Tampere University

ANNIKA KOHVAKKA

Characterization of Novel Long Noncoding RNAs in Prostate Cancer

Tampere University Dissertations 1026

Tampere University Dissertations 1026

ANNIKA KOHVAKKA

Characterization of Novel Long Noncoding RNAs in Prostate Cancer

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the F114
of the Arvo, Arvo Ylpön katu 34, Tampere,
on 24 May 2024, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
Finland

Responsible supervisor and Custos	Professor Tapio Visakorpi Tampere University Finland	
Supervisors	Adjunct Professor Leena Latonen University of Eastern Finland Finland	Doctor Kati Kivinummi Tampere University Finland
Pre-examiners	Adjunct Professor Maria Sundvall University of Turku Finland	Adjunct Professor Biswajyoti Sahu University of Oslo Norway University of Helsinki Finland
Opponent	Professor Guido Jenster Erasmus MC The Netherlands	

Cover design: Roihu Inc.

ISBN 978-952-03-3448-2 (print)
ISBN 978-952-03-3449-9 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-3449-9

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2024

"Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes, yet they are usually left unchronicled."

– William Ramsay

ACKNOWLEDGEMENTS

The research presented in this dissertation has been conducted during 2013–2024 in the Molecular Biology of Prostate Cancer research group at Tampere University in Faculty of Medicine and Health Technology, and at former University of Tampere in Faculty of Medicine and Life Sciences, BioMediTech, and Institute of Biotechnology. The research has been supported by Cancer Foundation Finland, the Finnish Cultural Foundation, Research Council of Finland, Sidrig Juselius Foundation, Business Finland, Tampereen Yliopiston Tukisäätiö foundation, and Doctoral Program of Tampere University.

First and foremost, I want to thank my supervisor Prof. Tapio Visakorpi for his guidance and leadership over the years and giving me freedom and trust to do research in my own way. The knowledge and expertise you have in prostate cancer research has no rival and I have learned so much about this field thanks to you. I also want to thank Dr. Kati Kivinummi for her supervision in the beginning of my doctoral journey. Thank you for seeing the potential in me to be included part of your fantastic lncRNA project and helping be to adjust to the cancer field. Your help in and outside the lab gave me integral building blocks to be the researcher I am today. Last but not least, I want to thank Adj. Prof. Leena Latonen for her invaluable supervision at the later stage of this never-ending story that finally, thanks to her, got an end. You have been my mentor, my critic and my supporter, my role model, my rock, my therapist, my friend. Your integrity and deep passion for research has been an example also for my own career. It has been a privilege and an absolute pleasure to know you and to have you as my supervisor.

As no tumor, neither is cancer research created in a void. I would like to thank all the co-authors that collaborated to the publications included in this thesis. Most importantly, I want to thank Prof. Matti Nykter and his current and previous team members, especially Antti Ylipää, Matti Annala, and Anastasia Shcherban for their computational wizardry and help over the years. Collaboration with Matti and his team has been invaluable for my research personally and for our whole research group. I also want to thank Dr. Mina Sattari for her computational analyses after she joined our group. It has been great to have another lncRNA person in the group besides (and beside) me. I am also grateful for the collaboration with Adj. Prof. Peter

Sarin and his group member Pavlína Gregorová for the polysome profiling experiments, and with Adj. Prof. Ulla Aapola and Dr. Janika Nättinen for mass spectrometry experiments and analyses. I would also like to thank Prof. Teuvo Tammela for his contribution, as without him, we would not have the clinical prostate cancer cohort that was integral for our studies. My deepest gratitude goes to my steering group members, Prof Matti Nykter and Prof. Jorma Palvimo, who have guided and supported me over the years. Matti, I am truly grateful for your help. I really admire the way you lead your group and appreciate your efforts to create a more unified cancer research environment, and all the joined after-work events that have been organized for the endeavor. Jorma, thank you for your excellent advice and opinions. I always felt more motivated after our steering group meetings and a big part of that was because of you.

If it takes a village to raise a child, I feel it takes a research group to raise a researcher. I have been so fortunate to work with amazing colleagues, many of which I can now call my friends. I want to thank all the past and present people working in our research group, including our current group members Dr. Hanna Rauhala, Dr. Mina Sattari, Konsta Kukkonen, and Kirsi Kaukoniemi, and previous group members Dr. Heini Kallio, Dr. Sini Eerola, Dr. Mauro Scaravilli, Dr. Merja Helenius, Dr. Benedikta Hafliðadóttir, Dr. Outi Saramäki, Liisa Sjöblom, Katri Leinonen, and Anniina Brofeldt, not forgetting our current, previous, and shared lab assistants Päivi Martikainen, Hanna Selin, Riina Kylätie, Marika Vähä-Jaakkola, and Paula Kosonen. You have helped me in so many ways during this journey, and most importantly, created a warm and supporting place to work in. I also want to thank all other people in our research community that have helped me over the years, especially Sari Toivola for her histological knowledge and work that greatly helped my studies, and Assoc. Prof. Pekka Ruusuvuori for all his attempts at helping me and his constant questions about my progress, which greatly motivated me to finalize the thesis.

Finally, I want to thank my family and friends for their support and loyalty. Sini and Heini, all the words in the world cannot describe how much our friendship mean to me. Together, we have conquered the world and supported each other through all the good and bad, and I am so, so grateful for that. I wish our bond won't diminish in the future, even now that the life will lead us in different ways. I also want to thank Sini, Heini, Mina, Gunilla, and Benny for all the after-work escapes we have had, Ismaïl and Dafne for our long discussions and delicious gatherings, Lissu and Anniina for continued friendship after your departure from the prostate team, and Matti for our countless trips and geocaching adventures. Big thanks go also to my friends from Jyväskylä times, Sanna, Eija, and Sakke, for all the times you have

hosted me. Pauliina, you are my oldest friend, and I'm grateful that we have stayed close over the years and honored to be a Godmother (in the loosest of terms) to Konsta. Thank you also to all the scouts in Metsän Siskot, especially Ultzi, Teeku, Alica, Noora, Paula, Kerttu, Jenni, and Jonna. The experiences and fun we have had, especially during our nightly last-minute planning sessions in the camps, have truly kept me sane during these long years. Haluan kiittää myös äitiäni Pirjoa, isääni Keijoa ja siskoani Hannaleenaa kaikesta tuesta ja rakkaudesta vuosien varrella. Äiti, kiitos että olet juuri sellainen kuin olet ja opettanut minulle, kuinka olla hyvä ihminen. Isä, olet näyttänyt esimerkilläsi, että ahkeruus, tiedonjano ja halu ymmärtää asioita ovat tutkinnosta riippumattomia ominaisuuksia. Hanna, olet opettanut minulle, että joskus on ihan terveellistä ottaa vähän rennommin. Tukesi, keskustelumme ja näyttämäsi esimerkki nuoruudestani lähtien ovat muovanneet minusta paremman ihmisen. Kiitos!

Annika Kohvakka

Annika Kohvakka

January 2023, Tampere

ABSTRACT

Cancer is the second leading cause of death globally, and the number of diagnosed cancer cases continues to grow, thus exerting substantial physical, emotional, and financial burdens on society as a whole. In men, prostate cancer (PC) is the most diagnosed cancer and the third leading cause of cancer death in Europe and the second most lethal cancer in Finland. While the majority of patients with PC are cured, 11–26% of PCs clinically relapse and eventually progress to lethal PC. Current diagnostic methods, most notably prostate-specific antigen (PSA), are very efficient at detecting PCs to the point where even nonthreatening, indolent PCs are diagnosed and treated. However, these treatments can cause unnecessary harm to patients and burden the health care system. Therefore, more specific and effective ways to diagnose and treat aggressive PCs are needed.

Long noncoding RNAs (lncRNAs) are a relatively recently identified group of RNAs that do not encode proteins. They participate in the regulation of genes in cells, including in cancers. LncRNAs are typically expressed specifically in certain tissues and cellular contexts, and many of them are aberrantly expressed in cancers. In addition, several lncRNAs have been shown to play roles in cancer development and progression. These attributes make lncRNAs ideal as prospective biomarkers and therapeutic targets in cancers.

The aim of this dissertation was to discover novel, unexplored lncRNAs that are specifically expressed in PC and to study their regulatory, functional, and biomarker potential in PC models. For this purpose, multiple next-generation sequencing datasets, both publicly available and produced by us, and several molecular, cellular, and biotechnological methods were used to evaluate both *in vitro* and clinical patient material. These methods were utilized to study the expression profiles of novel lncRNAs in different sample types and to understand the detailed molecular mechanisms underlying their aberrant expression and function. In total, we identified more than one hundred novel PC-associated lncRNAs, called Tampere PCATs (TPCATs). The aberrant expression of most TPCATs is caused by the dysregulation of PC-specific transcription factors, most notably androgen receptors. High expression levels of three TPCATs were independently associated with PC progression. The prognostic TPCATs include *EPCART*, which is androgen

regulated and associated with *TMPRSS2-ERG* fusion, the most common genetic alteration in PC. The functional role of *EPCART* was studied in more detail in PC cells, and *EPCART* was found to promote PC cell proliferation and migration through the regulation of protein translation via the PI3K/AKT/mTORC1 pathway. Overall, the findings in this dissertation revealed novel diagnostic and prognostic biomarker candidates for PC and identified new molecular pathways and potential therapeutic targets for PC treatment.

TIIVISTELMÄ

Syöpä on maailman toiseksi yleisin kuolinsyy ja uusien syöpätapausten määrä on kasvanut jatkuvasti, mikä aiheuttaa merkittävää fyysistä, emotionaalista ja taloudellista taakkaa koko yhteiskunnalle. Miehillä eturauhassyöpä on Euroopan diagnosoiduin syöpä ja kolmanneksi yleisin syöpäkuoleman syy; Suomessa eturauhassyöpä aiheuttaa toiseksi eniten syöpäkuolleisuutta miehillä. Vaikka suurin osa eturauhassyöpäpotilaista paranee, 11–26 % eturauhassyövistä uusiutuu ja lopulta etenee tappavaan eturauhassyövän muotoon. Nykyiset diagnostiset menetelmät, etenkin prostataspesifinen antigeeni (PSA), ovat erittäin tehokkaita tunnistamaan eturauhassyöpiä, mikä on johtanut myös vaarattomien, vähäoireisten syöpätapausten diagnosoimiseen ja hoitamiseen. Tämä tuottaa tarpeetonta fyysistä ja henkistä kuormaa potilaille ja lisäksi rasittaa terveydenhuoltojärjestelmää. Siksi tarkemmille ja tehokkaammille menetelmille diagnosoida ja hoitaa aggressiivisia eturauhassyöpiä on suuri tarve.

