



Clinical

# Distinct transcriptional repertoire of the androgen receptor in ETS fusion-negative prostate cancer

Anders E. Berglund<sup>1</sup> · Robert J. Rounbehler<sup>2,3</sup> · Travis Gerke<sup>4</sup> · Shivanshu Awasthi<sup>4</sup> · Chia-Ho Cheng<sup>1</sup> · Mandeep Takhar<sup>5</sup> · Elai Davicioni<sup>5</sup> · Mohammed Alshalalfa<sup>5</sup> · Nicholas Erho<sup>5</sup> · Eric A. Klein<sup>6</sup> · Stephen J. Freedland<sup>7</sup> · Ashley E. Ross<sup>8</sup> · Edward M. Schaeffer<sup>9</sup> · Bruce J. Trock<sup>10</sup> · Robert B. Den<sup>11</sup> · John L. Cleveland<sup>2</sup> · Jong Y. Park<sup>4</sup> · Jasreman Dhillon<sup>12</sup> · Kosj Yamoah<sup>4,13</sup>

Received: 10 July 2018 / Revised: 27 August 2018 / Accepted: 8 September 2018  
© The Author(s) 2018. This article is published with open access

## Abstract

**Background** Prostate cancer (PCa) tumors harboring translocations of *ETS* family genes with the androgen responsive *TMPRSS2* gene (ETS+ tumors) provide a robust biomarker for detecting PCa in approximately 70% of patients. ETS+ PCa express high levels of the androgen receptor (AR), yet PCa tumors lacking *ETS* fusions (ETS−) also express AR and demonstrate androgen-regulated growth. In this study, we evaluate the differences in the AR-regulated transcriptomes between ETS+ and ETS− PCa tumors.

**Methods** 10,608 patient tumors from three independent PCa datasets classified as ETS+ (samples overexpressing ERG or other ETS family members) or ETS− (all other PCa) were analyzed for differential gene expression using false-discovery-rate adjusted methods and gene-set enrichment analysis (GSEA).

**Results** Based on the expression of AR-dependent genes and an unsupervised Principal Component Analysis (PCA) model, AR-regulated gene expression alone was able to separate PCa samples into groups based on ETS status in all PCa databases. ETS status distinguished several differentially expressed genes in both TCGA (6.9%) and GRID (6.6%) databases, with 413 genes overlapping in both databases. Importantly, GSEA showed enrichment of distinct androgen-responsive genes in both ETS− and ETS+ tumors, and AR ChIP-seq data identified 131 direct AR-target genes that are regulated in an ETS-specific fashion. Notably, dysregulation of ETS-dependent AR-target genes within the metabolic and non-canonical WNT pathways was associated with clinical outcomes.

**Conclusions** ETS status influences the transcriptional repertoire of the AR, and ETS− PCa tumors appear to rely on distinctly different AR-dependent transcriptional programs to drive and sustain tumorigenesis.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41391-018-0103-4>) contains supplementary material, which is available to authorized users.

✉ Kosj Yamoah  
Kosj.Yamoah@moffitt.org

<sup>1</sup> Department of Biostatistics & Bioinformatics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA

<sup>2</sup> Department of Tumor Biology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA

<sup>3</sup> Department of Oncological Sciences, University of South Florida, Tampa, FL, USA

<sup>4</sup> Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA

<sup>5</sup> GenomeDx Biosciences Inc, Vancouver, BC, Canada

<sup>6</sup> Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA

<sup>7</sup> Department of Surgery, Division of Urology, Center for Integrated Research on Cancer and Lifestyle, Samuel Oschin Comprehensive Cancer Center, Cedars-Sinai Medical Center, Los Angeles, CA, USA

<sup>8</sup> Texas Urology Specialists, Dallas, TX, USA

<sup>9</sup> Department of Urology, Northwestern University, Chicago, IL, USA

<sup>10</sup> Department of Urology, Johns Hopkins, Baltimore, MD, USA

<sup>11</sup> Department of Radiation Oncology, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA, USA

<sup>12</sup> Department of Pathology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA

<sup>13</sup> Department of Radiation Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA

## Introduction

The androgen receptor (AR), a nuclear hormone transcription factor, directs the transcription of several genes implicated in the development and progression of prostate cancer (PCa). Given the critical roles AR plays in PCa progression, advanced PCa is often treated with an androgen deprivation therapy (ADT)-based regimen, which depletes the natural ligands of AR, testosterone and dihydrotestosterone. However, several acquired gain-of-function mutations and modifications of the *AR* gene can occur during ADT and these are major mechanisms that drive androgen-independent PCa progression [1]. Accordingly, development and progression of PCa is often studied in the context of AR signaling, yet less is known regarding AR-regulated molecular pathways in PCa progression.

Genomic aberrations in members of the E26 transformation-specific (ETS) family of oncogenic transcription factors (*ERG*, *ETV1*, *ETV4*, *ETV5*, and *FLI1*) are early carcinogenic events found in up to 70% of PCa [2]. *ETS* gene translocations are AR-regulated events, and most frequently this results in overexpression of ETS family genes by placing them under the control of the androgen-responsive promoter *TMPRSS2* [3–5]. Furthermore, ectopic activation of ETS transcription factors directs expression of several biomarkers, including AR-regulated genes [6–10]. Notably, up to 90% of aggressive early-onset PCa tumors appear to be ERG+, and ETS+ tumors have elevated AR expression and somatic alterations that are androgen-driven, including the expression of *TMPRSS2-ETS* fusions [8]. Finally, molecular and biological studies of ETS+ tumors have revealed that sites of DNA damage are often located near AR binding sites [11], that ERG and AR have overlapping transcription targets, and that ERG inhibits AR-directed differentiation of prostate epithelial cells [12].

Despite the fact AR is also expressed in PCa tumors lacking ETS fusions (ETS–) [13], the role of AR activity in ETS– tumors is poorly understood, and there are also very little data regarding the landscape of somatic alteration of ETS– tumors. Here we report striking differences in the AR transcriptional programs of ETS+ versus ETS– PCa that establish marked differences in the biology of these tumors.

