This is the accepted manuscript of the article, which has been published in
Oncogene. 2020. https://doi.org/10.1038/s41388-020-1365-6

# AR and ERG Drive the Expression of Prostate Cancer Specific Long Noncoding RNAs

Annika Kohvakka[1], Mina Sattari[1], Anastasia Shcherban[1], Matti Annala[1], Alfonso Urbanucci[2], Juha Kesseli[1], Teuvo L.J. Tammela[1,3], Kati Kivinummi[1], Leena Latonen[4], Matti Nykter[1], and Tapio Visakorpi[1,5]

[1]Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Center, Tampere University Hospital, Tampere, Finland.
[2]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
[3]Department of Urology, Tampere University Hospital, Tampere, Finland
[4]Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
[5]Fimlab Laboratories Ltd, Tampere University Hospital, Tampere, Finland

Running title: AR and ERG driven long noncoding RNAs in prostate cancer

Corresponding author: Tapio Visakorpi, mailing address: Tampere University, Kalevantie 4, 33100, Tampere, Finland, tel: +358-50-3185829, email: tapio.visakorpi@tuni.fi.

# Abstract

Long noncoding RNAs (lncRNAs) play pivotal roles in cancer development and progression, and some function in a highly cancer-specific manner. However, whether the cause of their expression is an outcome of a specific regulatory mechanism or nonspecific transcription induced by genome reorganization in cancer remains largely unknown. Here, we investigated a group of lncRNAs that we previously identified to be aberrantly expressed in prostate cancer (PC), called TPCATs. Our high-throughput real-time PCR experiments were integrated with publicly available RNA-seq and ChIP-seq data and revealed that the expression of a subset of TPCATs is driven by PC-specific transcription factors (TFs), especially androgen receptor (AR) and ETS-related gene (ERG). Our *in vitro* validations confirmed that AR and ERG regulated a subset of TPCATs, most notably for *EPCART*. Knockout of *EPCART* was found to reduce migration and proliferation of the PC cells *in vitro.* The high expression of *EPCART* and two other TPCATs (*TPCAT-3-174133* and *TPCAT-18-31849*) were also associated with the biochemical recurrence of PC in prostatectomy patients and were independent prognostic markers. Our findings suggest that the expression of numerous PC-associated lncRNAs is driven by PC-specific mechanisms and not by random cellular events that occur during cancer development. Furthermore, we report three prospective prognostic markers for the early detection of advanced PC and show *EPCART* to be a functionally relevant lncRNA in PC.

# Introduction

Prostate cancer (PC) is the most common cancer and the third leading cause of male cancer death in developed countries (1). Androgen receptor (AR) is a transcription factor (TF) that plays an important role in the growth and development of normal prostate cells, and in PC tumorigenesis and progression. While the mechanisms of AR signaling have been widely investigated and utilized for treatment in advanced PC, the role of AR in primary PC is less clear. Previous studies have indicated that the AR cistrome is reprogrammed to novel genomic loci during tumorigenesis by master regulators, most notably FOXA1, HOXB13, and ETS family TFs, particularly ERG (2-4). ERG is involved in AR cistrome modulation by recruiting AR to novel genomic loci and binding to the same binding sites as AR (2, 3). Recent findings also indicate that ERG binds and redirects FOXA1 and HOXB13 to new genomic loci in TMPRSS2-ERG gene fusion positive PC (5). TMPRSS2-ERG gene fusion is the most frequent genetic aberration in PCs; it is found in ~50% of cases (6, 7), and it is an early event in PC development (8, 9), leading to overexpression of ERG. High ERG expression has been suggested to promote invasion and progression of PC cells (10, 11).

Long non-coding RNAs (lncRNA) are over 200 nucleotide long nonprotein-coding transcripts that are involved in various biological and pathological processes, including cancer (12). In prostate cancer, several lncRNAs have been discovered to have a potential role in PC tumorigenesis, progression, and metastasis (13). Furthermore, lncRNA tissue- and cancer-specific expression makes them ideal biomarkers for cancer detection and prediction (14). For example, PCA3, a highly PC-specific lncRNA, is a potent diagnostic marker (15), and a few other lncRNAs have been proposed as prognostic markers for advanced disease (16-18).

Although several lncRNAs have been found to be aberrantly expressed in PC samples (19, 20), their functional roles in the development of PC are poorly understood. Here, we aim to assess the possibility of regulation of PC-specific lncRNAs by AR and ERG. We focused our research on PC-associated transcripts (PCATs) that we previously discovered in the Tampere RNA-seq cohort (named TPCATs) (20). We used high-throughput real-

time PCR to identify TPCATs associated with PC progression in primary tumors and integrated publicly available RNA-seq and chromatin immunoprecipitation sequencing (ChIP-seq) data from PC patient and cell line samples to examine the regulative processes behind the expression of TPCATs. We found that the majority of studied TPCATs were associated with ERG overexpression, and they were putative targets of AR regulation. We also experimentally validated the regulation of TPCATs by AR and ERG. Finally, we identified three TPCATs whose expression was associated with PC progression. These findings provide insight into the importance of AR in the regulation of lncRNAs in PC and introduce potential novel prognostic markers to be used in the early detection of advanced PC.

# Materials and Methods

## Clinical samples

Fresh-frozen tissue samples from 87 radical prostatectomies were obtained from Tampere University Hospital (Tampere, Finland). The samples were snap frozen and stored in liquid nitrogen. The percentage of cancer in the samples varied from 30% to 80% (**Supplementary Table S1**). The mean age at diagnosis was 62.3 years (range: 40.3-71.8) and the mean prostate-specific antigen (PSA) at diagnosis was 10.1 ng/ml (range: 3.1-48.1) (**Supplementary Table S1**). The biochemical progression was defined as two consecutive samples with PSA ≥0.5 ng/ml. The use of clinical material was approved by the ethics committee of the Tampere University Hospital (Tampere, Finland). Written informed consent was obtained from all subjects.

## Cell lines and xenografts

The prostate cancer cell line LNCaP was obtained from American Type Cell Collection (ATCC, Manassas, VA, USA), and VCaP and DuCaP cells were kindly provided by Dr. Jack Schalken (Radboud University Nijmegen Medical Center, Nijmegen, the Netherlands). Parental LNCaP cells that were transfected either with empty pcDNA3.1(+) (LNCaP-pcDNA3.1) or wild-type AR-cDNA (LNCaP-ARhi) were previously established by our group (21). All cell lines were cultured as recommended by the suppliers and tested for mycoplasma contamination regularly. Previously established xenografts, LuCaP69 and LuCaP73, were provided by Dr. Robert L. Vessella (University of Washington, Seattle, WA, USA).

## Data acquisition and analysis

Our previously generated RNA-seq data from 28 untreated primary PC, 13 castration resistant PC (CRPC), and 12 benign prostatic hyperplasia (BPH) specimens (20) was used to identify TPCATs that are overexpressed in primary PC. To analyze the expression of TPCATs in The Cancer Genome Atlas prostate adenocarcinoma (TCGA-

PRAD) samples (7), transcriptome sequencing data for those samples was downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) and aligned against the hg19 human reference genome using Tophat-2.1.1. A catalog of gene exons was built by taking the union of Ensembl 75 splice variants and adding the novel TPCAT genes. The number of reads aligned to each gene was quantified using bedtools-2.26.0. Expression levels were normalized between samples using median-of-ratios normalization.

Unsupervised hierarchical clustering was performed for the matrix of ΔCt values, which was quantified relative to the genes' median expression across 34 TPCATs in 87 samples. Clustering was performed using the complete-linkage agglomerative clustering method based on the Euclidean distance matrix and visualized using R package gplots version 3.0.1.

TCGA-PRAD expression of TPCATs and over 3000 human genes linked to transcriptional regulation from the TFcheckpoint database (22) were compared with each other. The expression values were converted to log2, and the Pearson correlation coefficient was calculated for each TPCAT and TF in a pairwise manner.