Pitkät ei-koodaavat RNA:t (lncRNA) ovat suhteellisen hiljattain löydetty ryhmä RNA:ita, joista ei tuoteta proteiineja. Ne toimivat osana solujen geenisäätelyä, myös syövissä. LncRNA:t ilmentyvät erityisesti spesifeissä kudoksissa ja solukonteksteissa, ja monien lncRNA:iden on havaittu ilmentyvän syövissä normaalista poikkeavasti. Lisäksi useilla lncRNA:illa on osoitettu olevan rooli syövän kehittymisessä ja leviämisessä. Nämä ominaisuudet tekevät lncRNA:ista erinomaisia kandidaatteja biomarkkereiksi ja mahdollisiksi terapeuttisiksi kohteiksi syövissä.

Tämän väitöskirjan tavoitteena oli löytää aiemmin tuntemattomia lncRNA:ita, jotka ilmentyvät spesifisti eturauhassyövässä, ja selvittää näiden lncRNA:iden säätelyä, toimintaa ja biomarkkeripotentiaalia eturauhassyöpämalleissa. Tätä varten hyödynsimme julkisesti saatavilla olevia ja itse tuotettuja uuden sukupolven sekvensointimenetelmillä valmistettuja aineistoja ja monia solu- ja molekyylibiologian sekä bioteknologian tutkimusmenetelmiä. Näitä menetelmiä käyttämällä tutkimme uusien lncRNA:iden ilmentymistä erilaisissa solumalleissa ja kliinisissä potilasnäytteissä, sekä näiden lncRNA:iden normaalista poikkeavaa ilmentymistä ja toiminnan taustalla olevia molekyylimekanismeja. Yhteensä löysimme yli sata uutta eturauhassyöpään assosioituvaa lncRNA:ta, joita kutsuimme TPCAT:eiksi. Normaalista poikkeavan ilmentymisen havaittiin johtuvan isolla osalla

TPCAT:eista eturauhassyöpäspesifisten transkriptiotekijöiden, etenkin androgeenireseptorin, poikkeavasta säätelystä. Havaitsimme, että kolmen TPCAT:n korkea ilmentyminen assosioi myös eturauhassyövän progression kanssa. Prognostisten TPCAT:ien joukosta erottautui *EPCART*, joka osoittautui myös androgeenien säätelemäksi ja assosioituvan eturauhassyövän yleisimmän geneettisen muutoksen, *TMPRSS2-ERG*-fuusion kanssa. *EPCART*:n toiminnallista roolia tutkittiin tarkemmin eturauhassyöpäsoluissa, joissa sen todettiin edistävän solujen jakautumista ja liikkumista. Tämän havaittiin ainakin osittain tapahtuvan proteiinien translaation säätelyn kautta, missä *EPCART* aktivoi PI3K/AKT/mTORC1 signalointireittiä. Tässä väitöskirjassa löysimme uusia potentiaalisia biomarkkereita eturauhassyövän toteamiseen ja aggressiivisuuden ennustamiseen, sekä tunnistimme uusia molekyylireittejä ja mahdollisia terapeuttisia kohteita eturauhassyövän hoitoon.

CONTENTS

ABBREVIATIONS

4E-BP	eIF4E binding proteins
ADT	androgen deprivation therapy
AKT	protein kinase B
AR	androgen receptor
ARBS	androgen receptor binding site
ARE	androgen response element
ARSI	androgen receptor signaling inhibitors
ASO	antisense oligonucleotide
BCR	biochemical recurrence
BPH	benign prostatic hyperplasia
ChIP-seq	chromatin immunoprecipitation sequencing
Cq	quantification cycle
CRE	*cis*-regulatory element
CRPC	castration resistant prostate cancer
CT	computed tomography
CTCF	CCCTC-binding factor
DBD	DNA binding domain
ddPCR	droplet digital PCR
del-4 or del-56	*EPCART*-deletion clone 4 or 56
DEPTOR	DEP domain containing mTOR interacting protein
DHT	dihydrotestosterone
DRE	digital rectal exam
DSB	DNA double strand break
DTT	dithiothreitol
eIF	translation initiation factor
EMA	European Medicines Agency
ENCODE	Encyclopedia of DNA Elements
EPCART	ERG-positive prostate cancer-associated androgen responsive transcript
EPCART-del	*EPCART*-deletion

ERG	ETS transcription factor ERG
ETS	E26 transformation-specific
ETV	ETS variant transcription factor
EZH2	enhancer of zeste 2 polycomb repressive complex
FC	fold change
FDA	U.S. Food and Drug Administration
FFPE	formalin-fixed paraffin-embedded
FISH	fluorescence in situ hybridization
FOXA1	forkhead box protein A1
fPSA	free prostate-specific antigen
gDNA	genomic DNA
GDP	guanosine diphosphate
GEO	Gene Expression Omnibus
GnRH	gonadotropin-releasing hormone
GS	Gleason score
GTP	guanosine triphosphate
GWAS	genome-wide association studies
H&E	hematoxylin and eosin
HDAC	histone deacetylase
HOXB13	homeobox protein Hox-B13
HR	homologous recombination
HSP	heat shock proteins
IHC	immunohistochemistry
IPA	Ingenuity Pathway Analysis
ISUP	International Society of Urological Pathology
KO	knockout
LBD	ligand-binding domain
LH	luteinizing hormone
LHRH	luteinizing hormone-releasing hormone
lincRNA	long intergenic noncoding RNA
LNA	locked nucleic acid
lncRNA	long noncoding RNA
mCRPC	metastatic castration resistant prostate cancer
miRNA	micro-RNA
mLST8	mammalian lethal with SEC13 protein 8
MRI	magnetic resonance imaging

mRNA	messenger RNA
MS	mass spectrometry
mTOR	mechanistic target of rapamycin
mTORC	mTOR complex
NC	negative control
NCOA	nuclear receptor coactivator
ncRNA	noncoding RNA
NHEJ	non-homologous end joining
NKX3-1	NK3 homeobox 1
nt	nucleotide
NT	no template control
NTD	NH2-terminal domain
ORF	open reading frame
p70S6K	ribosomal protein S6 kinase beta-1
PAGE	polyacrylamide gel electrophoresis
PARP	poly (ADP-ribose) polymerase
PC	prostate cancer
PCAT	prostate cancer-associated transcript
PCAT5	prostate cancer associated transcript 5
PCR	polymerase chain reaction
PDCD4	phosphorylates programed cell death 4
PDK-1	phosphoinositide-dependent kinase-1
PET-CT	positron emission topography- computed tomography
PH	pleckstrin homology
PI3K	phosphoinositide 3-kinase
PIN	prostatic intraepithelial neoplasia
PIP_2	phosphatidylinositol 4,5-bisphosphate
PIP_3	phosphatidylinositol (3,4,5)-trisphosphate
PolII	RNA polymerase II
PRAS40	proline-rich AKT substrate of 40 kDa
PSA	prostate-specific antigen
PSMA	prostate-specific membrane antigen
PTEN	phosphatase and tensin homolog
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RACE	rapid amplification of cDNA ends
RAPTOR	regulatory protein associated with mTOR

RHEB	Ras homolog enriched in brain
RICTOR	rapamycin insensitive companion of mTOR
RNA-ISH	RNA in situ hybridization
RNA-seq	RNA sequencing
RTK	receptor tyrosine kinase
SD	standard deviation
SDS	sodium dodecyl sulfate
sgRNA	single guide RNA
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SWATH-MS	sequential window acquisition of all theoretical fragment ion spectra mass spectrometry
TAD	topologically associated domain
TCGA-PRAD	The Cancer Genome Atlas Prostate Adenocarcinoma
TE	transposable element
TF	transcription factor
TMA	tissue microarray
TNM	primary tumor/regional lymph node/distant metastasis
TOP2B	topoisomerase II beta
TPCAT	Tampere prostate cancer associated transcript
tPSA	total prostate-specific antigen
TRUS	transrectal ultrasound
TSC	tuberous sclerosis complex
TSS	transcription start site
TURP	transurethral resection of the prostate
UTR	untranslated region
WGS	whole-genome sequencing
WT	*EPCART* wild-type clone

PUBLICATIONS

This thesis is based on the following original communications, referred to in the text by Roman numerals:

I Antti Ylipää*, Kati Kivinummi*, **Annika Kohvakka**, Matti Annala, Leena Latonen, Mauro Scaravilli, Kimmo Kartasalo, Simo-Pekka Leppänen, Serdar Karakurt, Janne Seppälä, Olli Yli-Harja, Teuvo L. J. Tammela, Wei Zhang, Tapio Visakorpi, and Matti Nykter. Transcriptome Sequencing Reveals *PCAT5* as a Novel ERG-Regulated Long Noncoding RNA in Prostate Cancer. (2015). *Cancer Research*, 75:4026–4031. DOI: 10.1158/0008-5472.CAN-15-0217.

II **Annika Kohvakka**, Mina Sattari, Anastasia Shcherban, Matti Annala, Alfonso Urbanucci, Juha Kesseli, Teuvo L. J. Tammela, Kati Kivinummi, Leena Latonen, Matti Nykter, and Tapio Visakorpi. AR and ERG drive the expression of prostate cancer specific long noncoding RNAs. (2020). *Oncogene*, 39:5241–5251. DOI: 10.1038/s41388-020-1365-6.

III **Annika Kohvakka**, Mina Sattari, Janika Nättinen, Ulla Aapola, Pavlína Gregorová, Teuvo L.J. Tammela, Hannu Uusitalo, Peter Sarin, Tapio Visakorpi*, and Leena Latonen*. Long noncoding RNA *EPCART* regulates translation through PI3K/AKT/mTOR pathway and PDCD4 in prostate cancer. Manuscript.