## Results

### AR-regulated genes differentiate ETS– and ETS+ PCa tumors

ETS status (ETS+ or ETS–) was determined for all PCa samples within the TCGA, GRID, and GRID-prospective datasets (Table 1). Clinical, demographic, and pathological characteristics of patients within the TCGA and GRID

**Table 1** ETS status of PCa tumors in TCGA and GRID datasets

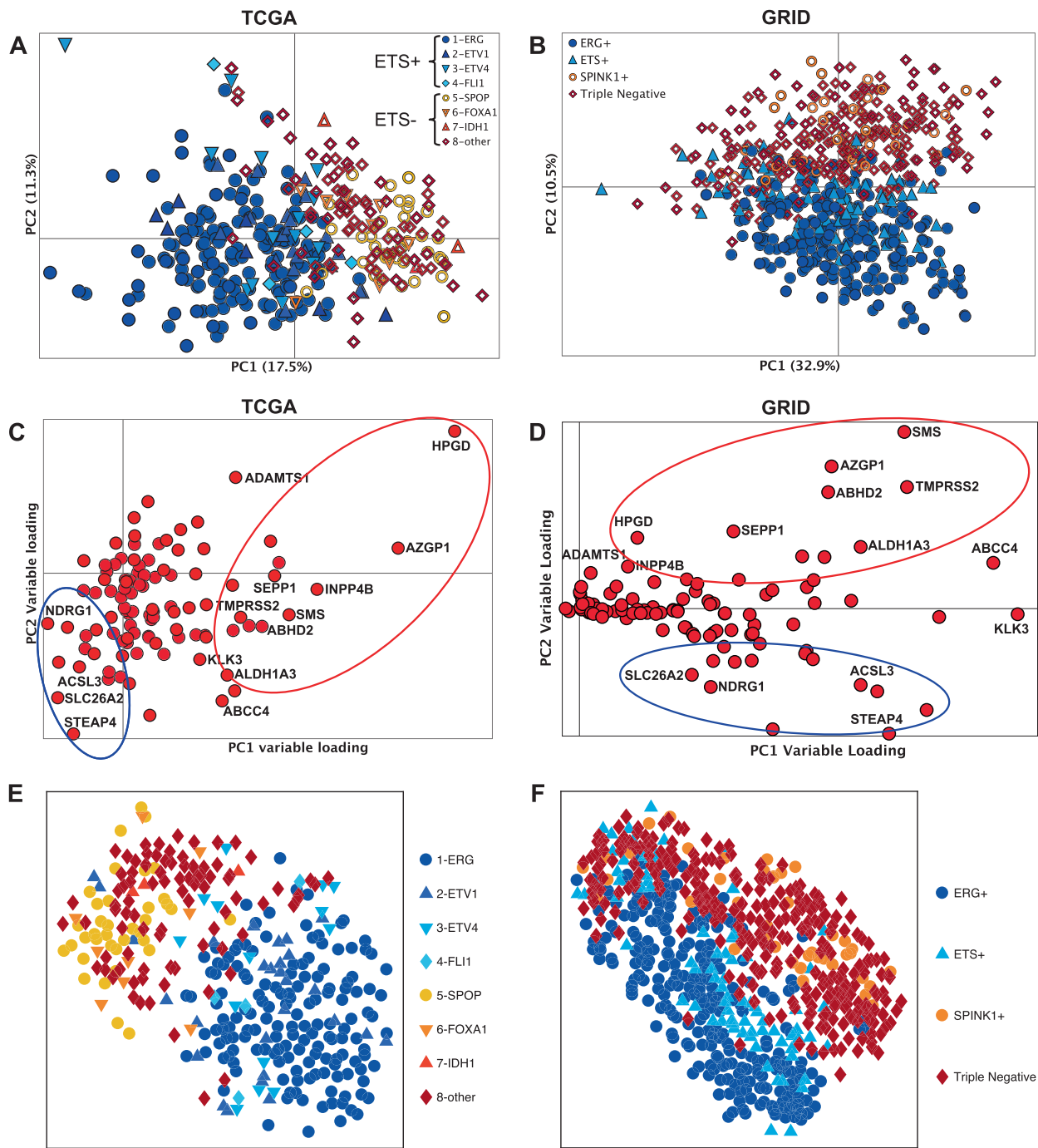
ETS subtypes	TCGA ( <i>N</i> = 333)	GRID ( <i>N</i> = 635)	GRID-prospective ( <i>N</i> = 9640)
ETS–	<i>N</i> = 135	<i>N</i> = 309	<i>N</i> = 4757
Variants	SPOP FOXA1 IDH1 Others	SPINK1+ Triple negative	SPINK1+ Triple negative Other
ETS+	<i>N</i> = 198	<i>N</i> = 326	<i>N</i> = 4883
Variants	ERG ETV1 ETV4 FLI1	ERG+ ETS+	ERG+ ETV1 ETV4 FLI1

TCGA cancer genome atlas research network; GRID genomic resource information database; *Triple negative* ERG–, ETS– and SPINK1–

cohorts stratified by ETS status is shown in Supplementary Table S1. An unsupervised principal component analyses (PCA) was performed in the TCGA (Fig. 1a) and GRID (Fig. 1b) dataset. PCA showed a robust separation of ETS+ and ETS– tumors based on their expression profile of genes present in the HALLMARK\_ANDROGEN\_RESPONSE gene list, which are well-validated AR regulated genes (Fig. 1a, b). The loading plots for the PCA models indicate the individual AR-dependent genes that are differentially expressed in both the TCGA and GRID datasets (Fig. 1c, d, respectively). Finally, a similar clustering of ETS+ versus ETS– samples based on AR-dependent genes was manifest using a second unsupervised method for sample clustering, *t*-SNE (*t*-distributed stochastic neighbor embedding) (Fig. 1e, f). Thus, known AR-regulated genes are differentially expressed in ETS– prostate tumors compared to ETS+ tumors, suggesting AR activates distinct biological pathways in ETS– versus ETS+ PCa.

### Unique androgen response genes are differentially upregulated in ETS– and ETS+ prostate tumors

To investigate if discrete biological pathways are upregulated in ETS– prostate tumors compared to ETS+ prostate tumors, we identified the genes that are differentially expressed between the ETS+ and ETS– cohorts in the TCGA and GRID dataset. This analysis found 1423 (Supplementary Table S2) genes in the TCGA dataset (6.9% of genes analyzed) and 3047 (Supplementary Table S3) genes in the GRID dataset (6.6% of genes analyzed) to be differentially expressed between ETS– and ETS+ tumors. Further, 413 genes were identified as differentially expressed in the two ETS groups in both the TCGA and GRID datasets (Fig. 2a, Supplementary Table S4), where 220 genes were overexpressed in ETS+ cases and 193 genes

