the role of radical prostatectomy as initial treatment. *Urology* 77: 946–950
- Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, Serandour AA, Birrell SN, Bruna A, Saadi A et al (2015) Progesterone receptor modulates ERalpha action in breast cancer. *Nature* 523: 313–317
- Nhili R, Peixoto P, Depauw S, Flajollet S, Dezitter X, Munde MM, Ismail MA, Kumar A, Farahat AA, Stephens CE et al (2013) Targeting the DNA-binding activity of the human ERG transcription factor using new heterocyclic dithiophene diamidines. *Nucleic Acids Res* 41: 125–138

- Norris JD, Chang CY, Wittmann BM, Kunder RS, Cui H, Fan D, Joseph JD, McDonnell DP (2009) The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol Cell* 36: 405–416
- Pilepich MV, Winter K, John MJ, Mesic JB, Sause W, Rubin P, Lawton C, Machtay M, Grignon D (2001) Phase III radiation therapy oncology group (RTOG) trial 86-10 of androgen deprivation adjuvant to definitive radiotherapy in locally advanced carcinoma of the prostate. *Int J Radiat Oncol Biol Phys* 50: 1243–1252
- Potter GA, Barrie SE, Jarman M, Rowlands MG (1995) Novel steroidal inhibitors of human cytochrome P45017 alpha (17 alpha-hydroxylase-C17,20-lyase): potential agents for the treatment of prostatic cancer. *J Med Chem* 38: 2463–2471
- Rahim S, Minas T, Hong SH, Justvig S, Celik H, Kont YS, Han J, Kallarakal AT, Kong Y, Rudek MA et al (2014) A small molecule inhibitor of ETV1, YK-4-279, prevents prostate cancer growth and metastasis in a mouse xenograft model. *PLoS ONE* 9: e114260
- Ramos-Montoya A, Lamb AD, Russell R, Carroll T, Jurmeister S, Galeano-Dalmau N, Massie CE, Boren J, Bon H, Theodorou V et al (2014) HES6 drives a critical AR transcriptional programme to induce castration-resistant prostate cancer through activation of an E2F1-mediated cell cycle network. *EMBO Mol Med* 6: 651–661
- Roach M III (2014) Current trends for the use of androgen deprivation therapy in conjunction with radiotherapy for patients with unfavorable intermediate-risk, high-risk, localized, and locally advanced prostate cancer. *Cancer* 120: 1620–1629
- Robinson JL, Macarthur S, Ross-Innes CS, Tilley WD, Neal DE, Mills IG, Carroll JS (2011) Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. *EMBO J* 30: 3019–3027
- Robinson JL, Carroll JS (2012) FoxA1 is a key mediator of hormonal response in breast and prostate cancer. *Front Endocrinol* 3: 68
- Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR et al (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481: 389–393
- Safe S, Imanirad P, Sreevalsan S, Nair V, Jutooru I (2014) Transcription factor Sp1, also known as specificity protein 1 as a therapeutic target. *Expert Opin Ther Targets* 18: 759–769
- Sahu B, Laakso M, Ovaska K, Mirtti T, Lundin J, Rannikko A, Sankila A, Turunen JP, Lundin M, Konsti J et al (2011) Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J* 30: 3962–3976
- Scher HI, Sawyers CL (2005) Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J Clin Oncol* 23: 8253–8261
- Schubert M, Joniau S, Contero P, Kneitz S, Scholz CJ, Kneitz B, Briganti A, Karnes RJ, Tombal B, Walz J et al (2012) The role of adjuvant hormonal treatment after surgery for localized high-risk prostate cancer: results of a matched multiinstitutional analysis. *Adv Urol* 2012: 612707
- Shang Y, Myers M, Brown M (2002) Formation of the androgen receptor transcription complex. *Mol Cell* 9: 601–610
- Sharma NL, Massie CE, Ramos-Montoya A, Zecchini V, Scott HE, Lamb AD, MacArthur S, Stark R, Warren AY, Mills IG et al (2013) The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer Cell* 23: 35–47
- Shemshedini L, Knauthe R, Sassone-Corsi P, Pornon A, Gronemeyer H (1991) Cell-specific inhibitory and stimulatory effects of Fos and Jun on transcription activation by nuclear receptors. *EMBO J* 10: 3839–3849
- Shen P, Sun J, Xu G, Zhang L, Yang Z, Xia S, Wang Y, Liu Y, Shi G (2014) KLF9, a transcription factor induced in flutamide-caused cell apoptosis, inhibits AKT activation and suppresses tumor growth of prostate cancer cells. *Prostate* 74: 946–958
- Siddiqui SA, Boorjian SA, Blute ML, Rangel LJ, Bergstralh EJ, Karnes RJ, Frank I (2011) Impact of adjuvant androgen deprivation therapy after radical prostatectomy on the survival of patients with pathological T3b prostate cancer. *BJU Int* 107: 383–388
- Takahashi S, Suzuki S, Inaguma S, Ikeda Y, Cho YM, Nishiyama N, Fujita T, Inoue T, Hioki T, Sugimura Y et al (2002) Down-regulation of human X-box binding protein 1 (hXBP-1) expression correlates with tumor progression in human prostate cancers. *Prostate* 50: 154–161
- Taslim C, Chen Z, Huang K, Huang TH, Wang Q, Lin S (2012) Integrated analysis identifies a class of androgen-responsive genes regulated by short combinatorial long-range mechanism facilitated by CTCF. *Nucleic Acids Res* 40: 4754–4764
- Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B et al (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18: 11–22
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global cancer statistics, 2012. *CA Cancer J Clin* 65: 87–108
- Tran C, Ouk S, Clegg NJ, Chen Y, Watson PA, Arora V, Wongvipat J, Smith-Jones PM, Yoo D, Kwon A et al (2009) Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 324: 787–790
- Tsurumaki Sato Y, Fukuhara H, Suzuki M, Fujimura T, Nakagawa T, Nishimatsu H, Kume H, Morikawa T, Fukayama M, Homma Y (2014) Long-term results of radical prostatectomy with immediate adjuvant androgen deprivation therapy for pT3N0 prostate cancer. *BMC Urol* 14: 13
- Valencia LB, Sweeney CJ, Pomerantz MM (2015) Sequencing current therapies in the treatment of metastatic prostate cancer. *Cancer Treat Rev* 41: 332–340
- Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, Chen Z, Beroukheim R, Wang H, Lupien M et al (2009) Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* 138: 245–256
- Xu Y, Chen SY, Ross KN, Balk SP (2006) Androgens induce prostate cancer cell proliferation through mammalian target of rapamycin activation and post-transcriptional increases in cyclin D proteins. *Cancer Res* 66: 7783–7792
- Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y et al (2010) An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 17: 443–454
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137
- Zincke H, Lau W, Bergstralh E, Blute ML (2001) Role of early adjuvant hormonal therapy after radical prostatectomy for prostate cancer. *J Urol* 166: 2208–2215
- Zwart W, Koornstra R, Wesseling J, Rutgers E, Linn S, Carroll JS (2013) A carrier-assisted ChIP-seq method for estrogen receptor-chromatin interactions from breast cancer core needle biopsy samples. *BMC Genom* 14: 232



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.