*These authors contributed equally to the work.

AUTHOR'S CONTRIBUTION

Publication I: The author designed and performed laboratory work and data analysis for *PCAT5 in vitro* experiments to address the reviewers' comments. The author also participated in the writing of those parts of the manuscript.

Publication II: The author participated in the study design with the supervisors. The author designed and performed the majority of the laboratory work (all except the creation of CRISPR-Cas9 cell clones) and some of the data and statistical analyses. The author was responsible for writing the first draft of the manuscript and finalizing the manuscript for submission.

Publication III: The author participated in the study design with the supervisors. The author designed and performed the majority of the laboratory work and data analyses (all except RNA sequencing library construction, polysome profiling, mass spectrometry experiments, immunohistochemical staining, and analyses of those data) and some of the statistical analyses. The author was responsible for writing the first draft of the manuscript and finalizing the manuscript for submission.

1 INTRODUCTION

Cancer is a disease of uncontrollable cell growth and division in which cells evolve to spread outside their usual boundaries to invade other parts of the body. Cancer formation and development are caused by inherited and/or somatically gained genetic and epigenetic alterations that accumulate over time. These alterations lead to the dysregulation of genes that modulate how and when cells proliferate, grow, interact with their surroundings, and die. While the majority of the known genes that drive cancer development are protein-coding genes, noncoding RNA (ncRNA) genes, which do not encode proteins, have also been found to play a significant role in cancer formation and progression. Approximately 15 years ago, a new group of ncRNAs consisting of more than 200-nucleotide (nt) long ncRNAs (lncRNAs) was recognized. Although many of their functions are still unknown, a large number of lncRNAs have been shown to play important roles in the regulation of cancer cells.

Prostate cancer (PC) is the most commonly diagnosed cancer in European men (Dyba et al., 2021). While a great majority of patients with PC are cured, 11–26% of PCs clinically progress within 15 years after initial treatment or active monitoring (Hamdy et al., 2023). PCs, as well as normal prostates, are dependent on androgens for their cell growth and survival. Hence, androgen deprivation is utilized as a therapeutic option for recurrent PCs. Cancers that become resistant to androgen deprivation (*i.e.*, castration-resistant PC, CRPC) are eventually lethal. Therefore, new treatments for CRPCs are desperately needed.

Improvements in diagnostic methods, mainly due to the invention of prostate-specific antigen (PSA) determination, have vastly increased the number of diagnosed PCs over the past three decades (Duffy, 2020). However, PSA is not a PC-specific marker, as inflammation in the prostate and benign prostatic hyperplasia (BPH) can also lead to the overexpression of PSA (Nadler et al., 1995); thus, PSA can also be used to diagnose indolent tumors (Pinsky et al., 2014). This has led to the overdiagnosis and overtreatment of PCs, which can cause unnecessary adverse effects for patients and additional costs for society (Fenton et al., 2018). Moreover, current prognostic markers cannot perfectly predict which PCs will relapse and which will not. Therefore, more precise diagnostic and prognostic markers are required to improve the diagnosis and treatment of PCs.

In the studies described in this dissertation, novel PC-associated lncRNAs (PCATs) were discovered and characterized in PC models to better understand how, where and why these lncRNAs are expressed in PC. These findings reveal that PC-specific dysregulation is a driving mechanism for PCAT expression and reveal a novel androgen-regulated lncRNA that promotes PC proliferation and migration through the modulation of protein translation. Moreover, we examined whether these novel PCATs could be utilized in the detection or treatment of PC and identified potential candidates for biomarkers and novel therapeutic targets.

2 REVIEW OF THE LITERATURE

2.1 Prostate cancer

2.1.1 Epidemiology

Prostate cancer is the most commonly diagnosed cancer in men in Europe (Dyba et al., 2021), including in Finland, where there were over 5200 new cases diagnosed and approximately 60 000 patients suffering from PC in 2021 (Seppä et al., 2023). Patients with PC are typically over 60 years old, and the average age at diagnosis is 66 years globally (Bray et al., 2018; Rawla, 2019). Often, PC is asymptomatic and indolent, and the 5-year survival rate for PC is high, at 94% in Finland (Seppä et al., 2023). Nevertheless, PC is the third most common cause of cancer death among men in Europe (Dyba et al., 2021) and the second most common cause of cancer death in Finland, with over 900 PC deaths in 2021 (Seppä et al., 2023).

2.1.2 Prostate anatomy and histology

The prostate is an exocrine gland of the male reproductive system located at the base of the urinary bladder. Its primary function is to secrete fluid that transports and nourishes semen. The prostate is surrounded by an integral fibromuscular layer called the capsule. The prostate gland wraps around the urethra and two ejaculatory ducts (Figure 1A). Histologically, the prostate gland can be divided into three zones: transitional, central, and peripheral (Figure 1A) (Lee et al., 2011; McNeal, 1981). Additionally, the periurethral stroma and anterior fibromuscular stroma can be defined (Figure 1A) (Lee et al., 2011; McNeal, 1981). The peripheral zone is the outermost layer that surrounds the transition and central zones and accounts for 70% of the prostatic volume (Lee et al., 2011; McNeal, 1981). It also accounts for approximately 70% of PCs (Lee et al., 2011; McNeal et al., 1988). The central zone surrounds the ejaculatory ducts and accounts for 25% of the prostatic volume (Lee et al., 2011; McNeal, 1981). The transitional zone wraps around the proximal urethra

and accounts for 5% of the prostatic volume (Lee et al., 2011; McNeal, 1981). It is the most common place of origin for BPH (McNeal, 1981).

At the cellular level, the prostate gland is composed of acini and ducts lined by pseudostratified epithelium upon the basal membrane and surrounded by stroma (Ittmann, 2018). There are three major types of epithelial cells observed in the prostate: luminal, basal, and neuroendocrine, of which luminal and basal cells are the most common (Figure 1B) (Abate-Shen & Shen, 2000). Recent single-cell RNA sequencing studies have revealed two additional epithelial cell types, club cells and hillock cells (Henry et al., 2018), the roles of which in the prostate have not yet been fully characterized. The majority of luminal cells are terminally differentiated and known for their high expression levels of androgen receptor (AR) and PSA (Abate-Shen & Shen, 2000; Ittmann, 2018). Luminal cells are also secretory cells that can express and secrete substances that liquefy the ejaculate, *e.g.,* PSA, into the lumen of the glands (Abate-Shen & Shen, 2000; Ittmann, 2018). In contrast, basal cells are proliferating cells that express low levels of PSA and AR. It is generally believed that basal cells possess the potential to differentiate into luminal and even neuroendocrine cells (Xin, 2019). Neuroendocrine cells comprise only up to 1% of the epithelial cells in the prostate and are scattered among the luminal and basal cells. They are suspected to perform both secretory and proliferative activities (Xin, 2019). The stroma consists of stromal cells, including fibroblasts, smooth muscle cells, endothelial cells, and immune cells, in addition to blood vessels, automatic nerve fibers, and extracellular matrix components (Barron et al., 2012). Stromal cells participate in the formation and remodeling of the extracellular matrix and the regulation of epithelial cell development and death and are also an essential part of the immune response in the prostate (Barron et al., 2012; Hägglöf & Bergh, 2012).

The vast majority of PCs are adenocarcinomas, meaning that they develop from the glandular epithelium. Prostate cancer has been suggested to arise from either luminal or basal cells (Xin, 2019). A small number of neuroendocrine cells can also be found in PC tumors (Huang et al., 2007). In addition, stromal cells can assist cancer cells by creating a suitable microenvironment for proliferation, differentiation, and movement (Barron et al., 2012).

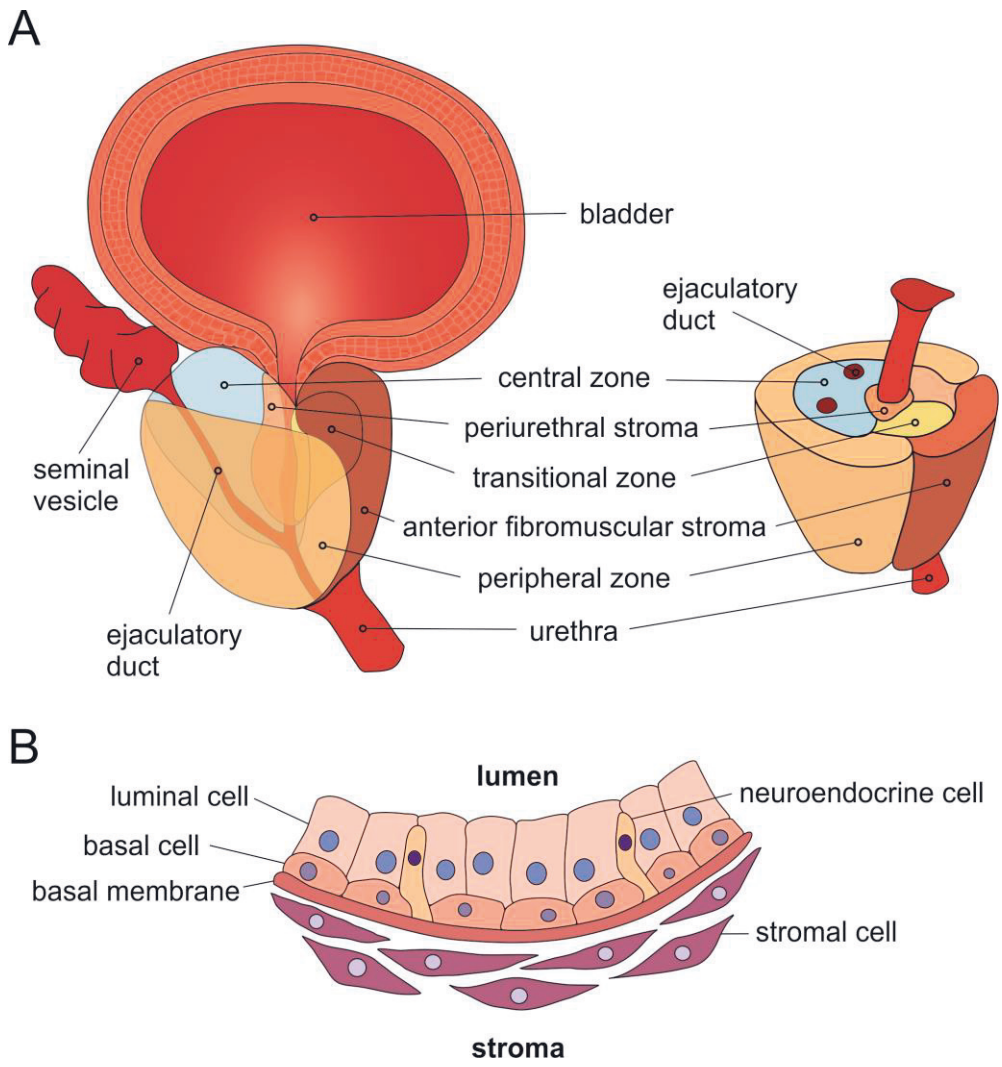


Figure 1. Prostate anatomy. A) Zonal anatomy of the prostate. B) Cellular architecture of the prostate epithelium. The figures are based on illustrations in Verze et al. (Verze et al., 2016).