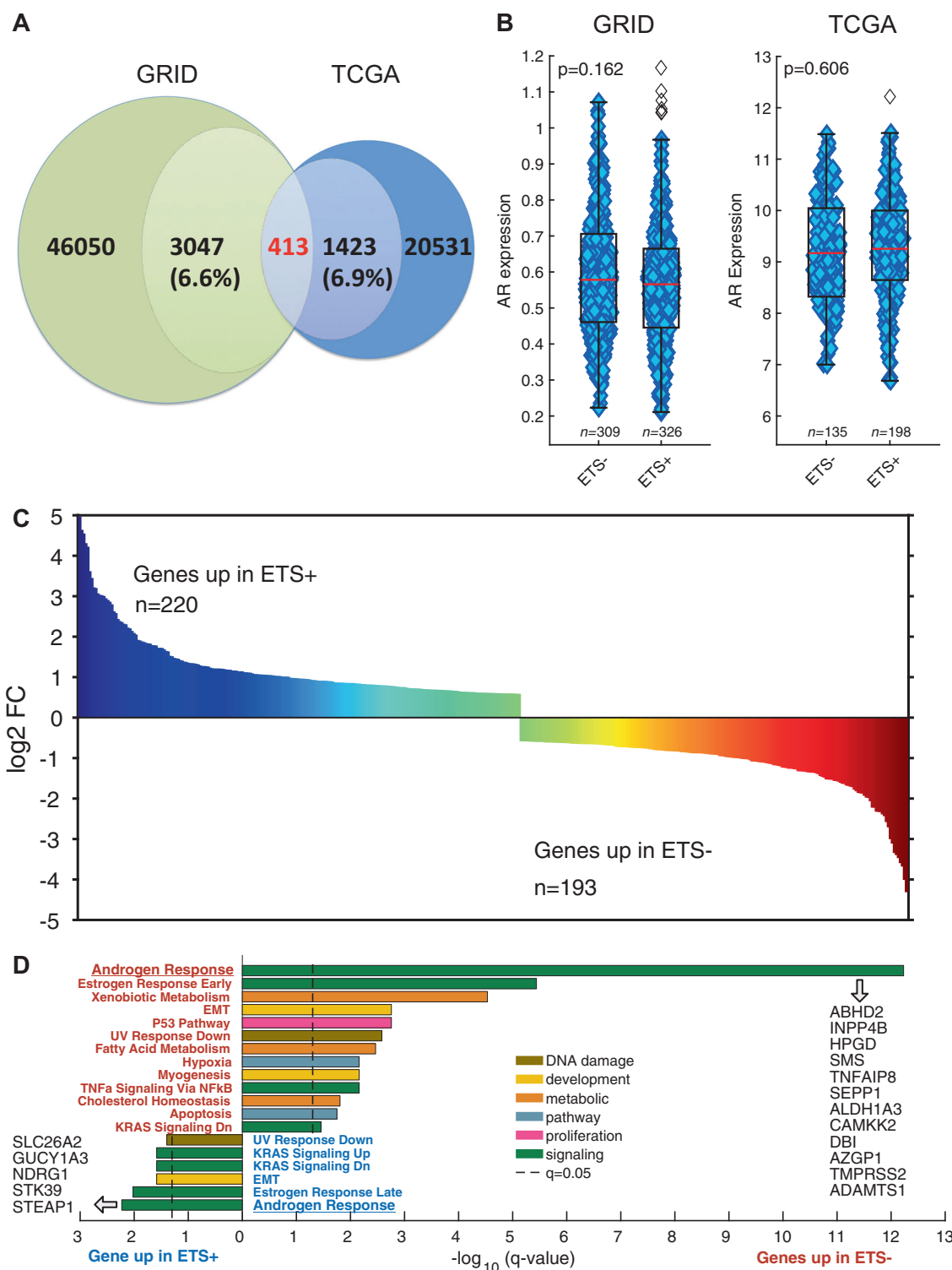


**Fig. 1** AR-regulated genes discriminate tumors based on ETS status. An unsupervised PCA model of 101 AR-regulated genes affirms the distinct AR signatures of ETS+ and the ETS- samples in the first principal component for the TCGA RNA-seq (a) and in the second principal component for the GRID microarray profiling (b) PCA datasets. c and d show the relative contribution of the individual AR-regulated genes to the PCA models in a and b, respectively. The t-SNE

model shows similar results to the PCA model for the TCGA (e) and the GRID (f) dataset, where there is a clear separation between the ETS- and ETS+ samples. The results also indicate that the molecular subtypes used for ETS+ (1-ERG, 2-ETV1, 3-ETV4, 4-FLI1) and ETS- (5-SPOP, 6-FOXA1, 7-IDH1, 8-other) fall into the correct ETS category

were overexpressed in ETS- cases. These 413 genes were validated as being differentially expressed based on ETS status in a third independent, non-overlapping cohort, specifically the prospective GRID cohort (GRID-prospective)

that is comprised of 9640 PCa patient samples. This analysis established a strong correlation in the expression of these genes between the TCGA dataset, where expression was measured by RNA sequencing, and the



**Fig. 2** Genes differentially expressed in ETS+ versus ETS– PCa tumors. **a** The GRID dataset has 46050 genes with 3047 (6.6%) being differentially expressed when comparing ETS+ vs. ETS– PCa. The TCGA dataset has 20531 genes with 1423 (6.9%) being differentially expressed based on ETS status. Using a false-discovery-rate adjusted ( $q < 0.05$ ) Mann–Whitney U test and a fold-change cut-off of 0.585 (TCGA) and 0.05 (GRID), 413 differentially expressed genes based on ETS status were defined in both PCa databases. **b** There is no

significant difference in AR expression between ETS– and ETS+ tumors in either TCGA or GRID. The 413 significant differentially expressed genes (**c**) in the TCGA and GRID PCa based on ETS status were analyzed by GSEA and the HALLMARK gene sets [43]. Significant gene sets (**d**) overexpressed in ETS+ are shown at the top and those gene sets for ETS– shown below. Androgen response was the top-ranked gene set for both ETS+ and ETS– significantly over-expressed genes



GRID-prospective dataset, where expression was measured using a genome-wide microarray platform ( $r > 0.8$ ; Supplementary Fig. S1A). This correlation was even more robust when comparing the GRID and GRID-prospective datasets ( $r \geq 0.9$ ; Supplementary Fig. S1B), which both used the same microarray platform for gene expression analysis. Percentage tumor purity was also similar between ETS- and ETS+ tumors in both TCGA and GRID databases (Supplementary Fig. S1C and D).

To determine which biological pathways, based on the Hallmark Gene Sets, are over-represented in the 413 genes that are up-regulated in either of the two ETS groups (Fig. 2c), gene-set enrichment analysis (GSEA) was performed. Intriguingly, GSEA revealed that independent, non-overlapping sets of androgen response genes are overexpressed in both ETS- and ETS+ tumors (Fig. 2d), suggesting ETS status alters the repertoire of androgen-regulated genes in PCa. Furthermore, many of the genes upregulated in ETS- tumors are involved in metabolic, p53 or hypoxia pathways, whereas this was not the case in ETS+ tumors that showed more prominent association with the epithelial mesenchymal transition (EMT) and KRAS signaling pathways (Fig. 2). Importantly, AR expression by itself is not dependent on ETS status as shown in Fig. 2b. Individual genes enriched in corresponding pathways are provided in Supplementary Fig. S2 A-B. Thus, ETS- and ETS+ prostate tumors both show comparable expression of AR, yet demonstrate unique AR-dependent alterations in biological pathways that drive tumorigenesis.