To investigate the binding sites of TFs, called ChIP-seq peaks were retrieved from following public databases: AR, FOXA1, and HOXB13 ChIP-seq peaks in human prostate tumor samples (GSE56288), and VCaP ERG ChIP-seq peaks (GSM353647 and GSM2612457). The number of peaks for each TF was counted in the regulatory regions of TPCATs (-15kb/+2kb from transcription start site (TSS)). Next, the ChIP-seq peaks for all four TFs (AR, FOXA1, HOXB13 and ERG) were combined into union peaks, and each of the sites from the union peaks was checked for overlaps.

For determination of open chromatin sites, DNase-seq data in LNCaP was used. The data was retrieved from ENCODE portal (23) (https://www.encodeproject.org/) with the following identifier: ENCSR000EPF.

## Real-time PCR

For PCR-based analyses, RNA was extracted by using TRIzol (Thermo Fisher Scientific) or TRI Reagent (Sigma-Aldrich) following the manufacturer's instructions. RNA from knockdown and hormone deprivation samples were treated with DNase I and purified with RNeasy Mini Spin Columns (Qiagen) according to manufacturer's instructions.

For gene expression studies with Fluidigm Biomark HD, cDNA synthesis (Reverse Transcription Master Mix) and pre-amplification (Preamp Master Mix) reagents were purchased from Fluidigm and used according to the manufacturer's instructions. Quantification of expression was performed using a 48.48. Dynamic Array on a BioMark HD system (Fluidigm) with an EvaGreen-based detection system (SsoFast EvaGreen Supermix with Low ROX, Bio-Rad) following Fluidigm's instructions for fast gene expression analysis using EvaGreen on the BioMark HD system. Experiments with prostatectomy samples were performed as technical duplicates, and biological and technical triplicates were performed for gene knockdown and hormone deprivation studies. The primers used for the Fluidigm BioMark HD experiments are listed in **Supplementary Table S2**.

Relative expression values were calculated from Ct values, and the target gene measurements were normalized to *TBP* values and were averaged. Relative gene expression changes were calculated using the 2^-ΔΔCt-method. For the gene expression study using prostatectomies, ΔCt expression ratios for each gene were calculated relative to the gene's median expression. The percentage of the tissue that was cancerous in the prostatectomies was taken into account in the calculations [2^ΔCt*(100/cancer%)].

## Droplet digital PCR

Absolute quantification of transcripts was performed using a QX200 droplet digital PCR (ddPCR) system (Bio-Rad). cDNA was synthesized by Maxima RT (Thermo Fisher Scientific), and ddPCR was conducted with QX200 ddPCR EvaGreen Supermix (Bio-Rad) following the manufacturer's instructions. PCR was performed in a T100

Thermal Cycler (Bio-Rad). Experiments were carried out in biological or technical duplicates, and each sample was partitioned over 12,000 droplets. For data analysis, QuantaSoft ddPCR software (Bio-Rad) was used to calculate the absolute quantity of gene transcripts in the samples. Relative quantities of transcripts were normalized to *TBP*. The primers used for ddPCR experiments are listed in **Supplementary Table S2**.

## ChIP-qPCR

AR chromatin immunoprecipitation (ChIP) was performed as in Urbanucci *et al*. (24). A CFX96 Real-Time PCR Detection System (Bio-Rad) with Maxima SYBR Green (Thermo Fisher Scientific) was used for ChIP-qPCR studies, which were performed according to manufacturer's instructions in technical duplicates. The enrichment relative to IgG control was calculated as 2^-ΔCt. The primers used for ChIP-PCR are listed in **Supplementary Table S2**.

## Transfections for gene knockdown

siRNAs targeting AR, ERG, and a negative control siRNA (MISSION siRNA Universal Negative Control #1 or #2) were purchased from Sigma-Aldrich (**Supplementary Table S2**). Transfection reagent Lipofectamine RNAiMAX (Thermo Fisher Scientific) was used for transfecting siRNAs according to the manufacturer's instructions. Cells were reverse transfected with 25 nM siRNA and grown for 48 hours before RNA extraction and 72 hours before protein extraction.

## Androgen induction studies

The effect of androgens on to expression of TPCATs was studied in hormone-deprived cells. Cells were grown in phenol red-free RPMI 1640 medium (Lonza) with 10% charcoal/dextran-treated (CCS) FBS (Thermo Fisher Scientific) and 1% glutamine (Thermo Fisher Scientific) for four days. Hormone deprived cells were treated with 0 or 10 nM of DHT for 24 h.

## Western blotting

After knockdown experiments, cells were lysed in Triton-X lysis buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0,5% Triton x-100, 1 mM PMSF, 1 mM DTT and 1× Halt protease inhibitor cocktail (Thermo Fisher Scientific), after which the lysates were sonicated four times for 30 s at medium power with Bioruptor equipment (Diagenode), and cellular debris was removed by centrifugation. Proteins were separated by polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to PVDF membrane (Immobilon-P; Millipore). Primary antibodies against AR (AR-441; NeoMarkers; dilution 1:200), ERG (EPR3864; Abcam; dilution 1:5000), and pan-actin (ACTN05; NeoMarkers; 1:10 000) were used and detected by anti-mouse HRP-conjugated antibody produced in rabbit (dilution 1:2000-1:5000; DAKO) or by anti-rabbit HRP-conjugated antibody produced in swine (dilution 1:5000; DAKO) and Clarity Western ECL Substrate (Bio-Rad) with autoradiography.

## CRISPR-Cas9 knockout

To knockout *EPCART* in a prostate cancer cell line, the area covering the promoter and the 1$^{st}$ and 2$^{nd}$ exon of *EPCART* was targeted by CRISPR-Cas9 system. We used GenScript's CRISPR Gene Editing Services to perform the gene editing for LNCaP cells. Two single guide RNAs (sgRNAs; sequences listed in **Supplementary Table S2**) were designed and cloned by CloneEZ (GenScript) into AIO-1.0-Cas9-GGG-2A-EGFP vector by GenScript. The two vectors were co-transfected by Celetrix electroporation into LNCaP cells, and single cell clones were produced by GenScipt. The full deletion of *EPCART* was confirmed by PCR and Sanger sequencing for two cell clones (del-1 and del-2) and one clone without the deletion (WT) by GenScript. The expression of *EPCART* in the cell clones was analyzed by us using ddPCR.

## Cell viability assay

The proliferation of the *EPCART* deletion clones and the WT control clone was measured by alamarBlue (Thermo Fisher Scientific) cell viability reagent. 20 000 cells were plated in a normal medium on a 48 well plates

as 8 technical replicates. The alamarBlue reagent was used according to manufacturer's instructions; the fluorescence was measured (excitation 570 nm, emission 585 nm) at day 1, 3, 4, and 5 after plating by EnVision 2104 Multilabel Reader (Perkin-Elmer). The relative viability was calculated in relation to day 1.

## Wound healing assay

The migration of the *EPCART* deletion clones and the WT control clone was analyzed by wound healing assay. 500 000 cells were plated in a normal medium on a 24 well plate as 6 technical replicates and growth for 2 days before the experiment. Before imaging, fresh media was changed and a pipette tip was used to scratch a wound on the cell layer. Time-lapse imaging was performed over 24 h by Cell-IQ Automated Imaging and Analysis System (CM Technologies). Cell-IQ's Analyzer program was used to analyze the wound closure rate.

## Statistical analyses

Mann-Whitney *U* tests were used to analyze the association between ERG-positive and ERG-negative samples. Unpaired two-tailed Student's t-tests were used to calculate the significance between control and experimental conditions in PCR, cell viability, and wound healing experiments. P values $<0.05$ were considered statistically significant.