2.1.3 Diagnostics, staging, and prognostics

The initial steps used for the diagnosis of PC have traditionally included digital rectal examination (DRE) for the detection of any abnormalities (*e.g.*, a hard mass or nodule, induration, or lobar asymmetry) on the prostate surface that might indicate PC and PSA level measurements in serum (EAU, 2023). High PSA levels can be an

indicator of PC, although BPH and prostatitis can also increase PSA levels (Nadler et al., 1995). When DRE or PSA tests indicate PC, PC diagnosis is confirmed by a core needle biopsy (EAU, 2023). Traditionally, 12 core samples are taken from different prostate locations via transrectal ultrasound (TRUS) or, more recently, magnetic resonance imaging (MRI) (EAU, 2023). The biopsies are analyzed by histopathological methods for cancer grading (Gleason grading), in which abnormalities in the cells and tissue structures are assessed (Gleason, 1966; Van Leenders et al., 2020). After PC diagnosis is confirmed, the extent of cancer spread can be further evaluated. As PC often metastasizes to bone, a bone scan by isotope labeling is typically performed on symptomatic patients (EAU, 2023). In addition, other imaging methods to evaluate the location and spread of cancer can be used, *e.g.,* MRI, computed tomography (CT) scan, or positron emission topography-CT (PET-CT) scan (EAU, 2023). All the test results from the different analyses are combined to assess the stage and prognosis of the cancer, which are then used to determine the method of treatment (EAU, 2023).

Although the current methods of diagnosis and prognosis are sufficient for detecting most PCs and predicting their aggressiveness, they have drawbacks. For example, PSA testing can lead to the detection of patients with indolent PC who need not be treated, *i.e.*, overdiagnosis, causing unnecessary stress for patients and a burden on health care (Donovan et al., 2016; Fenton et al., 2018). In addition, the current prognostic methods do not include all the significant risk factors that influence PC progression and could be utilized for better patient care. Thus, new biomarkers are needed for improved and more personalized diagnostics, prognostics, and treatment selection of PC.

Prostate-specific antigen

PSA is a serine protease that is part of the kallikrein-related peptidase family (Lilja, 1985) and is encoded by the *KLK3* gene located on chromosome 19q13.33. In a normal prostate gland, PSA is secreted by the ductal and acinar epithelium into the seminal fluid, where it functions as a liquefier of the seminal coagulum to allow the sperm to move freely.

PSA is highly prostate-specific (Wang et al., 1979) and can be detected in the serum (Kuriyama et al., 1980). PSA levels in the serum of healthy men are low, and high PSA serum levels are associated with PC (Kuriyama et al., 1980). Measuring total PSA (tPSA) levels in serum samples is the most commonly used diagnostic tool for PC. However, PC is not the only condition that can increase tPSA levels in serum,

as benign conditions such as prostatitis and BPH can also cause tPSA levels to increase (Nadler et al., 1995). In addition, age increases the likelihood of BPH, which causes the prostate volume to increase and causes changes in tPSA levels. To improve PC detection in younger men, age-adjusted guidelines for tPSA levels (Oesterling et al., 1995) are often used. To further enhance the distinction between PC and benign prostatic disease, the ratio of free PSA (fPSA) to tPSA may be measured, especially when tPSA levels are modest (<10 μg/l). The majority of serum PSA is bound to serum protease inhibitors, most often to alpha-1-anti-chymotrypsin (PSA-AC), in contrast to fPSA, which is noncomplexed in the serum (Lilja et al., 1991; Stenman et al., 1991). Higher tPSA and lower fPSA levels are associated with a greater risk of PC; it is estimated that 30–50% of men with an fPSA/tPSA ratio <15% will have cancer at biopsy (Catalona et al., 1998).

Since the approval of the serum PSA test for the diagnosis of PC in the clinic (Catalona et al., 1994), its use has significantly enhanced the detection of patients with PC due to its ability to recognize asymptomatic PCs early (Duffy, 2020). However, as PSA is not cancer-specific and its levels can increase for other reasons, PSA screening of asymptomatic men may lead to unnecessary biopsies, which can result in avoidable complications (Pinsky et al., 2014). In addition, through PSA screening, it is possible to recognize indolent PCs that would not cause clinical consequences for patients in their lifetime, which has led to the overdiagnosis and overtreatment of PC. It has been estimated that 16–41% of all patients with PC are overdiagnosed (Fenton et al., 2018) and undergo treatments such as radical prostatectomy or radiation therapy, which can lead to complications such as impotence and incontinence (Donovan et al., 2016).

In addition to the use of the PSA test as a diagnostic tool, it can also be used as a prognostic marker to assess the risk of PC recurrence, as a follow-up tool to monitor the recurrence of PC after treatment for the primary disease, and as a tool to monitor the response to treatment for advanced disease. The use of PSA as a prognostic factor relies on the discovery that there is a near linear correlation between PSA levels at initial diagnosis and outcome (Duffy, 2020). Generally, PSA levels of 10 to <20 μg/l are regarded as a marker for intermediate risk and levels ≥20 μg/l as a marker for high-risk recurrence (EAU, 2023). However, some patients with low PSA levels can have poor outcomes, which is why other clinical and histopathological factors are used to predict outcomes together with PSA (EAU, 2023).

After successful treatment of primary PC, PSA levels decline drastically, for example, in the case of radical prostatectomy, to undetectable levels (<0.1 μg/l)

within two months (Loblaw et al., 2017). An increase in the PSA concentration after initial treatment can be a sign of recurrence, *e.g.,* after radical prostatectomy, biochemical recurrence (BCR) is defined as two consecutive increases in the PSA concentration >0.2 μg/l (Amling et al., 2001). However, BCR does not equal clinical recurrence. For example, only approximately 30% of patients with BRC develop metastatic PC within 15 years following surgery (Pound et al., 1999). Additionally, due to the easy availability of PSA tests, they can be used to monitor the response to therapies in patients with advanced PC (EAU, 2023). Although PSA determination has been shown to be useful in predicting the responses to hormone therapy and most other treatments (Hussain et al., 2006), imaging is routinely used to follow responses together with PSA.

Gleason grading

Cancer histology is performed to assess abnormalities in tissue morphology and cells. The most commonly used grading system for PC is Gleason grading, in which microscopic examination of hematoxylin and eosin (H&E)-stained histological tissue samples is used to determine the architectural growth patterns of prostate adenocarcinomas.

Gleason grading is based on the recognition of five possible Gleason patterns (Table 1), with Gleason 1 being the most differentiated and Gleason 5 being the least differentiated glandular structures (Gleason, 1966; Van Leenders et al., 2020). As many prostate adenocarcinomas harbor more than one Gleason pattern, the sum of the primary and secondary patterns is calculated, giving the Gleason score (GS). The primary pattern is defined as the most prevalent pattern detected, and the secondary pattern may be either the second most prevalent pattern (in prostatectomies) or any identifiable high-grade pattern of any quantity (in systematic biopsies). If only one Gleason pattern is detected, the secondary pattern is considered the same as the primary pattern.

The GS can be used as a prognostic marker to predict the outcome of PC (Table 1); the higher the score is, the poorer the prognosis (Humphrey, 2004). In 2014, the International Society of Urological Pathology (ISUP) adopted a new prognostic grading system consisting of grade groups 1 to 5 (Pierorazio et al., 2013). The new grading system is based on Gleason grading, but it is simplified to include only those grades that accurately reflect prognosis (Table 1).

Although Gleason grading can predict PC outcome quite well, especially for lower and higher grades, it is not absolute. Gleason grading relies strongly on the

expertise of pathologists and is subject to interpretation. In addition, when based on GS only, treatment selection can be difficult, particularly for tumors with intermediate GSs. Thus, Gleason grading is often combined with other prognostic factors (EAU, 2023).

Table 1. Gleason grades and Grade groups

Gleason pattern	Description	Gleason score	Grade group	Probability of progression (%)*
1	Small, uniform glands	≤6	1	4
2	More stroma between glands			
3	Distinct infiltration of cells from glands at margins			
4	Irregular masses of neoplastic cells with few glands	3 + 4 = 7	2	12
		4 + 3 = 7	3	37
		4 + 4 = 8	4	52
5	Lack of or occasional glands; sheets of cells	3 + 5 = 8		
		5 + 3 = 8		
		4 + 5 = 9	5	74
		5 + 4 = 9		
		5 + 5 = 10		

*Calculations based on 5-year BCR-free progression probability for radical prostatectomy (Epstein *et al.*, 2016).

TNM staging and risk groups

TNM staging is a standardized system used for classifying the anatomical extent of cancers (Brierley et al., 2017). The system describes the size of the primary tumor (T) and whether it has invaded nearby tissue, the extent to which the cancer has spread to regional lymph nodes (N), and the presence of distant metastasis (M).