### AR directs distinct transcriptional programs in PCa based on ETS status

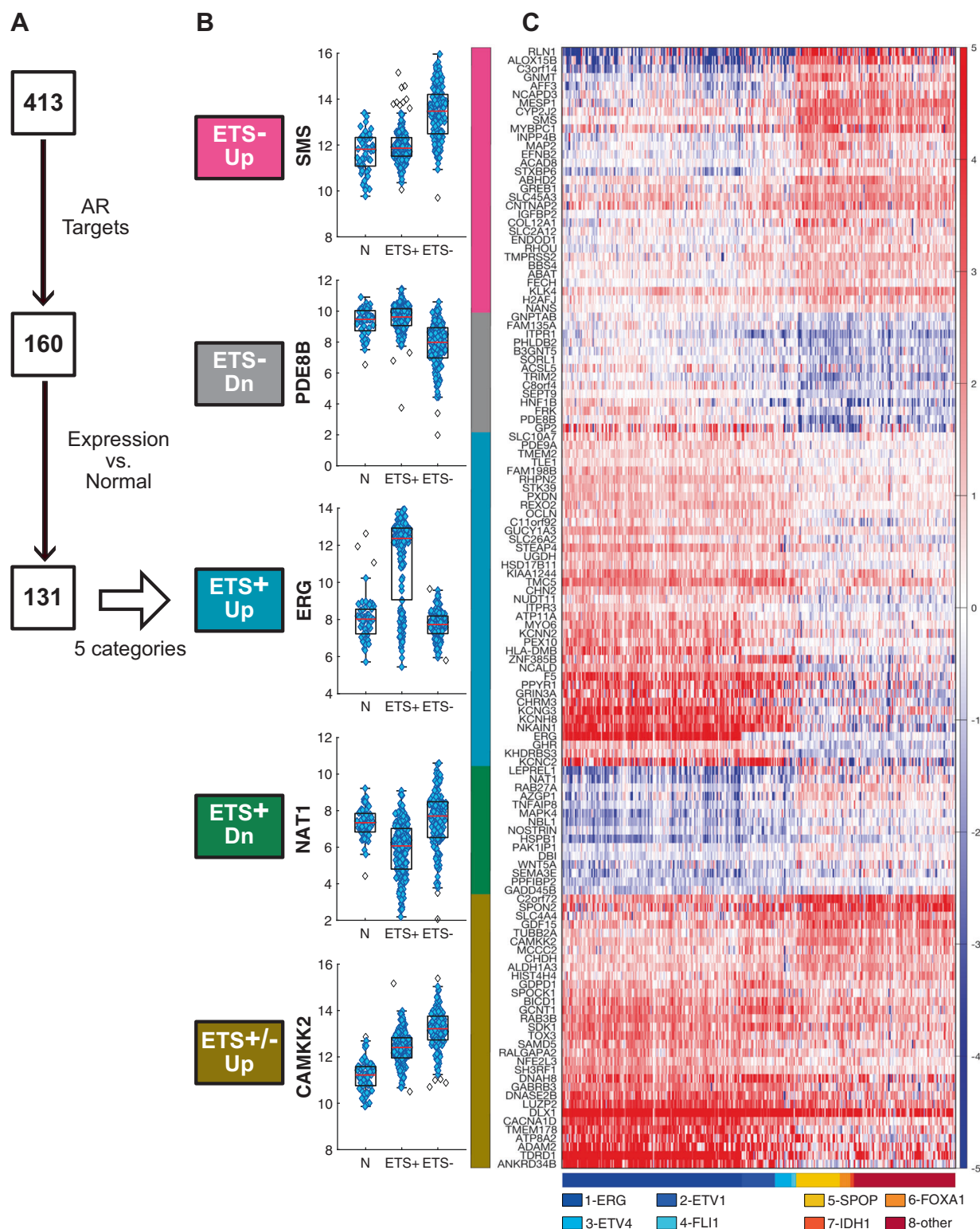
Since ETS- and ETS+ prostate tumors differentially expressed distinct sets of androgen response pathway genes, we assessed if the AR directs distinct transcriptional programs in PCa based on ETS status. To do so, we queried the 220 genes overexpressed in ETS+ cases and 193 genes overexpressed in ETS- cases in three publicly available AR ChIP-seq datasets [14], which defined direct AR targets as those genes that display AR binding to an androgen response element within 25 kb of the target gene's promoter, in several PCa cell lines [12, 14, 15], primary prostate tumor tissues [15], and metastatic PCa [16]. These analyses revealed that a total of 160 of these genes were not only direct AR target genes but were also regulated in an ETS-specific manner (Fig. 3a). To further analyze these 160 AR targets, their relative levels of expression in adjacent non-tumor prostate sample was compared to their expression in ETS- tumor samples and ETS+ tumor samples using the TCGA dataset. Using the subtractive genomics method, we removed 29 AR target genes that were not significantly different between the PCa tumors and the

adjacent normal prostate tissues. The remaining 131 AR target genes belonged to five distinct categories based on ETS status: (1) ETS- Up (genes upregulated only in ETS- tumors), (2) ETS- Dn (genes downregulated only in ETS- tumors) (3) ETS+ Up (genes upregulated only in ETS+ tumors), (4) ETS+ Dn (genes downregulated only in ETS+ tumors), and (5) ETS-/ETS+ Up (genes upregulated in both ETS- and ETS+ tumors). As shown in Fig. 3b, *SMS*, *PDE8B*, *ERG*, *NAT1* and *CAMKK2* are representative genes identified for each of these five categories, which were clearly evident in a clustering analyses of the 131 direct AR target genes that are differentially expressed based on ETS status in PCa, as well as by the relevant subtypes that have been identified for ETS+ and ETS- prostate tumors (Fig. 3c). Strikingly, all 131 AR target genes were validated by an independent subtractive genomics and clustering analyses performed in 9640 samples from the Microarray-based GRID-prospective dataset (Table 1; Supplementary Fig. S3)