Kaplan-Meir survival analysis and log-rank tests were used to determine the progression-free survival between samples divided by their median expression. A Cox-proportional hazard model was utilized to model progression-free survival by measuring the size effects of multiple factors, including age at diagnosis, Gleason score, pathologic T status and PSA levels (**Supplementary Table S1**); TPCAT transcript expression levels were also included. Age at diagnosis was incorporated into the regression model as a continuous covariate, whereas each of the remaining factors was categorized into two or three groups depending on the type of covariate. The expression of each TPCAT transcript was binarized as either low or high using the gene's median ΔCt expression value as a baseline. Similarly, pathologic T status was categorized as either low (pT levels from 2 to

4) or high (pT levels 5 and 6). Gleason scores were divided into three groups: low (scores less than 7), intermediate (scores equal to 7) and high (scores from 8 to 10). Similar to the Gleason score, diagnostic PSA values were divided into three groups: low (PSA less than or equal to 10), intermediate (PSA from 10 to 19.9) and high (PSA greater than 20). Cox regression analysis was performed using coxph function from the survival package version 2.41-3 in R.

# Results

## ERG expression drives the aberrant expression of several TPCATs

Using transcriptome sequencing of clinical patient samples, we previously identified 145 TPCATs that were expressed specifically in primary PC, CRPC, or both (20). Here, we used Fluidigm BioMark HD real-time PCR system to evaluate the expression of TPCATs in 87 specimens of prostatectomy-treated patients obtained from the Tampere University Hospital PC cohort. Only TPCATs that had multiple exons and were overexpressed in primary PC were selected to ensure that TPCATs were transcribed from genuine genes. In total, the expression of 34 TPCATs was investigated. Hierarchical clustering of the real-time PCR gene expression data of TPCATs and their expression relative to common PC-related TFs ERG, ETV1, FOXA1, and AR in the same samples revealed that expression of multiple TPCATs was associated with the expression of ERG (**Figure 1**).

To further assess the observed ERG association further, we divided the PC samples into ERG-positive and ERG-negative groups based on their ERG gene fusion status and expression (25) (**Supplementary Table S3**) and examined the expression of TPCATs in these two sample groups. Based on this analysis, we found 17 of the TPCATs to be differentially expressed ($p<0.05$) in ERG-positive vs. ERG-negative samples (**Supplementary Figure S1a**). To validate the identified ERG association in another dataset, we investigated the expression of TPCATs in the TCGA-PRAD data collection (7) (**Supplementary Table S3**). Indeed, all TPCATs found to associate with ERG expression based on our Tampere cohort were also found to be associated with ERG expression in the TCGA-PRAD dataset ($p<0.05$) (**Supplementary Figure S1b**). Furthermore, five additional TPCATs were discovered to be ERG-associated in the TCGA-PRAD dataset. In total, 22 out of 34 TPCATs were found to be associated with ERG expression.

Next, we compared the expression of the 34 TPCATs to expression of over 3000 validated human TFs (22) at the mRNA level in the expression data from TCGA-PRAD. Indeed, among the TFs, the expression of ERG showed

the strongest correlation with the expression of TPCATs, with 10 TPCATs positively correlating with ERG (Pearson's r>0.4 of log2 expression values) (**Supplementary Table S4**). When the expression of each of the TPCATs was compared to the expression of other TPCATs, 11 TPCATs showed positive correlation with each other (Pearson's r>0.4 of log2 expression values). Ten of these TPCATs were positively associated with ERG, and they only correlated with other ERG-associated TPCATs (**Supplementary Table S4**). Therefore, the similar expression profiles of TPCATs could be mostly explained by ERG overexpression. Together, these results imply that ERG has a significant role in the regulation of several TPCATs.

To assess how ERG regulates TPCAT expression, we used publicly available ERG ChIP-seq data to look specifically into the putative regulatory region (-15 kb/+2 kb from TSS) of TPCATs in VCaP cells. VCaP cells are a PC cell line harboring the TMPRSS2-ERG fusion gene and expressing ERG. Of the ERG-associated TPCATs, over 70% (16 out of 22) had at least one ERG binding site in their regulatory regions, but ERG binding sites in such regions were only found in one third of the TPCATs (4 out of 12) that were not associated with ERG expression ($p<0.05$, Fisher's exact test) (**Figure 2; Supplementary Table S5**). In addition, the vast majority of all the TPCAT-associated ERG peaks (31 out of 35) were located in the regulatory regions of ERG-associated TPCATs (**Supplementary Table S5**).

To validate that the expression of TPCATs was ERG-dependent, we performed siRNA knockdown of ERG in ERG-expressing PC cell lines (VCaP and DuCaP) and measured the gene expression by Fluidigm BioMark HD (**Supplementary Figure S2a-b**). When a log2-fold change <-1 or >1 was used as a cut-off value, nearly half of the TPCATs (16 out of 34) were verified to be ERG regulated in either VCaP or DuCaP cells (**Figure 2; Supplementary Table S6**). Ten of those were in the group of ERG expression-associated TPCATs.

## Majority of TPCATs are targets of AR

Since prior studies have indicated that ERG interacts with AR in early PC (2, 3, 5) and that multiple lncRNAs are part of the AR signaling pathway (26-29), we hypothesize that AR could also play a role in the regulation of TPCATs. First, we examined the publicly available AR ChIP-seq data from primary PC tumors as well as corresponding normal tissue (4) for AR binding sites (ARBS) in the regulatory region (-15 kb/+2 kb from TSS) of TPCATs. We found that nearly 70% of the TPCATs (23 out of 34) showed ARBS in PC (**Figure 2; Supplementary Table S5**). Of those TPCATs, two-thirds (22 out of 34) had more ARBSs in cancer tissues than they had in normal tissues (**Supplementary Table S5**). There were over 6 times more AR binding sites in the regulatory region of TPCATs present in PC than there were in normal samples ($p<0.001$, Mann-Whitney *U*-test) (**Supplementary Table S5**).

We further investigated the role of AR in the regulation of TPCATs in PC cell lines expressing AR (LNCaP, DuCaP, and VCaP). We performed AR knockdown and DHT stimulation experiments, followed by gene expression analysis by Fluidigm BioMark HD. We verified the success of the AR knockdown and DHT stimulation by monitoring AR levels and the stimulation of target genes, respectively (**Supplementary Figure S3a-c**). More than half of TPCATs were found to be strongly affected (log2-fold change <-1 or >1) by either AR knockdown (21 out of 34) or DHT stimulation (19 out of 34) (**Supplementary Table S6**). Of these, 7 TPCATs were affected in opposite ways by both treatments in the same cell line; however, a similar but weaker effect was also noticeable with several additional TPCATs (**Figure 2, Supplementary Table S6**).

## AR and ERG colocalize in the regulatory regions of TPCATs together with FOXA1 and HOXB13

AR and ERG partially target the same genes (3), and FOXA1 and HOXB13 are colocalized with both AR and ERG (4, 5); therefore, we investigated whether FOXA1 and HOXB13 also regulate TPCATs. We located their binding sites in TPCAT regulatory regions (-15 kb/+2 kb from TSS) as described above for AR and ERG. For FOXA1 and HOXB13, we used previously established ChIP-seq data in PC tumor specimens (4). The vast majority of all the

TPCAT-related ERG binding sites (28 out of 35) were co-occupied by AR (**Figure 3a**). These shared binding sites were found in among half of the TPCATs (17 out of 34), of which nearly all (15 out of 17) were associated with ERG expression (**Figure 2**). In addition, the majority of these TPCATs had FOXA1 and/or HOXB13 bound in their regulatory regions (22 out of 34), and nearly half (16 out of 34) were co-occupied by both TFs (**Figure 2; Supplementary Table S5**). HOXB13 binding (39 peaks) was observed more frequently than FOXA1 binding (22 peaks) (**Figure 3a**), which is concordant with the previous results from the whole PC genome (4). The number of FOXA1 and HOXB13 binding sites co-occupied by AR (78%) in TPCAT regulatory regions (**Figure 3a**) was slightly, but not significantly, higher than what was globally detected in PC (62%) (**Figure 3b**).

In total, we found AR, ERG, FOXA1, and HOXB13 to co-occupy 25% (15 out of 61) of all TPCAT-related binding sites; there were only 7% global co-binding of these TFs ($p<0.0001$, Pearson chi-square with Yates' correction) (**Figure 3a-b**). One third of the TPCATs (13 out of 34) had at least one binding site from one of the four TFs (**Figure 2**). These findings suggest that all four TFs are involved in the regulation of TPCATs.