In PC, before initial treatment, clinical TNM staging is carried out by DRE and sometimes TRUS and/or other imaging techniques. In clinical staging, the T status for PC can be categorized into four stages: T1, clinically inapparent tumor not palpable or visible by imaging; T2, tumor confined within prostate; T3, tumor extending through the prostatic capsule; and T4, tumor fixed or invading adjacent structures other than seminal vesicles (Brierley et al., 2017). Additionally, pathologic staging is determined after radical prostatectomy where the prostate and sometimes

the regional lymph nodes are dissected but can also be determined by other means, *e.g.,* by a biopsy of the extraprostatic soft tissue. For pathologic staging, the extracted prostatic tissue and/or the surrounding tissue are processed, stained with H&E, and analyzed by pathologists. This leads to slightly modified staging categories compared to clinical staging (Brierley et al., 2017).

By itself, TNM staging in PC has been shown to have prognostic value (Epstein, 2011). However, the TNM stage combined with the GS grade and diagnostic PSA has been found to lead to a more reliable prognosis (D'Amico et al., 1998; Roach et al., 2006). Various cancer and urological organizations have developed risk group classification systems that consider all three of these factors. These risk groups are generally used when choosing the best therapy for the patient. The classification system implemented by the European Association of Urology (Table 2) is consistent with that of D'Amico (D'Amico et al., 1998; EAU, 2023).

Table 2. European Association of Urology risk groups for the biochemical recurrence of localized and locally advanced PC

Risk group	PSA		GS		TNM	
Low	<10 ng/ml	AND	<7	AND	T1 - T2a	
Intermediate	10 - 20 ng/ml	OR	7	OR	T2b	**Localized**
High	>20 ng/ml	OR	>7	OR	T2c	
	any		any		T3-4 or N+	**Locally advanced**

Survival analysis

In clinical cancer research, understanding the factors influencing the survival time of patients after diagnosis or treatment is crucial for determining patient prognosis, evaluating treatment effectiveness, and estimating survival rates. Hence, survival analysis is employed to analyze the time until an event of interest occurs. The events of interest are often recurrence or relapse-free survival, response to treatment, or death. In PC research, BCR-free survival (also known as biochemical progression-free survival), metastasis-free survival, PC-specific survival, and overall survival are commonly used. BRC-free survival measures the time until PSA levels increase after treatment, metastasis-free survival evaluates the time until cancer spreads beyond the prostate, PC-specific survival tracks survival specifically due to prostate cancer, and overall survival reflects the time until death from any cause. Survival analysis also takes into account the censoring of data, which occurs when individuals are lost to

follow-up or when the event of interest has not occurred by the end of the study period (Clark et al., 2003; EAU, 2023).

Three main methods commonly used in cancer survival analysis are Kaplan–Meier analysis, the log-rank test, and Cox regression analysis. Kaplan–Meier analysis is used to estimate the survival probability over time for different groups of individuals, such as patients with different genetic backgrounds or treatment regimens. This analysis generates a survival curve, which illustrates the probability of survival as a function of time. It is a nonparametric method and can effectively handle censored data. The log-rank test is a statistical hypothesis test used to compare the survival curves of two or more groups. It determines whether there is a significant difference in survival between the groups while accounting for censoring. The test assesses whether the observed differences in survival curves are greater than would be expected by chance alone. Cox regression analysis, also known as the proportional hazards model, is a multivariable regression technique used to assess the association between one or more predictor variables and survival outcomes while adjusting for other covariates. It estimates hazard ratios, which quantify the relative risk of experiencing the event of interest between groups or levels of a predictor variable. Cox regression is widely used in cancer research to identify independent prognostic factors that influence patient survival. For instance, researchers may use Cox regression to determine whether age, diagnostic PSA level, Gleason score, TNM stage, and genetic alteration or expression of a target gene independently predict survival in a cohort of patients with PC (Bradburn et al., 2003; Clark et al., 2003).

2.1.4 Treatment

Treatment selection for PC depends not only on the spread and prognosis of the cancer but also on other factors, such as the life expectancy of the patient, the wish of the patient, or treatment availability. In the treatment selection for tumors that are confined to the prostate, also called localized PCs, much emphasis is placed on the risk group of the cancer (Table 2). PCs that are small, have not extended outside of the prostate gland and have a low-risk prognosis usually grow slowly and are asymptomatic. For these patients, treatment is not necessarily needed, and observation (also called watchful waiting) or active surveillance is an option. This option is also suggested for all patients with PC who are older and/or have other serious health problems, as they are more likely to die from causes other than PC.

During active surveillance, PSA tests and DRE exams are taken regularly, and additional needle biopsies and imaging tests may also be performed to determine if the cancer has progressed and if curative treatment is needed. Watchful waiting is less intensive and relies more on the patients' symptoms that might occur and on treating the symptoms rather than curing the disease. Alternatively, low-risk cancers can be treated. The most common treatment options for localized PCs are radical prostatectomy and radiation therapy. For intermediate- and high-risk patients, these treatment options are preferred over active surveillance. For high-risk patients, additional treatments might be included, *e.g.,* hormonal therapy in addition to radiation therapy or combining radical prostatectomy with radiation therapy (EAU, 2023).

Radical prostatectomy is a surgical procedure for removing the whole prostate gland and the attached seminal vesicles. Frequently, regional lymph nodes are also removed, especially for cancer staging purposes. Radical prostatectomy may be carried out as an open surgery in which the prostate is removed through an incision in the lower abdomen. Alternatively, a laparoscopic, robotically assisted radical prostatectomy may be performed, in which small incisions are made in the abdomen into which the robot's arms are inserted. As the recovery time is shorter in laparoscopic procedures than in open surgery, laparoscopic radical prostatectomy is generally preferred (EAU, 2023).

Another commonly used treatment method for localized PC is radiation therapy, which utilizes ionizing radiation to kill cancer cells. The primary mechanism by which ionizing radiation kills cells is by inducing DNA damage, either by the direct action of radiation or indirectly by free radicals. Too much DNA damage leads to an impaired cell cycle and eventually programmed cell death. As cancer cells proliferate more rapidly than normal cells, they are more susceptible to radiation than are normal cells. In addition, cancer cells frequently have defective DNA damage response pathways, which decreases cell survival after radiation. External beam radiation, in which radiation is targeted to mapped cancer areas from an external radiation source, is the most common type of radiation therapy used to treat PC. The treatments are usually delivered during many therapy sessions over many weeks, with the number of sessions varying depending on the technique used. Another alternatively used radiation therapy type is brachytherapy, in which a sealed radiation source is placed internally directly into the prostate. This radiation source can be either permanent low-dose-rate seeds that slowly release radiation over time or temporary high-dose-rate catheters that are removed after treatment. Brachytherapy

is used for patients with small, low-risk PCs or combined with external radiotherapy and hormone therapy for higher-risk patients (EAU, 2023).

There seems to be little difference in efficacy between the treatments for localized PC. The survival and metastasis-free survival of patients with localized disease were similar to those of patients receiving both radical prostatectomy and external beam radiation therapy in a randomized controlled trial performed in the United Kingdom (Neal et al., 2020). However, in retrospective studies, the risk of cancer progression was lower in patients treated with radical prostatectomy than in those treated with external beam radiation therapy, although the difference was small (Nepple et al., 2013; Zelefsky et al., 2010). Nevertheless, both treatments provided a survival benefit compared to active surveillance, but with increased rates of adverse effects (Neal et al., 2020). In brachytherapy, in which the radiation dose to surrounding tissues is minimal, there seem to be fewer adverse effects (Morris et al., 2017).

In high-risk patients who have locally advanced PC, the tumor extends through the prostatic capsule and, in some cases, invades adjacent tissues. For these patients, radical prostatectomy or radiation therapy alone might not be sufficient to cure the cancer, and additional treatments are used. The most common additional treatment for locally advanced PC is hormone therapy, also known as androgen deprivation therapy (ADT), which is most often used alongside radiotherapy. Sometimes ADT is used on its own if surgery or radiotherapy is not suitable, although ADT alone is not sufficient to cure the disease but rather slows the spread and allows for the management of symptoms. For PCs that metastasize to distant organs, ADT is currently used as the first-line treatment together with second-generation antiandrogens and the chemotherapeutic agent docetaxel. Eventually, metastatic PCs, also called castration-resistant PCs, become resistant to initial treatments, and other therapeutic options are needed (EAU, 2023).

ADT is based on the blockage of androgen function in the body. In the testes, androgen deprivation can be achieved by either surgical or chemical castration by suppressing gonadotropin-releasing hormone (GnRH) production by the pituitary gland. Generally, chemical castration is preferred because of the permanent nature of surgical castration. Most chemicals widely used for castration are GnRH agonists that bind to GnRH receptors, which are responsible for the secretion of luteinizing hormone (LH) and follicle-stimulating hormone. Initial binding of GnRH antagonists releases high levels of LH, which causes a surge in serum testosterone levels, an effect called a flare. However, continued stimulation desensitizes the pituitary gland to GnRH, which leads to a reduction in LH secretion and eventually a reduction in gonadal testosterone. GnRH antagonists are also available, although

they are less commonly used. The binding of these proteins to the GnRH receptor leads to a reduction in LH. Unlike GnRH agonists, GnRH antagonists do not cause a flare-up effect that can, in some circumstances, cause cancer to develop (EAU, 2023).

Antiandrogens are regularly used together with GnRH agonists or antagonists to achieve maximal androgen blockage. Antiandrogens affect the AR signaling pathway by either affecting androgen biosynthesis in both the testes and adrenal glands or inhibiting AR. Androgen biosynthesis inhibitors target the CYP17 enzyme, which is essential for the biosynthesis of androgens. Conversely, AR antagonists inhibit the transcription factor (TF) function of AR; AR antagonists interact with the androgen-binding domain of AR and thus affect the nuclear transport of AR and the ability of AR to bind DNA. Moreover, second-generation antiandrogens, also known as AR signaling inhibitors (ARSIs), most notably the CYP17 inhibitor abiraterone acetate and the AR antagonists enzalutamide, apalutamide, and darolutamide, are currently used for the treatment of metastatic PCs and nonmetastatic CRPCs. For metastatic hormone-sensitive PCs, double combination therapy (ADT + abiraterone, apalutamide, or enzalutamide) or triple combination therapy (ADT + docetaxel + abiraterone or darolutamide) are options chosen based on patient characteristics and preference, the nature of the disease (e.g., high vs. low volume) and availability. For nonmetastatic CRPCs, the use of apalutamide, darolutamide, or enzalutamide is recommended (EAU, 2023).