### AR transcriptional programming in prostate tumor is ETS-dependent

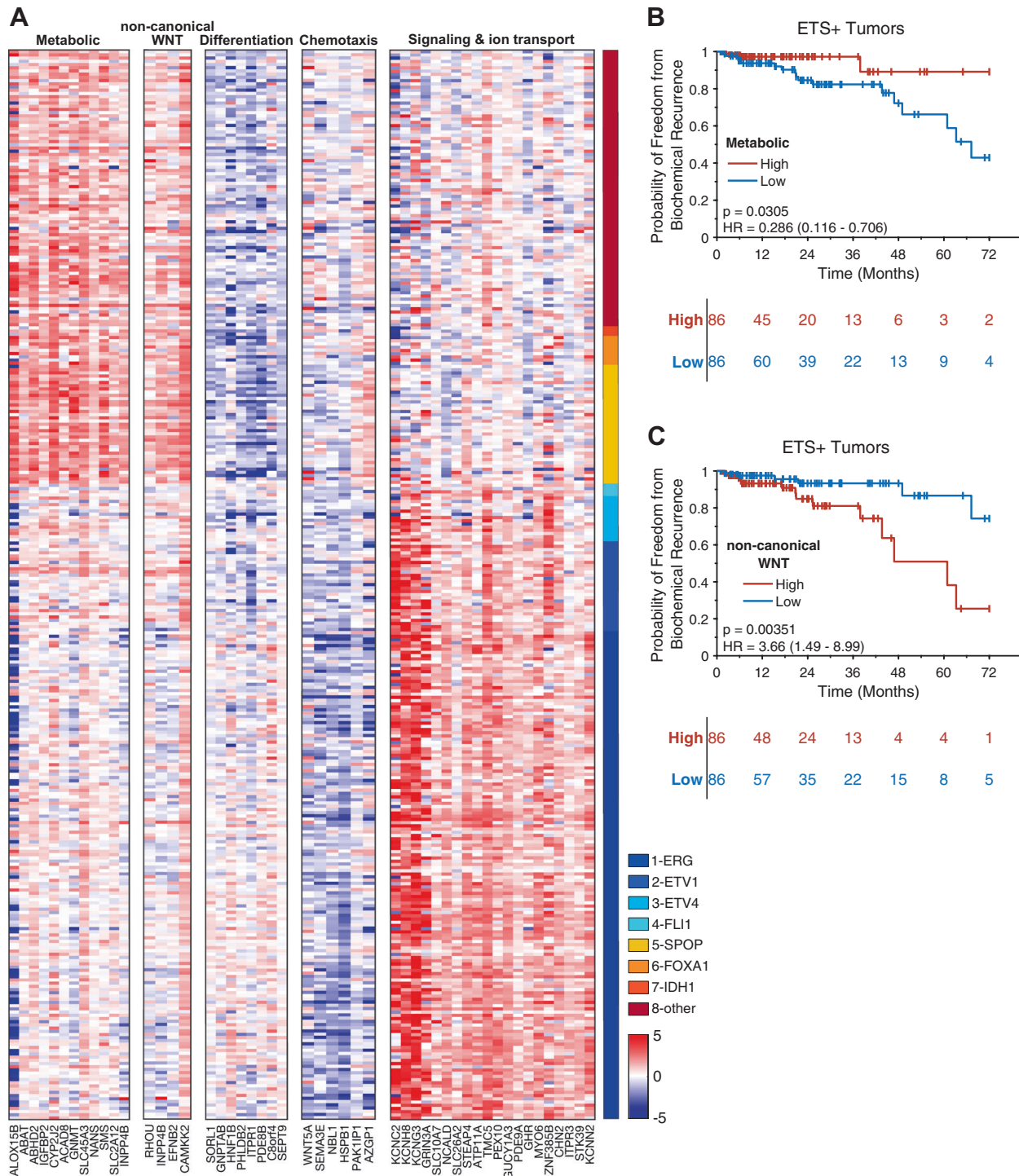
To identify biological pathways that may be involved in ETS-dependent PCa development, extensive literature searches coupled with gene ontology analysis was performed using GO biological process gene sets on the five distinct ETS-dependent AR target genes categories (Supplementary Table S5). Four of the five categories of ETS-dependent PCa AR target genes were enriched for specific biological pathways (Supplementary Table S5). These five pathways (metabolic, non-canonical WNT, differentiation, chemotaxis, and signaling & ion transport) with their corresponding genes are presented in Fig. 4a as a heatmap, where the expression values were normalized using median expression in normal adjacent tissue. This heatmap highlighted the up-regulation of metabolic pathway genes and the non-canonical WNT pathway, and downregulation of genes involved in cellular differentiation, in ETS- tumors. Conversely, ETS+ tumors have reduced expression of genes that suppress chemotaxis and cell motility and an upregulation of genes involved in signal transduction and ion transport (Fig. 4). Thus, there are profound differences in the AR repertoire of ETS+ and ETS- PCa that suggest marked differences in the biology of these tumors.

To test if these gene signature pathways were related to biochemical recurrence (BCR) a PCA model was applied to data from the TCGA cohort. The first principal component (PC1) of the non-canonical PCA model explains 55.9% of the variation, which is better than 99.6% of all random gene PCA models. The ratio of PC1/PC2 is 3.3 which is better than 99.99% of the random gene model. The corresponding results for the metabolic pathway is 51.5% explained the



**Fig. 3** Distinct direct AR target genes are regulated in ETS+ and ETS- PCa tumors. **a** Schematic of pipeline used to define 5 categories from the 131 direct AR transcription targets in ETS+ and ETS- PCa tumors. **b** Examples of genes whose expression is significantly different in adjacent normal tissue (N) in ETS+ PCa or ETS- PCa. **c** Heatmap of differentially expressed direct AR target genes in ETS-

Up pink, ETS- Down gray, ETS+ Up blue, ETS+ Down green, and that are upregulated in both ETS+ and ETS- PCa (brown). Each row/gene is normalized to median expression in adjacent normal tissue. The 8 subtypes of ETS+ and ETS- PCa, as defined by their expression of *ERG*, *ATV1*, *ETV4*, *FLI1*, *SPOP*, *FOXA1*, *IDH1* and “other” are shown beneath the heatmap



**Fig. 4** Specific AR target pathways show ETS status and BCR dependency. Gene ontology analysis and literature searches revealed that among ETS status-dependent AR target genes, metabolic pathway genes and non-canonical WNT pathway genes are up-regulated in

ETS- tumors, whereas signaling and ion transport pathway genes are up-regulated in ETS+ tumors (a). Kaplan-Meier curves for metabolic (b) and non-canonical WNT (c) genes show significance to BCR in ETS+ tumors

variance for PC1 (better than 99.99% of the random models), and with a PC1/PC2 ratio of 5.9 (better than 99.99% of the random models). These results indicate that the generated gene signatures are robust and better than random gene

signatures [17, 18]. Notably, two of these pathways, the non-canonical WNT pathway and the metabolic pathway genes, showed a significant association with BCR only in the ETS+ cohort. Specifically, low expression of metabolic



pathway genes and high expression of non-canonical WNT pathway genes was associated with worse biochemical recurrence (Fig. 4b, c); the corresponding results for the ETS<sup>−</sup> tumors are shown in supplementary figure S4.