## *EPCART* is a clinically relevant lncRNA that is regulated by prostate cancer-driving TFs

From our experiments, it became evident that *TPCAT-2-180961*, officially termed ERG-positive PC-associated androgen responsive transcript (EPCART), was highly expressed in PCs overexpressing ERG (**Figure 1; Supplementary Figure S1a-b**), and data suggested that it was regulated by both AR and ERG (**Figure 2**). According to our previously generated RNA-seq data, *EPCART* is located in chromosome 2 and has five exons (**Figure 4a**). Publicly available DNase-seq data in LNCaP cells (30) showed chromatin to be open where there were three ARBS located in the regulatory region of *EPCART* (**Figure 4a**). These ARBS were also highly PC-associated and were co-occupied by FOXA1 and/or HOXB13 (**Figure 4a**). To investigate AR binding to the TSS of *EPCART* in greater detail, we used AR ChIP-qPCR to analyze AR binding in LNCaP cells with and without DHT stimulation, and we analyzed AR binding in LuCaP xenografts with and without AR gene amplification. We demonstrated increased AR binding upon DHT stimulation in LNCaP cells overexpressing AR (LNCaP-ARhi)

compared to that of the parental LNCaP cells (**Figure 4b**). Additionally, LuCaP69 xenograft containing AR gene amplification (31) showed more AR binding to *EPCART* compared to what was observed in the LuCaP73 xenograft without amplification (**Figure 4c**). To thoroughly investigate whether *EPCART* is regulated by AR, we performed AR knockdown and DHT induction experiments in DuCaP cells and analyzed the variations in gene expression by ddPCR. In these experiments, the expression of *EPCART* was significantly downregulated after AR knockdown (**Figure 4d**), while DHT induced the expression of *EPCART* (**Figure 4e**). These results confirm that *EPCART* is an AR-regulated lncRNA.

To further elaborate the functional role of *EPCART* in the PC cells, we deleted *EPCART* form LNCaP cells (*EPCART*-del) using CRISPR/Cas9. Two sgRNAs were designed to target the area covering the promoter, the 1$^{st}$ exon, and the 2$^{nd}$ exon of *EPCART* (**Supplementary Figure S4a**). The full deletion of this area was confirmed by PCR and Sanger sequencing in two clones, and a wild type (WT) clone was used as a control (**Supplementary Figure S4b**). To verify the decrease of the *EPCART* expression, we quantified the absolute amount of *EPCART* transcripts by ddPCR by using two primer pairs, pair #1 targeting the deleted exon 2 and pair #2 targeting exons outside of the deleted area (**Supplementary Figure S4a**). We detected a considerable reduction, although not a full abolition, of the *EPCART* transcript in both *EPCART*-del clones when compare to the WT clone (**Figure 4f**). To assess whether this reduction influenced cell functions, we performed cell viability and wound healing assays for all three clones. Indeed, both cell proliferation (**Figure 4g**) and migration (**Figure 4h, Supplementary Figure S4c**) were significantly reduced in both *EPCART*-del clones as compared to the control cells. This indicates that *EPCART* has functions that may contribute to PC progression.

As some lncRNAs have been proposed as prognostic biomarkers of PC (16, 17), we were interested in testing whether *EPCART* could be utilized for the same purpose. Therefore, we assessed the association of TPCAT expression with the prognosis in prostatectomy-treated patients. Kaplan-Meier analysis revealed that high expression of *EPCART* was associated with short biochemical progression-free survival (**Figure 4i**). Furthermore, multivariate Cox regression analysis showed that the expression of *EPCART* had independent prognostic value

306 (other parameters included were age, Gleason score, diagnostic PSA, and pathological T stage (pT)) (**Table 1**).
307 Prompted by this, we further investigated whether the expression of other TPCATs was associated with PC
308 progression. We found that *TPCAT-3-174133* and *TPCAT-18-31849* were also associated with a short
309 biochemical progression-free survival in PC patients (**Supplementary Figure S5**). Both of these lncRNAs also had
310 independent prognostic value (**Supplementary Table S7**).

# Discussion

Various transcriptome studies in recent years have shown that lncRNAs are aberrantly expressed in cancers, and this expression is often cancer type-specific (19, 32-34). However, it is largely unknown whether a specific mechanism drives the expression of these lncRNAs, or whether it is the result of the genome reorganization in cancer cells that leads to nonspecific transcription. Previously, we discovered 145 lncRNAs (TPCATs) to be associated with primary PC and/or CRPC (20). Here, we showed that the expression of a selection of TPCATs is regulated by TFs that drive PC, especially AR and ERG, which could explain the high PC specificity of these TPCATs. Thus, this data suggests that the expression of at least these identified TPCATs is not the result of random transcriptional events and might have mechanistic significance for PC biology.

TMPRSS2-ERG gene fusion has previously been associated with early-onset PC and high-risk tumors as a result of ERG overexpression (9, 35-37), although the exact mechanisms behind its function are still unclear. In the current study, we showed a strong association between the expression of ERG and PC-associated lncRNAs in primary tumors. In addition to PCAT5, which we previously discovered to be an ERG-regulated TPCAT (20), we found that the majority (65%) of the investigated TPCATs were associated with overexpression of ERG. ERG also directly bound to the regulatory regions of more than half (59%) of the TPCATs, and it was primarily associated with those that were ERG-associated. Together, these results revealed that ERG had a regulatory role in the expression of TPCATs, which we confirmed for ten of the ERG-associated TPCATs by ERG *in vitro* knockdown studies. However, this portion could potentially be even greater, as we experienced some technical variation in the results that was most likely due to the very low expression level of some of the TPCATs (including *EPCART*) in the cell lines used for these studies. The same applies for ERG ChIP-seq data that has thus far only been generated from VCaP cells, while no data has been generated from patient samples. This could also explain why a prior study did not find a significant association between ERG and PC-associated lncRNAs (38).

Previous studies have shown several lncRNAs to be associated with AR signaling in PC (26-29), and our results suggest the same for most TPCATs. Nearly 70% of the TPCATs had ARBS in their regulatory region in PC, and there was significantly less in the benign prostate, in which the expression of TPCATs is also less abundant (20). We found that the expression of most TPCATs (62%) are androgen sensitive, and that AR knockdown had an effect on the majority of the TPCATs (56%). However, only seven TPCATs were oppositely affected by both androgen induction and AR knockdown. This could be due to the exceptionally high expression of AR in these cells. The high AR levels also explain why we could not demonstrate the reduction of *KLK3*, a well-known target gene of AR, in DuCaP and VCaP cells. On the other hand, we could detect a significant reduction of *TMPRSS2*, another target gene of AR, in VCaP cells, indicating that at least some of the AR downstream targets are efficiently affected by AR silencing in these cells. Thus, it is plausible that AR knockdown was not efficient enough to affect the expression of all the AR-regulated TPCATs in these experiments.

Because ERG is known to physically interact with AR and to bind to the downstream AR genes (2), we investigated whether this could also be the case for TPCATs. Indeed, we found that over 80% of ERG binding sites were co-occupied by AR within the regulatory regions of TPCATs, and the majority of those shared sites were located near ERG-associated TPCATs. In addition, we discovered that FOXA1 and HOXB13 co-occupy the majority of AR and ERG binding sites, implying that regulatory mechanisms that have been found to play a role in primary PC (4, 5), have a similar role in the regulation of TPCATs.

One of the TPCATs, *EPCART*, stood out early on in our analysis as being highly associated with ERG overexpression as well as being regulated by the AR signaling pathway. Our *EPCART* knockout studies found *EPCART* to effect the migration and proliferation of the PC cells, indicating *EPCART* to have a function in PC progression. Furthermore, in our prostatectomy cohort, we discovered that the high expression of *EPCART* and two other TPCATs were independent prognostic factors for biochemical recurrence. Interestingly, *EPCART* has also been previously associated with the development of clinical metastasis and PC-related death (38). Jointly, these results indicate that *EPCART* is a potential prognostic marker and therapeutic target for aggressive PC.