In addition to the treatment options described above, many other therapy options have been studied for PC, mainly for the treatment of metastatic CRPCs (mCRPCs) after first-line therapies, with varying results. In patients with symptomatic CRPCs with bone metastasis, radium-223 injections were found to improve overall survival and thus have been approved as a therapeutic option for these patients (EAU, 2023). Lutetium-177-PSMA-617 is a radioligand therapy that utilizes the ^{177}Lu-labeled PSMA (prostate-specific membrane antigen) ligand 617 to target PSMA-positive PC cells, and it has been approved for the treatment of mCRPCs that express PSMA (EAU, 2023; Sartor et al., 2021). PARP (poly (ADP-ribose) polymerase) inhibitors that target cancers with defective DNA repair pathways, *e.g.*, due to *BRCA* alterations, are commonly used in breast cancer but have also been studied in PC. The PARP inhibitors olaparib, talazoparib, and niraparib have been approved for the treatment of patients with mCRPC in Europe and the USA, as has rucaparib (EAU, 2023; EMA, 2023a, 2023b). PARP inhibitors are often used in combination with ARSIs (*e.g.*, olaparib or niraparib + abiraterone and talazoparib + enzalutamide) for patients with *BRCA* alterations (EAU, 2023; EMA, 2023a, 2023b).

Immunotherapy, most notably immune checkpoint inhibitors and vaccines, has also been studied in PC, but with less success (Rebuzzi et al., 2022). However, the PD-1 (programmed cell death protein 1) inhibitor pembrolizumab has been approved by the U.S. Food and Drug Administration (FDA) but not by the European Medicines Agency (EMA) for the treatment of patients with advanced PC with high microsatellite instability/mismatch repair-deficient tumors, although the prevalence of such deficiencies is only 3–5% in CRPCs (EAU, 2023; Graham & Schweizer, 2022). Furthermore, an autologous cell-based therapeutic vaccine, sipuleucel-T, has been approved by the FDA for the treatment of mCRPCs with few or no symptoms due to its survival benefit (EAU, 2023).

2.1.5 General mechanisms of carcinogenesis

Uncontrollable cell division is the fundamental abnormality that results in cancer development; in cancer cells, growth control mechanisms fail, and the cells become able to proliferate indefinitely. The mechanisms that drive cancer growth can vary greatly between different tumor types and individual tumors. However, there are many similarities in the processes by which normal cells evolve into cancer cells. According to Hanahan and Weinberg, there are common physiological changes that cells acquire during the course of carcinogenesis, called cancer hallmarks (Hanahan, 2022; Hanahan & Weinberg, 2000, 2011). These include 1) self-sufficiency in growth signals, 2) insensitivity to growth-inhibitory signals, 3) evasion of apoptosis, 4) limitless replicative potential, 5) sustained angiogenesis, 6) tissue invasion and metastasis, 7) deregulation of energy metabolism, 8) avoidance of immune destruction, and the recently proposed 9) phenotypic plasticity and 10) disrupted differentiation (Hanahan, 2022; Hanahan & Weinberg, 2000, 2011). These hallmarks can occur at various times during cancer evolution in different cancers. Moreover, the hallmarks are enabled by several acquired capabilities and characteristics, such as genome instability due to genetic alterations, epigenetic reprogramming, tumor-promoting inflammation, and polymorphic microbiomes (Hanahan, 2022; Hanahan & Weinberg, 2011).

Cancer development typically requires multiple genetic and/or epigenetic changes over time. The majority of these alterations occur in somatic cells, but some can be acquired through germline inheritance. While alterations in the germline are selected against because of evolutionary pressure, they are highly tolerated in somatic cells (Martincorena et al., 2017). Nevertheless, only a minority of the alterations

affect cell proliferation and differentiation and undergo positive selection in premalignant cells. The genes affected by these alterations are called cancer driver genes.

Genetic alterations

Genetic alterations occurring in carcinogenesis can be either large chromosomal aberrations that arise during cell division or smaller mutations caused by environmental factors (*e.g.,* ultraviolet light or viral infection) or spontaneous errors, *e.g.*, during DNA replication. Chromosomal aberrations can either be differences in chromosome number (gain or loss) or genomic rearrangements in the structure of chromosomes in which large pieces of chromosomes are deleted, duplicated, inverted inside a single chromosome or inserted or translocated between two chromosomes (Figure 2). These changes can cause either a dysfunctional or absent gene product, abnormal gene dosage, or aberrant expression or regulation of the gene.

Smaller mutations can be either point mutations, where a single base is substituted by another, or insertions or deletions (indels) of single nucleotides or larger segments (1–10 000 nts) (Figure 3A). The majority of these mutations are neutral, as they do not cause any effects on the functions of either protein-coding or noncoding genes. Protein-coding genes are especially susceptible to mutations because even small changes in codon sequence can have major ramifications on protein structure. Within coding regions, point mutations can be divided based on their effect on the amino acid sequence into silent (no amino acid change), nonsense (stop codon change) and missense (amino acid change) mutations (Figure 3B). While silent mutations do not cause a change in gene function, nonsense mutations generate a truncated protein with potentially altered properties, and missense mutations may affect protein structure and thus gene function. Additionally, indels can cause frameshift mutations by shifting the open reading frame (ORF). Frequently, frameshifting leads to early termination of translation and therefore to a truncated protein product, which has a few changed amino acids near the C-terminus. While mutations in the coding regions of protein-coding genes can have a great impact on gene function, mutations in the regulatory regions of genes can also alter gene expression levels. The same can also apply to ncRNA genes. As ncRNAs function through their RNA form, the mechanisms by which mutations alter their function are less understood. However, some evidence of mutations in ncRNA exons that could cause structural changes in the folding of RNAs and potential alterations in

gene function has been reported (Bhartiya et al., 2012; Blakely et al., 2013; He et al., 2019).

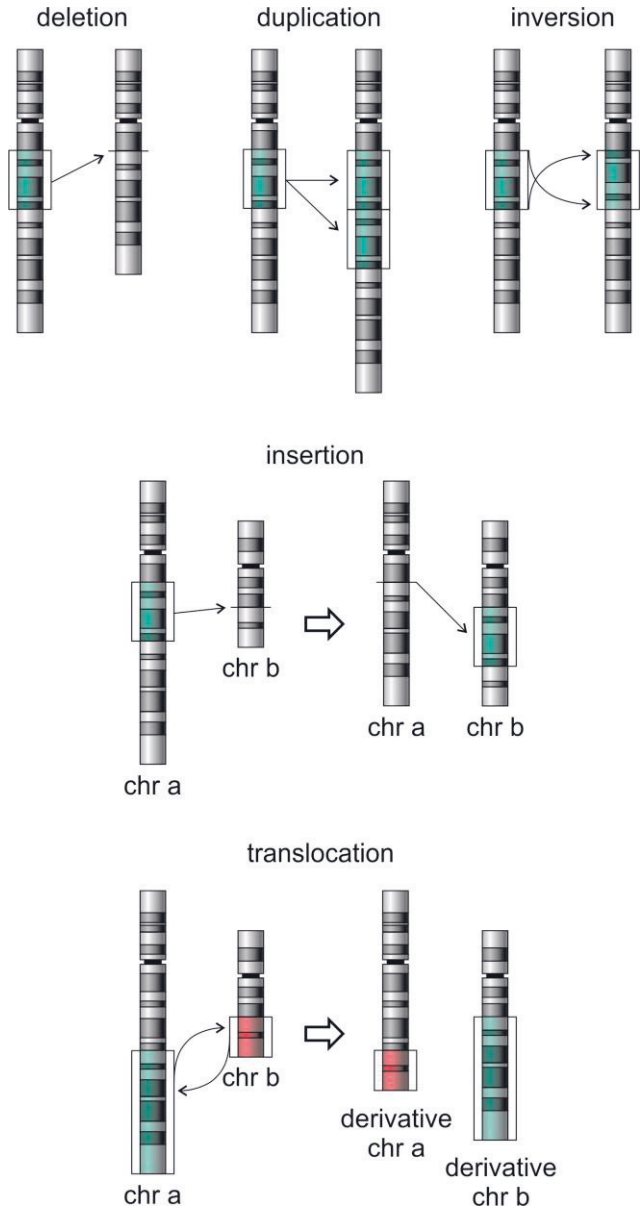


Figure 2. Genetic rearrangements. Figures are based on illustrations distributed under CC0 license.

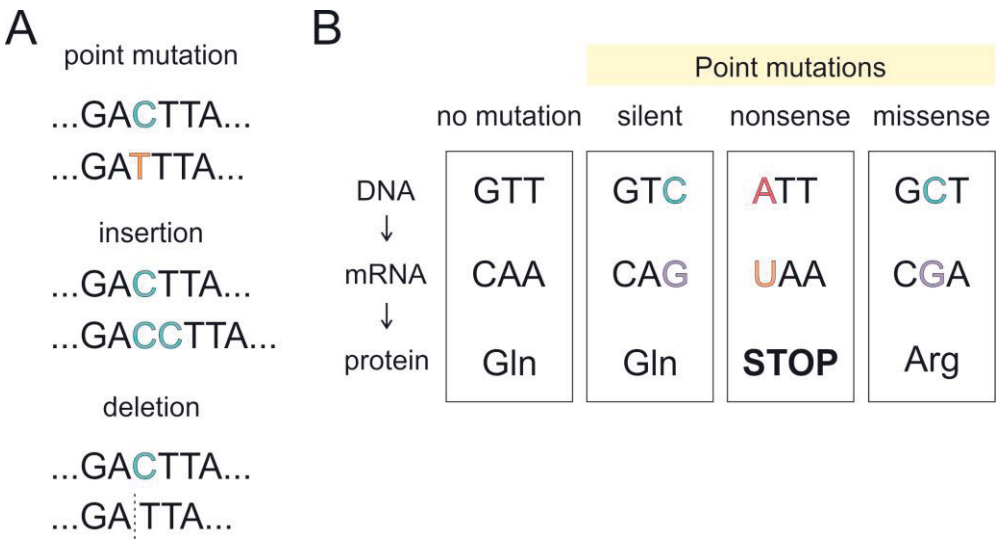


Figure 3. Point mutations and indels. A) Different mutation types. B) Effects of point mutations on mRNA and amino acid sequences.