## Discussion

The systematic analysis of the transcriptional landscape of PCa tumors based on ETS status reported herein reveals important insights into the biology of ETS<sup>+</sup> versus of ETS<sup>−</sup> tumors, and establishes that the repertoire of AR target genes is dramatically affected by ETS fusion oncoproteins. For example, GSEA analysis demonstrated that metabolic pathway genes, including genes involved in xenobiotic and fatty acid metabolism, were preferentially enriched in ETS<sup>−</sup> but not in ETS<sup>+</sup> PCa tumors. Further, although ETS<sup>+</sup> and ETS<sup>−</sup> PCa are both androgen dependent, different sets of androgen response genes are enriched based on ETS status, and analyses of previously published AR ChIP-seq datasets revealed that AR target genes are dysregulated in uniquely different ways in ETS<sup>−</sup> versus ETS<sup>+</sup> tumors. These findings explain the apparently conflicting reports regarding AR activity in ETS<sup>−</sup> PCa [8], where AR target genes that are dysregulated in ETS<sup>−</sup> tumors drive metabolism and suppress differentiation. Thus, ETS<sup>−</sup> tumors have comparable AR expression when compared to ETS<sup>+</sup> tumors (as shown in Fig. 2b), but rather utilize distinctly different AR-dependent transcriptional programs to sustain tumor growth and survival.

In an in-depth analysis of clinically relevant ETS-dependent, AR-regulated biological pathways we have identified 5 distinct pathways including the metabolic, non-canonical WNT, differentiation, chemotaxis, and signal transduction and ion transport. ETS<sup>+</sup> tumors showed reduced levels of AR target genes within the metabolic pathway that drive polyamine synthesis [19–21], fatty acid and arachidonic acid metabolism [22–24], and phosphoinositol metabolism [25, 26], non-canonical WNT, and chemotaxis pathways [10], and increased expression in differentiation and signal transduction and ion channel transport pathways. The reverse scenario was manifest in ETS<sup>−</sup> tumors.

Notably, AR regulates the expression of genes involved in polyamine biosynthesis such as spermine synthase (*SMS*) and Glycine-N-methyltransferase (*GNMT*), an enzyme involved in S-adenosyl-methionine (*SAM*) homeostasis. These findings are in accord with those showing that androgen tightly regulates polyamine biosynthesis via control of ornithine decarboxylase (*ODCI*) and SAM-decarboxylase (*AMDI*) [21, 27], and magnetic resonance spectroscopy analyses of clinical samples have suggested spermine as a biomarker for the malignant behavior of PCa [28].

The expression analyses reveal that AR coordinately induces targets involved in arachidonic acid metabolism in both ETS<sup>−</sup> and ETS<sup>+</sup> PCa tumors. First, it is interesting that glycerol and arachidonic acid produced via *ABDH2*, promote prostate tumor growth [23]. Indeed, immunohistochemical analysis of tumor specimens has shown a positive correlation of *ABDH2* levels with high Gleason score, pathological nodal stage, low cancer-specific survival rates, and a resistance to docetaxel-based chemotherapy [23]. Second, *ALOX15B* (arachidonate 15-lipoxygenase type II) known to metabolize arachidonic acid to fatty acid hydroperoxides, is selectively regulated in PCa in an ETS-dependent fashion [29]. Finally, the expression of *CYP2J2*, which promotes tumor cell growth by converting arachidonic acid to epoxyeicosatrienoic acids [24], is also up-regulated in ETS<sup>−</sup> tumors. Thus, given several other interesting metabolic targets are manifest in both ETS<sup>−</sup> and ETS<sup>+</sup> PCa tumors (Supplementary Fig. S2A-B; Table S5), targeting arachidonate metabolism may represent a particularly attractive therapeutic vulnerability for these tumors.

Of note, our analyses revealed that the WNT pathway is also up-regulated in ETS<sup>−</sup> PCa. Heretofore WNT/ $\beta$ -catenin signaling has been implicated in PCa tumor cell self-renewal, pathogenesis, and aggressive disease [30], and activated WNT/ $\beta$ -catenin signaling has been described to be among the most highly enriched pathways in ERG over-expressing PCa tumors [31]. However, our analyses also revealed ETS-dependent dysregulation of several AR targets within non-canonical WNT pathway genes, including *RHOA*, *INPP4B*, *EFNB2*, and *CAMKK2* (Fig. 4; Supplementary Table S5). These findings suggest distinct roles for WNT signaling in driving ETS<sup>+</sup> versus ETS<sup>−</sup> PCa, and further underscore the need for developing alternative strategies for treating these tumor types. Interestingly, our analyses also revealed that reduced expression of metabolic pathway genes and increased expression of non-canonical WNT pathway genes is associated with worse biochemical recurrence. These clinically relevant observations were only evidence in ETS<sup>+</sup> PCa tumors (Fig. 4b, c) but not in ETS<sup>−</sup> tumors (Supplementary Fig. S4). Caveats are that these differences may be due to lower numbers of ETS<sup>−</sup> tumors within the TCGA cohort and the relatively short median follow up time of 24 months. Alternatively, other yet unidentified pathways may be more clinically relevant among ETS<sup>−</sup> tumors.

The findings reported herein also have importance regarding health disparities in African American men (AAM), as ETS fusion events only occur in a minority of these prostate tumors [32–36]. Similarly, the proportion of AAM within the TCGA and GRID datasets used for the analyses in this current study had predominantly ETS<sup>−</sup> tumors (Supplementary Table S1). Our findings further emphasizes the lack of the clinical utility of the ETS<sup>+</sup> status

as a biomarker for PCa in AAM. While in-depth expression analyses are needed in this at-risk population to support this notion, we were unable to perform these studies due to insufficient number of AAM cases within each of the ETS subgroups.

In conclusion, this study identified distinct androgen-responsive genes in both ETS<sup>−</sup> and ETS<sup>+</sup> prostate tumors, and validated 131 AR-target genes that are regulated in an ETS-specific fashion. Importantly, AR-target genes in ETS<sup>−</sup> tumors were involved in metabolic and non-canonical WNT pathways, thus reclassifying these genes for use in targeted therapy discovery.