Further studies are warranted to test the specificity and sensitivity of *EPCART* and to analyze its performance in a larger cohort, and to analyze the downstream mechanisms of its action more in depth.

In summary, we report that the majority of TPCATs investigated here are strongly associated with AR and other cooperative TFs, most importantly with ERG, in fusion-positive tumors. We found that the expression of many of the TPCATs was regulated by these TFs. Additionally, three of the TPCATs were independently associated with PC progression, most notably *EPCART* that we also found to promote the migration and proliferation of the PC cells *in vitro*. Together, these findings demonstrate that *EPCART* has functions relevant for PC progression. Thus, we conclude that *EPCART* is a prospective prognostic marker for advanced PC and an intriguing candidate for further functional studies investigating its potential function as a therapeutic target in PC.

# Acknowledgements

This study was supported by grants from the Academy of Finland (TV 317755, MN 310829, LL 317871), Sigrid Juselius Foundation (TV, LL), Cancer Society of Finland, Business Finland, the Finnish Cultural Foundation (AK), the European Union's Horizon 2020 (MS, TransPot - 721746), Norwegian Cancer Society grant (AU 198016-2018), Research collegium of the University of Tampere/IASR (KK). The authors want to thank Jenni Jouppila, Paula Kosonen, Riina Kylätie, Päivi Martikainen, Hanna Selin, and Marika Vähä-Jaakkola for their technical assistance and Tampere Imaging Facility (TIF) for their service. The results published here are in part based upon data generated by The Cancer Genome Atlas project (dbGaP Study Accession: phs000178.v9.p8) established by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov. We acknowledge ENCODE Consortium and the ENCODE production laboratories for generating the DNase-seq data.

# Conflict of interest

The authors declare no potential conflicts of interest.

# References


(1) Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65: 87-108.

(2) Yu J, Yu J, Mani R, 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: 443-454.

(3) Chen Y, Chi P, Rockowitz S, Iaquinta PJ, Shamu T, Shukla S, *et al*. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. Nat Med 2013;19: 1023-1029.

(4) 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. Nature Genet 2015;47: 1346.

(5) Kron KJ, Murison A, Zhou S, Huang V, Yamaguchi TN, Shiah YJ, *et al*. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nature Genet 2017;49: 1336-1345.

(6) Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X, *et al*. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310: 644-648.

(7) Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell 2015;163: 1011-1025.

(8) Cerveira N, Ribeiro FR, Peixoto A, Costa V, Henrique R, Jerónimo C, *et al*. TMPRSS2-ERG gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. Neoplasia 2006;8: 826-832.

(9) Perner S, Mosquera J, Demichelis F, Hofer MD, Paris PL, Simko J, *et al*. TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. Am J Surg Pathol 2007;31: 882-888.

(10) Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, *et al*. Role of the TMPRSS2-ERG Gene Fusion in Prostate Cancer. Neoplasia 2008;10: 17-IN9.

(11) Carver BS, Tran J, Gopalan A, Chen Z, Shaikh S, Carracedo A, *et al*. Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. Nat Genet 2009;41: 619-624.

(12) Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. Cancer Cell 2016;29: 452-463.

(13) Martens-Uzunova ES, Böttcher R, Croce CM, Jenster G, Visakorpi T, Calin GA. Long Noncoding RNA in Prostate, Bladder, and Kidney Cancer. Eur Urol

2014;65: 1140-1151.

(14) Sahu A, Singhal U, Chinnaiyan AM. Long Noncoding RNAs in Cancer: From Function to Translation. Trends in Cancer 2015;1: 93-109.

(15) Hessels D, Klein Gunnewiek, Jacqueline M T, van Oort I, Karthaus HFM, van Leenders, *et al*. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol 2003;44: 16.

(16) Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekaran SM, *et al*. RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. Lancet Oncol 2014;15: 1469-1480.

(17) Shukla S, Zhang X, Niknafs YS, Xiao L, Mehra R, Cieslik M, *et al*. Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer. Neoplasia 2016;18: 489-499.

(18) White NM, Zhao SG, Zhang J, Rozycki EB, Dang HX, McFadden SD, *et al*. Multi-institutional Analysis Shows that Low PCAT-14 Expression Associates with Poor Outcomes in Prostate Cancer. Eur Urol 2017;71: 257-266.

(19) Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, *et al*. Transcriptome Sequencing Identifies PCAT-1, a Novel lincRNA Implicated in Prostate Cancer Progression. Nat Biotechnol 2011;29: 742-749.

(20) Ylipää A, Kivinummi K, Kohvakka A, Annala M, Latonen L, Scaravilli M, *et al*. Transcriptome Sequencing Reveals PCAT5 as a Novel ERG-Regulated Long Noncoding RNA in Prostate Cancer. Cancer Res 2015;75: 4026-4031.

(21) Waltering KK, Helenius MA, Sahu B, Manni V, Linja MJ, Jänne OA, *et al*. Increased expression of androgen receptor sensitizes prostate cancer cells to low levels of androgens. Cancer Res 2009;69: 8141-8149.

(22) Chawla K, Tripathi S, Thommesen L, Lægreid A, Kuiper M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. Bioinformatics 2013;29: 2519-2520.

(23) Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, *et al*. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 2018;46: D794-D801.

(24) Urbanucci A, Sahu B, Seppälä J, Larjo A, Latonen LM, Waltering KK, *et al*. Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer. Oncogene 2012;31: 2153-2163.

(25) Boormans JL, Porkka K, Visakorpi T, Trapman J. Confirmation of the association of TMPRSS2(exon 0):ERG expression and a favorable prognosis of primary prostate cancer. Eur Urol 2011;60: 183-184.

(26) Liuqing Yang, Chunru Lin, Chunyu Jin, Joy C Yang, Bogdan Tanasa, Wenbo Li, *et al*. lncRNA-dependent mechanisms of androgen-receptor- regulated gene activation programs. Nature 2013;500: 598.

(27) Takayama K, Horie-Inoue K, Katayama S, Suzuki T, Tsutsumi S, Ikeda K, *et al*. Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. EMBO J 2013;32: 1665-1680.

(28) Zhang A, Zhao JC, Kim J, Fong K, Yang YA, Chakravarti D, *et al*. LncRNA HOTAIR Enhances the Androgen-Receptor-Mediated Transcriptional Program and Drives Castration-Resistant Prostate Cancer. Cell Rep 2015;13: 209-221.

(29) Zhang Y, Pitchiaya S, Cieślik M, Niknafs YS, Tien JC-, Hosono Y, *et al*. Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. Nat Genet 2018;50: 814-824.

(30) The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489: 57-74.

(31) Linja MJ, Savinainen KJ, Saramäki OR, Tammela TL, Vessella RL, Visakorpi T. Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res 2001;61: 3550-3555.

(32) White NM, Cabanski CR, Silva-Fisher JM, Dang HX, Govindan R, Maher CA. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. Genome Biol 2014;15: 429.

(33) Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, *et al*. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. Oncotarget 2014;5: 9864-9876.

(34) Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, *et al*. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell 2015;28: 529-540.

(35) Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera J, Setlur S, *et al*. TMPRSS2:ERG Fusion-Associated Deletions Provide Insight into the Heterogeneity of Prostate Cancer. Cancer Res 2006;66: 8337-8341.

(36) Nam RK, Sugar L, Wang Z, Yang W, Kitching R, Klotz LH, *et al*. Expression of TMPRSS2:ERG gene fusion in prostate cancer cells is an important prognostic factor for cancer progression. Cancer Biology & Therapy 2007;6: 40-45.

(37) 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: 159-170.

(38) Böttcher R, Hoogland AM, Dits N, Verhoef EI, Kweldam C, Waranecki P, *et al*. Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. Oncotarget 2015;6: 4036-4050.