Epigenetic alterations

In addition to genetic changes, epigenetic alterations can modify gene expression. Epigenetics is defined as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs et al., 1996). The signals establishing this stably heritable epigenetic state are thought to consist of environmental signals that trigger the intracellular pathways that initiate epigenetic processes, epigenetic initiators that respond to environmental signals and define the locations of epigenetic alterations in chromatin, and epigenetic maintainers that sustain epigenetic alterations through cell division (Berger et al., 2009). Mechanisms promoting these epigenetic processes include DNA methylation, histone modifications, and regulation by ncRNAs (Figure 4A–C) (Berger et al., 2009; Felsenfeld, 2014). When DNA methylation and histone modifications more often contribute to the maintenance of epigenetic alterations, ncRNAs assist in the initiation of these alterations (Berger et al., 2009; Felsenfeld, 2014).

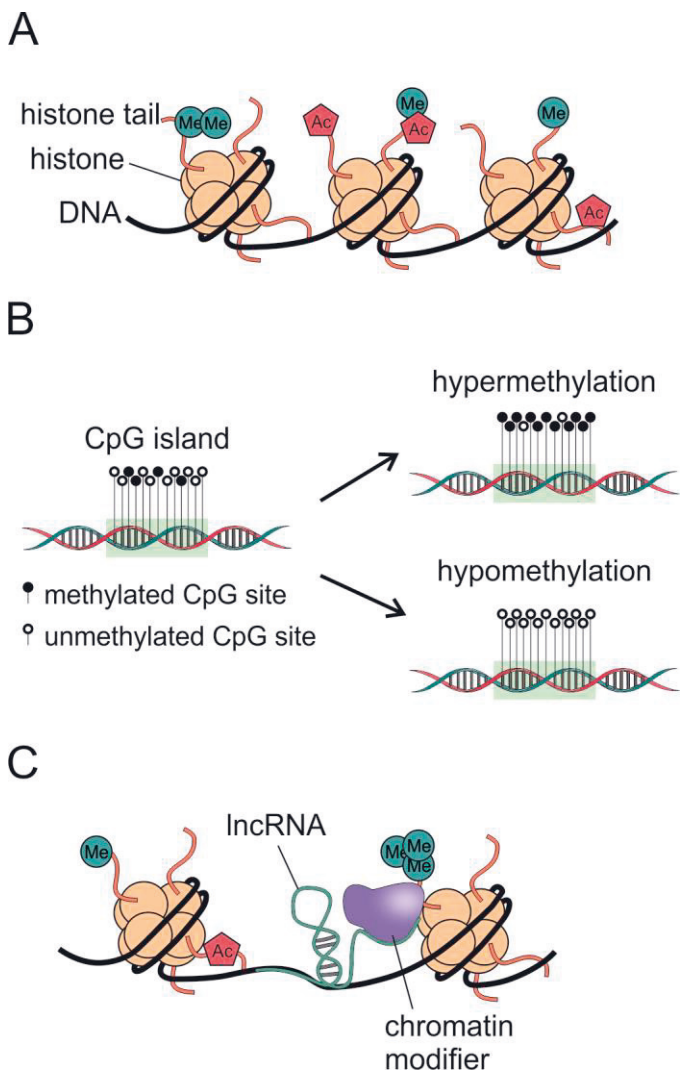


Figure 4. Epigenetic modifications. A) Histone modifications. Methylation (Me) and acetylation (Ac) are shown as examples. B) DNA methylation at CpG islands (marked as green). C) Epigenetic effects of ncRNAs. A lncRNA guiding a chromatin modifier to a correct site is shown as an example. Figures are based on illustrations in Coco *et al.* (Coco et al., 2019).

The methylation of cytosine in CG-rich sites, also called CpG islands, is the most widely studied type of DNA methylation. The majority of the CpG islands are located in gene promoters, and the methylation of these sites often leads to gene repression. In cancer, hyper- and hypomethylation of CpG islands are known to contribute to cancer through the disruption of DNA methyltransferases and the direct mutagenesis of CpG islands (Baylin & Jones, 2016).

Histone modifications include various posttranslational modifications, *e.g.,* methylation and acetylation, in the N-terminal tails or core domains of histones. These modifications and their combinations can have various effects on chromatin structure and thus on gene expression through transcriptional activation or repression. In cancer, alterations in several histone modifications contribute to carcinogenesis (Baylin & Jones, 2016).

NcRNAs are able to modulate the formation of epigenetic alterations by different means, including the downregulation of chromatin modifiers by microRNA (miRNA) silencing and the regulation of the binding and activity of chromatin modifiers by lncRNAs (Figure 4C) (Kumar et al., 2020). In cancer, the differential expression of ncRNAs is known to change the expression of their target genes that promote or suppress cancer at the epigenetic level (Kumar et al., 2020). Recently, chemical modifications of both coding and noncoding RNAs have been shown to be dysregulated in cancers and have been proposed to be involved in carcinogenesis (Barbieri & Kouzarides, 2020; Yang et al., 2020). More knowledge of the roles of RNA modifications in the epigenetic regulation of cancers is needed. It is also important to remember that not all alterations in RNA, DNA, histones or other proteins promote epigenetic alterations, as these mechanisms are also involved in more transient regulation of gene expression.

Cancer driver genes

Together, gene mutations, chromosomal aberrations, and epigenetic alterations can induce cancer by affecting cancer driver genes. These genes can either activate cell growth and differentiation (proto-oncogenes) or suppress them (tumor suppressor genes). In normal cells, the expression and function of cancer driver genes are highly regulated, but in cancer, these genes become dysregulated or defective. Proto-oncogenes and their tumor-inducing forms, oncogenes, are often growth factors or their receptors, genes involved in signal transduction, transcriptional regulators, cell cycle regulators, or antiapoptotic genes. The overexpression, gene amplification, and gain-of-function point mutations of proto-oncogenes are typical alterations that can promote oncogene activation. The presence of only one dysfunctional gene copy is often sufficient for oncogenic function. Tumor suppressor genes can also function in pathways similar to those of proto-oncogenes but also function as negative regulators or inhibitors. Generally, loss of function involving the loss of both copies by either deletion or point mutation and/or gene silencing by DNA hypermethylation is needed to promote tumor development. In addition, defects in

the DNA repair machinery can cause the accumulation of genetic changes, which may further lead to the inactivation of other tumor suppressors and the activation of oncogenes.

While alterations in malignant cells can strongly influence cancer behavior, cancer cells do not grow in a void. Rather, malignant cells grow in a complex environment full of several nonmalignant cell types (stromal cells), vasculature, extracellular matrix, and signaling molecules that together influence the growth and evolution of cancer. The most abundant stromal cell types in the tumor microenvironment are fibroblasts, endothelial cells, T cells, and macrophages. The tumor microenvironment also includes senescent cells that are nonproliferating but still metabolically active and can be disrupted to stimulate cancer cells (Hanahan, 2022). The interaction of stromal cells with malignant cells has been widely studied and is sometimes found to aid in tumor growth, maintenance, and progression (Lau & Heiden, 2020). In specific cases, the tumor microenvironment may even be an initiative force for cancer formation, *e.g.,* through inflammation, which induces genetic alterations (Greten & Grivennikov, 2019).

2.1.6 Molecular biology of prostate cancer

Prostate cancer carcinogenesis

Prostate adenocarcinoma is the most common form of PC, in which the epithelial cells of the prostatic glands begin to grow uncontrollably. As the cancer progresses, cancer cells infiltrate the stroma, and the tissue becomes increasingly less differentiated. Eventually, malignant cells may invade the surrounding tissues and metastasize to other organs. The progression from normal prostate cells to localized PC and then to advanced PC is a multistep process that consists of various genetic and epigenetic changes.

Prostate adenocarcinomas have a dominant luminal phenotype and disrupted basal layer, which is why luminal cells have been suggested to be the cells of origin for prostatic tumors. However, *in vivo* studies performed in mice suggest that basal cells might also possess the ability to become cancer progenitors that differentiate into malignant cells with luminal features (Strand & Goldstein, 2015). Notably, these studies drastically alter the microenvironment, which may also change cell behavior (Taylor et al., 2006). Recent single-cell RNA-seq studies have revealed that the prostate gland consists of different subpopulations of luminal cells, some of which

have more stem cell-like properties, while others are more differentiated (Guo et al., 2020; Karthaus et al., 2020). Further *in vivo* mouse studies revealed that some of these stem cell-like luminal cells are more prone to serve as PC initiators (Guo et al., 2020). It is also postulated that PC cells arise from premalignant lesions, of which prostatic intraepithelial neoplasia (PIN) is the most studied. High-grade PIN lesions exhibit enhanced luminal cell proliferation and can exhibit genetic alterations common to PC, such as *TMPRSS-ERG* fusion and *NKX3.1* loss (Bowen et al., 2000; Perner et al., 2007), but still contain intact basal epithelia. They are also often found in the same histological samples together with malignant lesions (Eminaga et al., 2013).

In PC, several somatic genetic alterations have been shown to participate in PC carcinogenesis. Overall, genomic rearrangements are common in localized PC. The most prevalent genetic alteration in primary and advanced PC is the *TMPRSS2-ERG* gene fusion. In addition to *ERG*, other E26 transformation-specific (ETS) family TFs have also been detected in PC gene fusions, and together, ETS fusions constitute 50–60% of the localized and advanced PCs (Robinson et al., 2015; The Cancer Genome Atlas Research Network, 2015). Other notable genomic rearrangements in primary PC include several deletions, most notably of *PTEN* (10–20%), *NKX3–1* (5–30%), *TP53* (4–20%), *FOXO1* (5–15%), *RB1* (5–15%), *CHD1* (8–10%), *CDH1* (4–5%), *CDKN1B* (2–5%), *BRCA2* (3–5%), and *BRCA1* (1–2%), and certain amplifications, mainly of *MYC* (6–10%) and *NBN* (4–5%) (Fraser & Rouette, 2019).