## Methods

### Prostate cancer tumor samples and microarray data

A total of 10,608 radical prostatectomy (RP) tumor expression profiles were used in this analysis. RNA-sequencing (RNAseqV2) gene expression data from the TCGA data portal for PCa and adjacent normal prostate tissue samples for 333 primary PCa tumors and associated clinical information, including molecular subtypes, was retrieved from the TCGA PRAD333 study [3, 37]. Retrospective and prospective microarray gene expression data and associated clinical information, including molecular subtypes was retrieved from Decipher GRID registry (NCT02609269). The retrospective samples (GRID) were from a matched cohort of African American (AAM) and European American (EAM) men microarray data from a previously published study [38, 39]. The prospective GRID cohort (GRID-prospective) was from clinical use of the Decipher test (GenomeDx Biosciences Laboratory, San Diego, CA). Samples from the GRID and GRID-prospective cohorts are non-overlapping and were analyzed using the same Affymetrix microarray platform. The SCAN algorithm was used for individual patient profile pre-processing and normalization [40]. COMBAT was used for debatching [41].

### ETS expression profiling and molecular characterization

ETS<sup>+</sup> status was assigned to samples that express high levels of *ERG* and the ETS family members *ETV1-5* and *FLII*. All other tumors were assigned as ETS<sup>−</sup>. For TCGA the following molecular subgroups were included in the ETS<sup>+</sup> group: 1-*ERG*, 2-*ETV1*, 3-*ETV2*, 4-*FLII*, and for the ETS<sup>−</sup> group: 5-*SPOP*, 6-*FOXA1*, 7-*IDH1*, 8-other, which were retrieved from the TCGA PRAD33 study [3]. The ETS status for GRID was determined as previously described for microarray-based ETS<sup>+</sup> (*ERG* or *ETV1/4/5*) subtyping [42]. For both TCGA and GRID the expression

levels and cut-points used for the molecular subtyping for ETS<sup>+</sup> and ETS<sup>−</sup> status were pre-specified in the datasets used for this analysis [3, 42].

AR targets genes were defined as those genes that were located within 25 kb downstream of an established AR binding site, based on publicly available AR ChIP-seq data from the studies of Massie et al. [14], Sharma et al. [16] and Pomerantz et al. [15].

### Statistical analysis

Principal component analysis (PCA) was performed using 101 androgen response genes as defined in the HALLMARK\_ANDROGEN\_RESPONSE gene-set [43]. The PCA models were calculated using Evince (Prediktera AB, Sweden). *t*-SNE (*t*-distributed stochastic neighbor embedding) was performed in MATLAB using a perplexity value of 30. All pairwise comparisons were performed using a two-tailed Mann–Whitney U test. Multiple testing was adjusted using FDR ( $q < 0.05$ ). Fold change was also used to define significantly expressed genes ( $\log_2$  FC > 0.585 for TCGA using RNAseq-based normalization and  $\log_2$  FC > 0.05 for GRID using SCAN normalization [40]). All statistical tests were performed using MATLAB R2017a (The MathWorks, Inc.). Gene set enrichment analysis (GSEA), using the molecular signatures database (MSigDB), was used to evaluate biologic pathway differences based on ETS status [44]. The HALLMARK [43] and the GO biological process gene sets were used for the GSEA analysis. Heatmaps (Fig. 3c and Fig. 4a) were generated as follows: samples and genes were ordered based on input-order, and for each gene the median expression in normal adjacent tissue was subtracted. All heatmaps were generated in MATLAB R2017a (The MathWorks, Inc.) and R version 3.3.3. The pathways shown in Fig. 4 were derived using gene ontology (GO) annotation and manual literature searches of the 131 selected genes (listed in Supplementary Table S5). The first PCA component was used to summarize gene sets (Fig. 4) and validated to ensure the robustness of the PCA model [17, 18]. Median cut was used to define high and low groups in the BCR analysis. Kaplan–Meier curves and log rank test was performed using MatSurv (github.com/aeberg1/MatSurv). This study has been approved by the Institutional Review Board.

**Acknowledgements** The authors thank the Prostate Cancer Foundation and the American Cancer Society for their generous support for this study. We also thank Genomedx Biosciences for generously providing the prostate tumor microarray data from the Decipher Genomic Resource Information Database registry (NCT02609269) for this study.

**Funding** This work was supported by grants from the Prostate Cancer Foundation Young Investigator Award (to K. Yamoah), American



Cancer Society MRS-17-108-01-TBG (to K. Yamoah), the V Foundation (K. Yamoah, J. Park), and by the NCI Comprehensive Cancer Center P30-CA076292 grant to the Moffitt Cancer Center. This work was also supported by the Cortner-Couch Endowed Chair for Cancer Research of the University of South Florida School of Medicine (J.L. Cleveland).