# Figure Legends

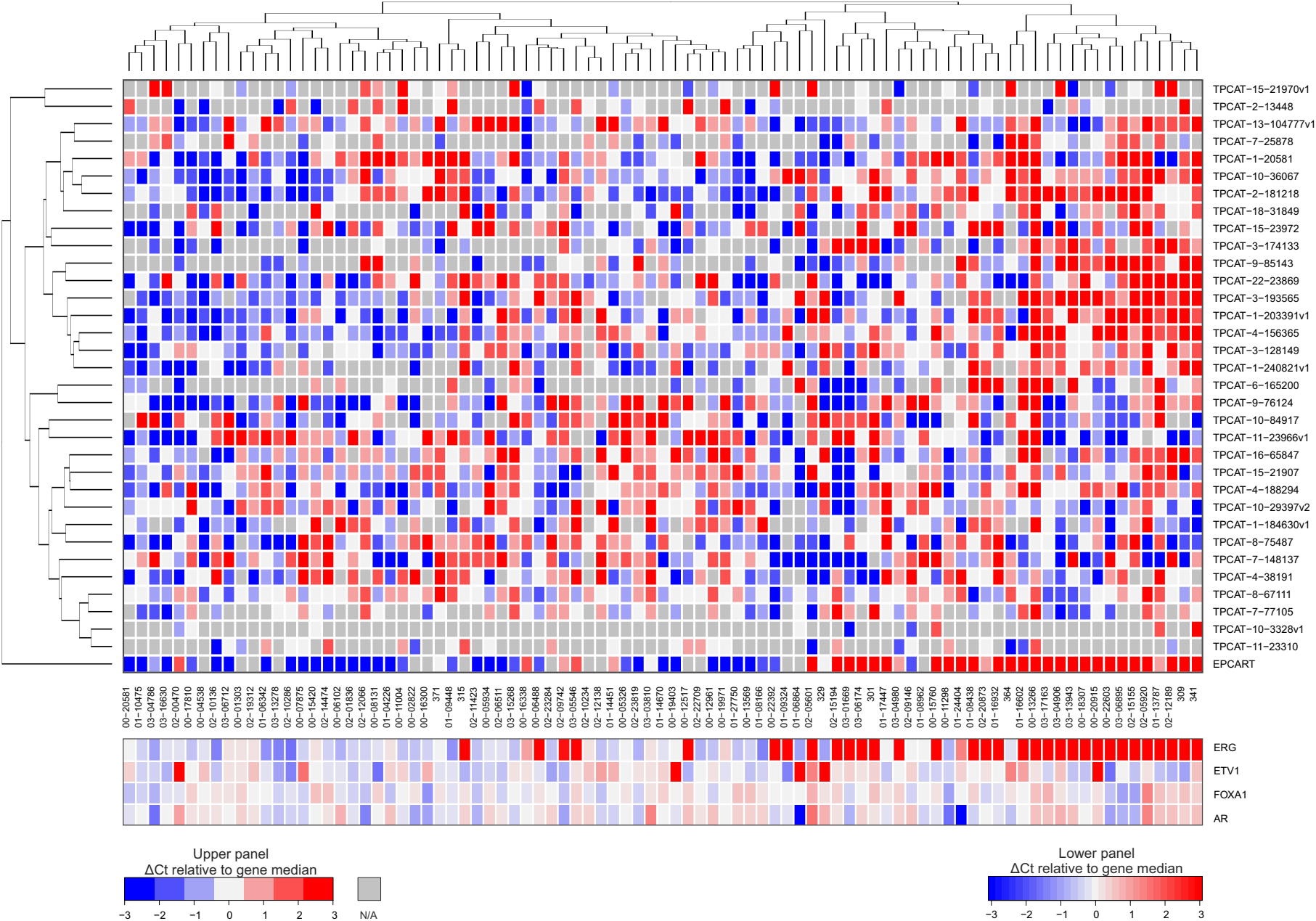
**Figure 1.** ERG overexpression correlates with the expression TPCATs. The expression of 34 TPCATs was analyzed in 87 prostatectomy specimens by qRT-PCR using Fluidigm Biomark HD. Hierarchical clustering revealed multiple TPCATs that were abundantly expressed in samples overexpressing ERG.

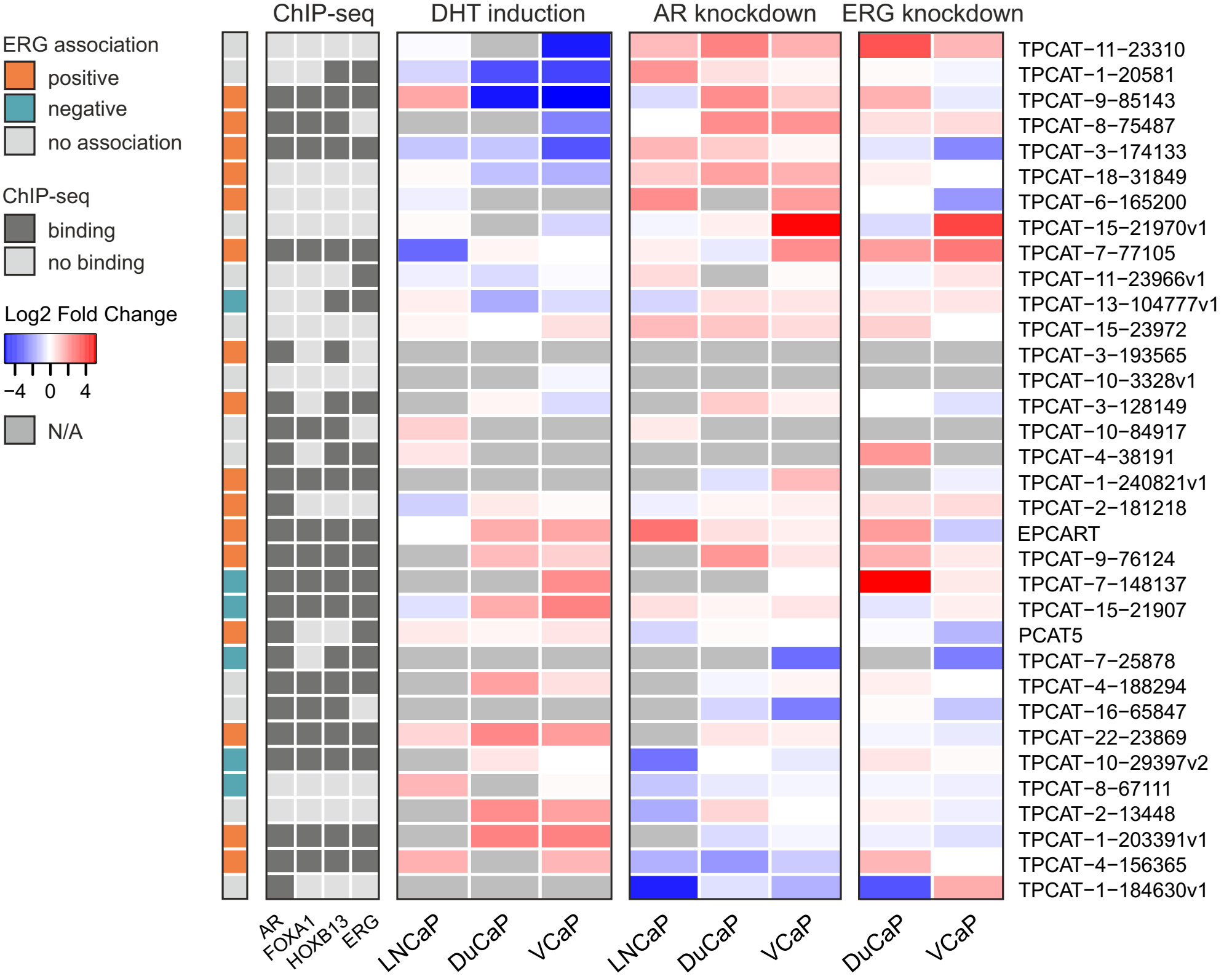
**Figure 2.** Several TPCATs are regulated by AR and ERG. The ERG association of TPCATs based on the expression of TPCATs in clinical samples (**Supplementary Figure S1a-b**) is marked in the column on the left. ChIP-seq peaks for different TFs (AR, ERG, FOXA1, and HOXB13) found in the regulatory region (-15 kb/+2 kb from TSS) of TPCATs are marked in the ChIP-seq panel. DHT induction was performed on hormone deprived cells after day 4 with 0 nM or 10 nM of DHT for 24 h. For AR and ERG knockdown experiments, cells were treated with target or control siRNA (25 nM) for 48 h. In both induction and knockdown experiments, the expression of TPCATs was measured in three biological and technical replicates by qRT-PCR using Fluidigm Biomark HD, and levels were normalized against *TBP*. Differential expression was calculated as log2-fold change between control and treated samples.

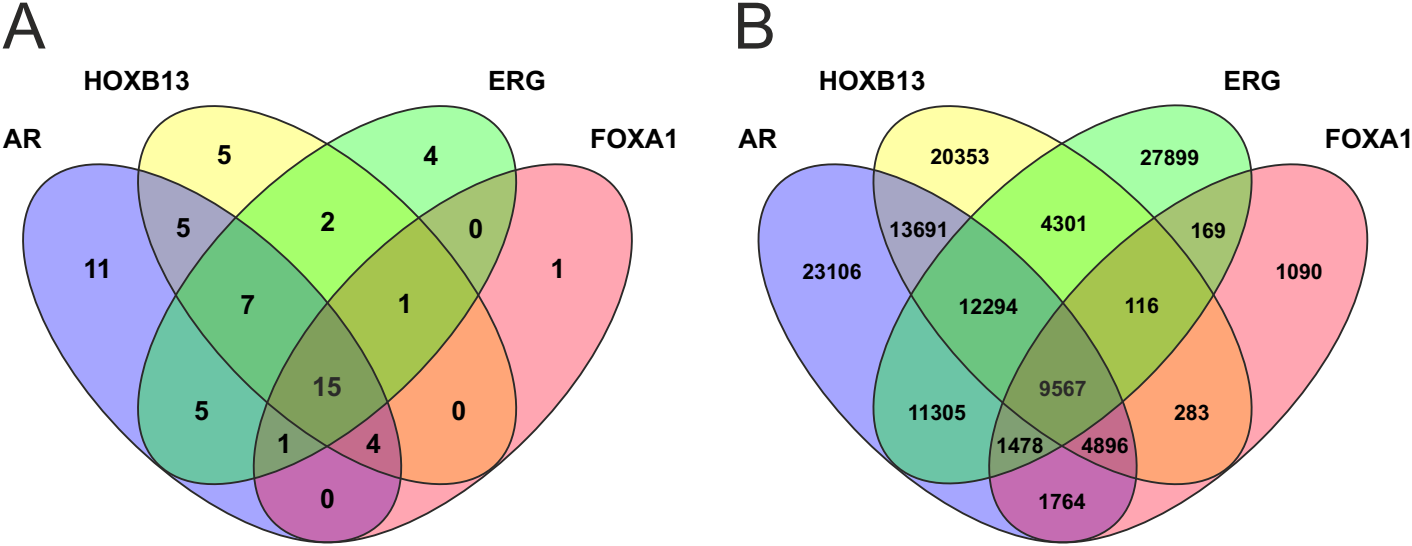
**Figure 3.** TFs that drive PC colocalize in the regulatory regions of TPCATs. **a**, Number of peaks detected in ChIP-seq data for AR, ERG, FOXA1, and HOXB13 in the regulatory region (-15 kb/+2 kb from TSS) of TPCATs. **b**, Total number of AR, ERG, FOXA1, and HOXB13 ChIP-seq peaks detected in the genome.

**Figure 4.** *EPCART* is an androgen responsive lncRNA that associates with PC progression. **a**, Publicly available ChIP-seq data was used to determine the binding sites for AR, ERG, FOXA1, and HOXB13 in the regulatory region of *EPCART*. DNase-seq data from LNCaP cells (by ENCODE) revealed the open chromatin sites co-occupied by TFs, and RNA-seq data from a primary PC sample in the Tampere cohort identified the transcript structure of *EPCART*. **b-c**, qPCR was performed following AR-ChIP from LNCaP (B) and LuCaP (C) samples using primers designed for AR peaks near the TSS of *EPCART*. LNCaP cells were hormone starved 4 days before they

were treated with either 0 nM of DHT (-DHT) or 1 nM of DHT (+DHT) for 24 h. LuCaP69 and LuCaP73 are CRPC-derived xenografts, of which LuCaP69 exhibits AR amplification, while LuCaP73 does not (31). The fold enrichment was calculated relative to IgG control (not shown in B) in technical duplicates. LNCaP-crtl, LNCaP cells stably expressing empty pcDNA3.1(+) vector; LNCaP-ARhi, LNCaP cells stably expressing high wt-AR from a pcDNA3.1(+) vector. Error bars, SD; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; data was assessed with an unpaired two-tailed t-test. **d**, AR siRNA (siAR) knockdown (25 nM) in DuCaP cells led to decrease of *EPCART* and *AR* expression when compared to control siRNA (NC). Expression of both *EPCART* and *AR* was analyzed by ddPCR in biological duplicates using *TBP* as a reference gene. Error bars, SD; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; data was assessed with an unpaired two-tailed t-test. **e**, DHT induction in DuCaP cells led to an increase in *EPCART* expression. DuCaP cells were hormone starved 3 days before they were treated with either with 0 nM of DHT (-DHT) or 10 nM of DHT (+DHT) for 24 h. Expression of *EPCART* was analyzed by ddPCR in biological duplicates, in which *TBP* was used as a reference gene. Error bars, SD; *, $p<0.05$; **, $p<0.01$; ***, and $p<0.001$; data was assessed with an unpaired two-tailed t-test. **f**, *EPCART* deletion in LNCaP cells (del-4 and del-56) led to a decrease in the amount of *EPCART* transcripts. Absolute quantification of *EPCART* transcripts was performed by ddPCR by using two primer pairs (ex 2-3 and ex 3-4) in technical duplicates. The relative concentration of *EPCART* transcripts was calculated in relation to *TBP*. Error bars, SD; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; data was assessed with an unpaired two-tailed t-test. **g-h**, Proliferation (G) and migration (H) was decreased in *EPCART*-del cells when compared to WT LNCaP cells. Cell viability was measured by alamarBlue over 5 days, and wound healing was analyzed by Cell-IQ time-lapse imaging over 24h. Error bars, range; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; data was assessed with an unpaired two-tailed t-test. **i**, Kaplan-Meier analysis was used for progression-free survival of PC patients who were grouped based on median expression of *EPCART*. P values were calculated by log-rank test. HR = hazard ratio.