## Compliance with ethical standards

**Conflict of interest** Anders E. Berglund. Patents, Royalties, Other Intellectual Property: H. Lee Moffitt Cancer Center & Research Institute. Mandeep Takhar Employment: GenomeDx Biosciences, Mohammed Alshalalfa Employment: GenomeDx Biosciences, Elai Davicioni Employment: GenomeDx Biosciences, Leadership: GenomeDx Biosciences, Stock or Other Ownership: GenomeDx Biosciences, Nicholas Erho, Employment: GenomeDx Biosciences, Ashley E. Ross Stock or Other Ownership: GenomeDx Biosciences, Consulting or Advisory Role: GenomeDx Biosciences, Research Funding: GenomeDx Biosciences, Edward M. Schaeffer Consulting or Advisory Role: GenomeDx Biosciences, Metama, Eric A. Klein Consulting or Advisory Role: GenomeDx Biosciences, Berg, Genomic Health, Speakers' Bureau: Genomic Health, Robert B. Den Research Funding: GenomeDx Biosciences, Bruce J. Trock Consulting or Advisory Role: GenomeDx Biosciences, Research Funding: Myriad Genetics, John L. Cleveland: No relationship to disclose, Jong Y. Park: No relationship to disclose, Jasreman Dhillon: No relationship to disclose, Stephen J. Freedland: No relationship to disclose, Shivanshu Awasthi: No relationship to disclose, Chia-Ho Cheng: No relationship to disclose, Travis Gerke: No relationship to disclose, Robert J. Rounbehl: No relationship to disclose, Kosj Yamoah: No relationship to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Brooke GN, Bevan CL. The role of androgen receptor mutations in prostate cancer progression. *Curr Genom*. 2009;10:18–25.
- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310:644–8.
- Cancer Genome Atlas Research N. The molecular taxonomy of primary prostate. *Cancer Cell*. 2015;163:1011–25.
- Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*. 2007;448:595–9.
- Paulo P, Barros-Silva JD, Ribeiro FR, Ramalho-Carvalho J, Jeronimo C, Henrique R, et al. FLI1 is a novel ETS transcription factor involved in gene fusions in prostate cancer. *Genes Chromosomes Cancer*. 2012;51:240–9.
- Kunderfranco P, Mello-Grand M, Cangemi R, Pellini S, Mensah A, Albertini V, et al. ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer. *PLoS ONE*. 2010;5:e10547.
- Sun C, Dobi A, Mohamed A, Li H, Thangapazham RL, Furusato B, et al. TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation. *Oncogene*. 2008;27:5348–53.
- Weischenfeldt J, Simon R, Feuerbach L, Schlangen K, Weichenhan D, Minner S, et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell*. 2013;23:159–70.
- Kron K, Trudel D, Pethe V, Briollais L, Fleshner N, van der Kwast T, et al. Altered DNA methylation landscapes of polycomb-repressed loci are associated with prostate cancer progression and ERG oncogene expression in prostate cancer. *Clin Cancer Res*. 2013;19:3450–61.
- Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, et al. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet*. 2007;39:41–51.
- Berger MF, Lawrence MS, Demicheli F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470:214–20.
- Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*. 2010;17:443–54.
- Mani RS, Tomlins SA, Callahan K, Ghosh A, Nyati MK, Varambally S, et al. Induced chromosomal proximity and gene fusions in prostate cancer. *Science*. 2009;326:1230.
- Massie CE, Lynch A, Ramos-Montoya A, Boren J, Stark R, Fazli L, et al. The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *EMBO J*. 2011;30:2719–33.
- Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet*. 2015;47:1346–51.
- Sharma NL, Massie CE, Ramos-Montoya A, Zecchini V, Scott HE, Lamb AD, et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer Cell*. 2013;23:35–47.
- Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol*. 2011;7:e1002240.
- Berglund AE, Welsh EA, Eschrich SA. Characteristics and validation techniques for PCA-based gene-expression signatures. *Int J Genom*. 2017;2017:2354564.
- Khan AP, Rajendiran TM, Ateeq B, Asangani IA, Athanikar JN, Yocum AK, et al. The role of sarcosine metabolism in prostate cancer progression. *Neoplasia*. 2013;15:491–501.
- Lucarelli G, Ditonno P, Bettocchi C, Spilotros M, Rutigliano M, Vavallo A, et al. Serum sarcosine is a risk factor for progression and survival in patients with metastatic castration-resistant prostate cancer. *Future Oncol*. 2013;9:899–907.
- Massie CE, Mills IG, Lynch AG. The importance of DNA methylation in prostate cancer development. *J Steroid Biochem Mol Biol*. 2017;166:1–15.
- Tang S, Bhatia B, Maldonado CJ, Yang P, Newman RA, Liu J, et al. Evidence that arachidonate 15-lipoxygenase 2 is a negative cell cycle regulator in normal prostate epithelial cells. *J Biol Chem*. 2002;277:16189–201.
- Obinata D, Takada S, Takayama K, Urano T, Ito A, Ashikari D, et al. Abhydrolase domain containing 2, an androgen target gene, promotes prostate cancer cell proliferation and migration. *Eur J Cancer*. 2016;57:39–49.

24. Chen C, Wei X, Rao X, Wu J, Yang S, Chen F, et al. Cytochrome P450 2J2 is highly expressed in hematologic malignant diseases and promotes tumor cell growth. *J Pharmacol Exp Ther*. 2011;336:344–55.
25. Chen H, Li H, Chen Q. INPP4B overexpression suppresses migration, invasion and angiogenesis of human prostate cancer cells. *Clin Exp Pharmacol Physiol*. 2017;44:700–8.
26. Hodgson MC, Deryugina EI, Suarez E, Lopez SM, Lin D, Xue H, et al. INPP4B suppresses prostate cancer cell invasion. *Cell Commun Signal*. 2014;12:61.
27. Cyriac J, Haleem R, Cai X, Wang Z. Androgen regulation of spermidine synthase expression in the rat prostate. *Prostate*. 2002;50:252–61.
28. van der Graaf M, Schipper RG, Oosterhof GO, Schalken JA, Verhofstad AA, Heerschap A. Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer. *MAGMA*. 2000;10:153–9.
29. Li J, Ren S, Piao HL, Wang F, Yin P, Xu C, et al. Integration of lipidomics and transcriptomics unravels aberrant lipid metabolism and defines cholesteryl oleate as potential biomarker of prostate cancer. *Sci Rep*. 2016;6:20984.
30. Robinson DR, Zylstra CR, Williams BO. Wnt signaling and prostate cancer. *Curr Drug Targets*. 2008;9:571–80.
31. Brase JC, Johannes M, Mannsperger H, Falth M, Metzger J, Kacprzyk LA, et al. TMPRSS2-ERG -specific transcriptional modulation is associated with prostate cancer biomarkers and TGF-beta signaling. *BMC Cancer*. 2011;11:507.
32. Petrovics G, Liu A, Shaheduzzaman S, Furusato B, Sun C, Chen Y, et al. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. *Oncogene*. 2005;24:3847–52.
33. Mosquera JM, Mehra R, Regan MM, Perner S, Genega EM, Bueti G, et al. Prevalence of TMPRSS2-ERG fusion prostate cancer among men undergoing prostate biopsy in the United States. *Clin Cancer Res*. 2009;15:4706–11.
34. Magi-Galluzzi C, Tsusuki T, Elson P, Simmerman K, LaFargue C, Esgueva R, et al. TMPRSS2-ERG gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients. *Prostate*. 2011; 71:489–97.
35. Hu Y, Dobi A, Sreenath T, Cook C, Tadase AY, Ravindranath L, et al. Delineation of TMPRSS2-ERG splice variants in prostate cancer. *Clin Cancer Res*. 2008;14:4719–25.
36. Yamoah K, Johnson MH, Choerung V, Faisal FA, Yousefi K, Haddad Z, et al. Novel biomarker signature that may predict aggressive disease in African American men with prostate cancer. *J Clin Oncol*. 2015;33:2789–96.
37. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6:8971.
38. Yamoah K, Johnson MH, Choerung V, Faisal FA, Yousefi K, Haddad Z, et al. Novel biomarker signature that may predict aggressive disease in African American men with prostate cancer. *J Clin Oncol*. 2015;33:2789–96.
39. Alshalalfa M, Verhaegh GW, Gibb EA, Santiago-Jimenez M, Erho N, Jordan J, et al. Low PCA3 expression is a marker of poor differentiation in localized prostate tumors: exploratory analysis from 12,076 patients. *Oncotarget*. 2017;8:50804–13.
40. Piccolo SR, Withers MR, Francis OE, Bild AH, Johnson WE. Multiplatform single-sample estimates of transcriptional activation. *Proc Natl Acad Sci USA*. 2013;110:17778–83.
41. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–27.
42. Torres A, Alshalalfa M, Tomlins SA, Erho N, Gibb EA, Chelliserry J, et al. Comprehensive determination of prostate tumor ETS gene status in clinical samples using the CLIA decipher assay. *J Mol Diagn*. 2017;19:475–84.
43. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–25.
44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50.