TPCAT−15−21970v1
TPCAT−2−13448
TPCAT−13−104777v1
TPCAT−7−25878
TPCAT−1−20581
TPCAT−10−36067
TPCAT−2−181218
TPCAT−18−31849
TPCAT−15−23972
TPCAT−3−174133
TPCAT−9−85143
TPCAT−22−23869
TPCAT−3−193565
TPCAT−1−203391v1
TPCAT−4−156365
TPCAT−3−128149
TPCAT−1−240821v1
TPCAT−6−165200
TPCAT−9−76124
TPCAT−10−84917
TPCAT−11−23966v1
TPCAT−16−65847
TPCAT−15−21907
TPCAT−4−188294
TPCAT−10−29397v2
TPCAT−1−184630v1
TPCAT−8−75487
TPCAT−7−148137
TPCAT−4−38191
TPCAT−8−67111
TPCAT−7−77105
TPCAT−10−3328v1
TPCAT−11−23310
EPCART
00−20581 01−10475 03−04786 03−16630 02−00470 00−17810 00−04538 02−10136 03−06712 00−01303 02−19312 01−06342 03−13278 02−10286 00−07875 00−15420 02−14474 00−06102 02−01836 02−12066 00−08131 01−04226 00−11004 00−02822 00−16300 371 01−09448 315 02−11423 00−05934 02−06511 03−15268 00−16338 00−06488 02−23284 02−09742 03−05546 02−10234 02−12138 01−14451 00−05326 02−23819 03−03810 01−14670 00−19403 00−12517 02−22709 00−12961 00−19971 01−27750 00−13569 01−08166 00−22392 01−09324 01−06864 02−05601 329 02−15194 03−01669 03−06174 301 01−17447 03−04980 02−09146 01−08962 00−15760 00−11298 01−24404 01−08438 02−20873 01−16932 364 01−16602 00−13266 03−17163 03−04906 03−13943 00−18307 00−20915 00−22603 03−06895 02−15155 02−05920 01−13787 02−12189 309 341
ERG
ETV1
FOXA1
AR
Upper panel
ΔCt relative to gene median
−3 −2 −1 0 1 2 3
N/A
Lower panel
ΔCt relative to gene median
−3 −2 −1 0 1 2 3


ChIP-seq
DHT induction
AR knockdown
ERG knockdown
ERG association
positive
negative
no association
ChIP-seq
binding
no binding
Log2 Fold Change
−4
0
4
N/A
TPCAT−11−23310
TPCAT−1−20581
TPCAT−9−85143
TPCAT−8−75487
TPCAT−3−174133
TPCAT−18−31849
TPCAT−6−165200
TPCAT−15−21970v1
TPCAT−7−77105
TPCAT−11−23966v1
TPCAT−13−104777v1
TPCAT−15−23972
TPCAT−3−193565
TPCAT−10−3328v1
TPCAT−3−128149
TPCAT−10−84917
TPCAT−4−38191
TPCAT−1−240821v1
TPCAT−2−181218
EPCART
TPCAT−9−76124
TPCAT−7−148137
TPCAT−15−21907
PCAT5
TPCAT−7−25878
TPCAT−4−188294
TPCAT−16−65847
TPCAT−22−23869
TPCAT−10−29397v2
TPCAT−8−67111
TPCAT−2−13448
TPCAT−1−203391v1
TPCAT−4−156365
TPCAT−1−184630v1
AR
FOXA1
HOXB13
ERG
LNCaP
DuCaP
VCaP
LNCaP
DuCaP
VCaP
DuCaP
VCaP


A
HOXB13
ERG
AR
FOXA1
5
4
5
2
0
11
1
7
1
15
5
0
1
4
0
B
HOXB13
ERG
AR
FOXA1
20353
27899
13691
4301
169
23106
1090
12294
116
9567
11305
283
1478
4896
1764

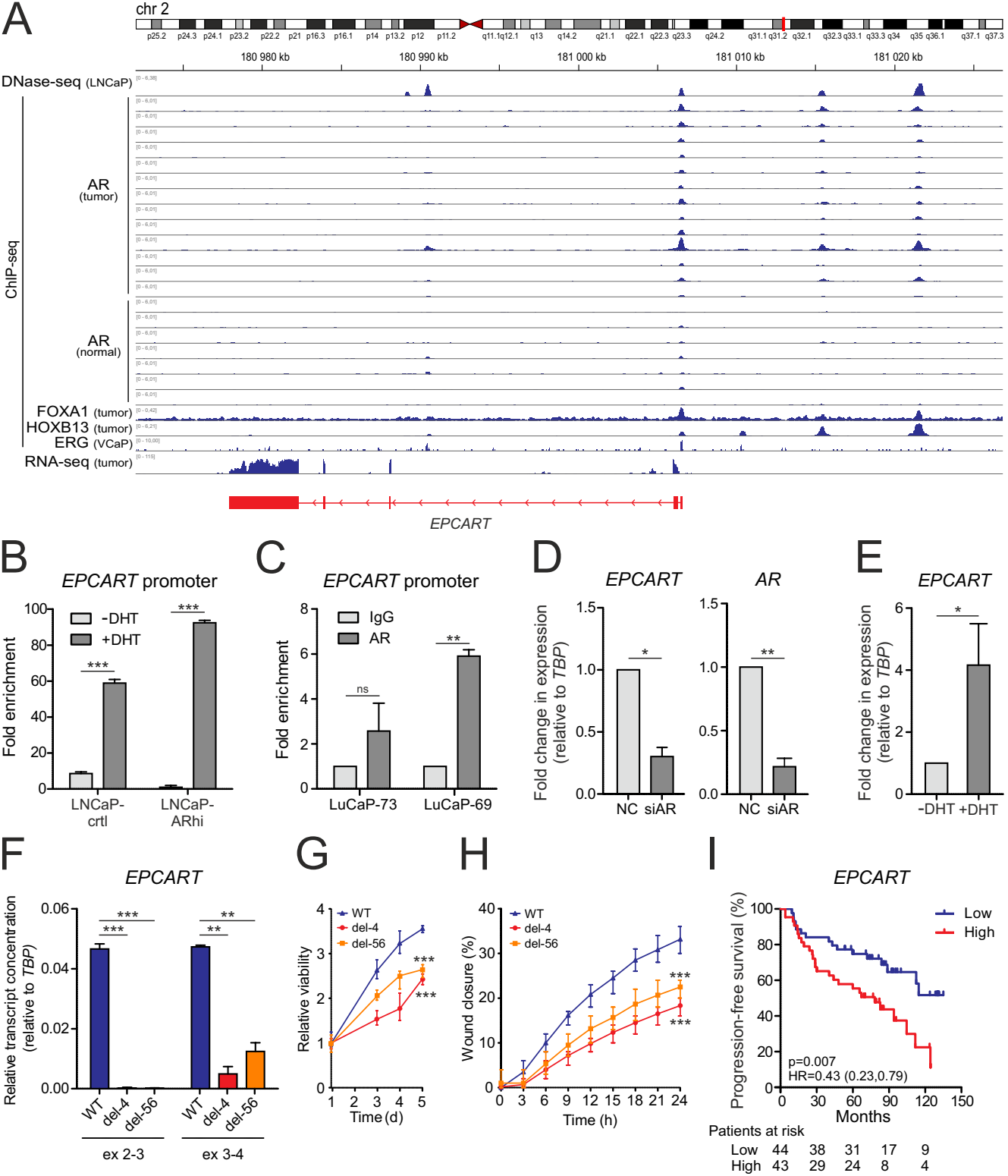

A
chr 2
180 980 kb
180 990 kb
181 000 kb
181 010 kb
181 020 kb
DNase-seq (LNCaP)
ChIP-seq
AR (tumor)
AR (normal)
FOXA1 (tumor)
HOXB13 (tumor)
ERG (VCaP)
RNA-seq (tumor)
EPCART
B
EPCART promoter
Fold enrichment
-DHT
+DHT
LNCaP-crtl
LNCaP-ARhi
C
EPCART promoter
IgG
AR
ns
LuCaP-73
LuCaP-69
D
EPCART
AR
Fold change in expression (relative to TBP)
NC siAR
E
EPCART
-DHT +DHT
F
EPCART
Relative transcript concentration (relative to TBP)
WT
del-4
del-56
ex 2-3
ex 3-4
G
Relative viability
Time (d)
H
Wound closure (%)
Time (h)
I
EPCART
Progression-free survival (%)
Low
High
p=0.007
HR=0.43 (0.23,0.79)
Months
Patients at risk
Low 44 38 31 17 9
High 43 29 24 8 4

**Table 1.** Multivariate Cox regression analysis.

| Variable | P-value | HR (95% CI) |
|---|---|---|
| *EPCART* | 0.027 | 2.06 (1.09-3.9) |
| Age at diagnosis | 0.3544 | 1.03 (0.97-1.10) |
| PSA at diagnosis | 0.0009 | 2.38 (1.43-3.97) |
| Gleason Score | 0.0023 | 2.16 (1.32-3.55) |
| pT | 0.001 | 3.10 (1.58-6.09) |

HR, hazard ratio
pT, pathological T stage