Coding somatic single-nucleotide mutations are relatively rare in PC, as the mutation frequency for any gene is less than 10% of that of localized PCs (Fraser et al., 2017). The most commonly mutated genes in PC are *SPOP*, *FOXA1*, and *TP53* (Fraser et al., 2017). There are multiple ways a particular gene can be altered; *e.g., PTEN* or *TP53* loss can occur due to homozygous deletion, missense mutation, or truncating mutation (The Cancer Genome Atlas Research Network, 2015).

Complex structural rearrangements, most notably chromoplexy and chromothripsis, have also been detected in PC. In chromoplexy, specific loci across multiple chromosomes are disrupted, and novel inter- and intrachromosomal rearrangements are formed in a single cell cycle (Baca et al., 2013). Chromoplexy is thought to occur at multiple times during cancer evolution. In chromothripsis, one chromosome is fragmented into thousands of small pieces that are joined randomly together during one or a few cell divisions. Chromothripsis has been found to be relatively common in localized prostate cancer (20% of cases) and is thought to occur early in PC development (Fraser et al., 2017).

The formation of genomic rearrangements in localized PCs usually originates from either DNA double-strand breaks (DSBs) or the misrepair of DSBs (Ramanand & Mani, 2019). All cells in the human body are exposed to endogenous and exogenous agents that induce DNA damage. Cellular stress damage caused by ionizing radiation, oxidative stress, or topological stress can all cause DSBs. Furthermore, the formation of DBSs by these stressors is dictated by DNA sequence and epigenetic features (Mani & Chinnaiyan, 2010). Ionizing radiation together with dihydrotestosterone (DHT), an AR ligand, has been shown to form *TMPRSS2-ERG* fusions in PC cell lines (Lin et al., 2009; Mani et al., 2009), although the prostate gland is rarely exposed to radiation. Oxidative stress can create DSBs through reactive oxygen species, *e.g.,* free radicals. Chronic inflammation, which induces oxidative stress, has been shown to be a risk factor for the initiation of PC (Marzo et al., 2007; Sfanos et al., 2018). Environmental factors, such as diet and infections, as well as aging, can trigger inflammation and thus may be a relevant source of DSBs in PC. One potential target for oxidative stress is NK3 homeobox 1 (NKX3.1), which has been shown to be downregulated by inflammatory cytokines (Markowski et al., 2008). In addition, inflammation-induced oxidative stress may also play a role in the formation of *TMPRSS2-ERG* fusions (Mani et al., 2016). Topological stress can also disrupt DNA through the action of topoisomerase II beta (TOP2B). TOP2B can cause temporary DSBs to release topological stress originating from changes in three-dimensional genomic organization and transcriptional activity, *e.g.,* in gene promoters and enhancers. In the normal prostate and in PC, a significant part of topological stress is induced by AR due to its important role as a transcriptional regulator. AR and TOP2B are highly coexpressed in PIN lesions, and androgen signaling promotes the corecruitment of AR and TOP2B to *TMPRSS2-ERG* fusion break points (Haffner et al., 2010).

As cells are constantly facing cellular stressors that cause DNA damage, cells have evolved a DNA damage response pathway to recognize the damage and activate specific DNA repair pathways to fix the damage (Chatterjee & Walker, 2017). There are several DNA repair pathways available for different types of DNA damage, and these pathways are active throughout distinct cell cycle stages and contexts (Chatterjee & Walker, 2017). The two major DBS repair pathways are homologous recombination (HR) and nonhomologous end joining (NHEJ). While NHEJ repairs the vast majority of DSBs by religating the DSB ends together, it is inhibited by HR components during the collapse of DNA replication forks in S phase and during programmed DBSs induced in meiosis, as NHEJ generates chromosomal rearrangements between different chromosomes and thus disrupts the cells (Groelly

et al., 2023). In contrast, HR uses a homologous template for repair and thus is not only less prone to error but also more complex than NHEJ (Chatterjee & Walker, 2017). Occasionally, erroneous DNA repair can lead to genomic rearrangements (Barbieri & Rubin, 2015). This may be due to the high number of DSBs that can overload the DNA repair system because of defective DNA repair machinery or because of the spatial proximity of DBSs. For example, the *TMPRSS2* and *ERG* genes are located only 2.8 Mb apart from each other in the human genome, and their close proximity, induced by androgen signaling, is thought to play a significant role in fusion formation (Mani et al., 2009).

In addition to somatic genetic alterations, certain genetic alterations in the germline have been associated with PC. Notably, PC has been found to be one of the most heritable human cancers, with 57% of the variability in disease risk in a population being due to genetic factors (Mucci et al., 2016). Nevertheless, only a few individual genes have been shown to have great significance. One of the most notable examples is *BRCA2*, which can be found in 1.3–7.9% of PCs (Lang et al., 2019). Carriers of deleterious *BRCA2* mutations are more than 4 times more likely to develop PC (The Breast Cancer Linkage Consortium, 1999). In addition, these PCs tend to be more aggressive (Risbridger et al., 2015). Similar to *BRCA2*, other hereditary mutations in genes involved in the DNA damage response and DNA repair, *e.g.,* mutations in *ATM*, *CHEK2*, *BRCA1*, *PALB2*, and *NBN*, are reported to increase PC risk (Pritchard et al., 2016). The prevalence of germline mutations in genes related to these pathways is 4.6% in localized PC but increases to 11.8% in metastatic PC (Pritchard et al., 2016). However, these mutations are still relatively rare in the population. Genome-wide association studies (GWASs) have mapped more than 100 single-nucleotide polymorphisms (SNPs), *i.e.,* common inherited single-nucleotide substitutions, which contribute to PC risk among multiancestral populations (Dadaev *et al.*, 2018; Schumacher *et al.*, 2018) (Dadaev et al., 2018). Together, these SNPs account for more than one-fourth of the familial relative risk of PC (Dadaev *et al.*, 2018; Schumacher *et al.*, 2018) (Dadaev et al., 2018). While some of these SNPs have been traced in gene bodies or regulative regions of well-known genes, the functional roles of many of these SNPs remain unknown.

In addition to genetic alterations, epigenetic modifications have also been shown to take part in PC development (Natesan et al., 2019). Both DNA hypomethylation and hypermethylation have been observed in PC. DNA hypermethylation of gene promoters has been commonly observed in tumor suppressor genes such as *APC* and *RARβ*, DNA repair genes such as *GSTP1* and *MGMT*, cell cycle control genes such as *CCND2* and *CCNA1*, and genes associated with apoptosis such as *PYCARD*

and *DAPK* and maintenance of cell–cell contacts such as *CDH1* and *CD44*, the repression of which promotes the growth and stabilization of cancer cells (Natesan et al., 2019). Alternatively, DNA hypomethylation in gene promoters is often detected in oncogenes such as *MYC* and *RAS* and leads to the activation of the transcription of those genes (Natesan et al., 2019). Hypomethylation is more common in mCRPC than in primary PC and in primary PC than in benign prostate cancer (Zhao et al., 2020). In addition to DNA methylation, certain histone modifications, especially methylation and acetylation, have been associated with PC. For example, the methylation levels of H3K4me1, H3K9me2, and H3K9me3 and the acetylation levels of H3ac and H4ac are decreased in PCs compared with benign prostates, and H3K4me1, H3K4me2, and H3K4me3 levels are increased in CRPC (Ellinger et al., 2010). Furthermore, certain histone markers, such as H3K4me2 and H3K18ac, are associated with PC progression (Natesan et al., 2019). Alterations in histone modifications affect the transcription of PC-associated driver genes (Natesan et al., 2019). For example, androgens can induce the recruitment of histone demethylases to AR target genes to mediate transcriptional activation (Metzger et al., 2005; Yamane et al., 2006). The cause of aberrant histone modification has often been traced to dysregulation or mutation in the chromatin modification machinery, either in readers, writers, or erasers of histone modifications (Natesan et al., 2019).

Epigenetic regulation by ncRNAs also occurs in PC (Liao & Xu, 2019). Among ncRNAs, miRNAs and lncRNAs have been studied the most in PC. There are examples of both oncogenic and tumor-suppressive effects that have been associated with aberrantly expressed miRNAs and lncRNAs in PC (Liao & Xu, 2019). The role of lncRNAs in PC will be discussed in Chapter 2.2.2.

ETS gene fusions

ETS family proteins are DNA-binding TFs that are encoded by 28 genes in humans and can be divided into 12 subfamilies based on their sequence similarities (Hollenhorst, McIntosh, et al., 2011). In PC, certain ETS family members are expressed abundantly due to genetic rearrangements, most notably ERG (ETS transcription factor ERG, 40–50% of PC tumors), ETV1 (ETS variant transcription factor 1, 8-10%), ETV4 (2-5%), and ETV5 (<1%) (Hu et al., 2018; Robinson et al., 2015; The Cancer Genome Atlas Research Network, 2015). These rearrangements are typically mutually exclusive and involve the fusion of the 3' ends of ETS genes with 5' fusion partners, which are frequently an androgen-responsive genes, thus resulting in ETS gene expression under AR regulation (Kumar-Sinha et al., 2008).

The most common 5' fusion partner for ETS genes in PC is *TMPRSS2* (The Cancer Genome Atlas Research Network, 2015; Tomlins et al., 2005). Among the ETS rearrangements, the great majority are *TMPRSS2-ERG* fusions (The Cancer Genome Atlas Research Network, 2015), most likely because of the close genomic proximity, 2.8 Mb, of these two genes. The region flanking exon 2 of *TMPRSS2* is most often fused with the third intron of *ERG* (Krumbholz et al., 2019; Weier et al., 2013), resulting in an N-terminally truncated ERG protein (Figure 5). However, the location of the fusion breakpoints can vary, *e.g.,* in the case of *TMPRSS2-ERG,* the exons included in the fusion range from exons 1 to 3 for *TMPRSS2* and from exons 2 to 6 for *ERG* (Clark et al., 2007). The result of ETS fusion is increased expression of either full-length or truncated ETS proteins in PC cells (Clark et al., 2007; Tomlins et al., 2007). This expression profile contrasts with that in normal prostate epithelial cells, in which PC-associated ETS proteins are not expressed (Hollenhorst et al., 2004).

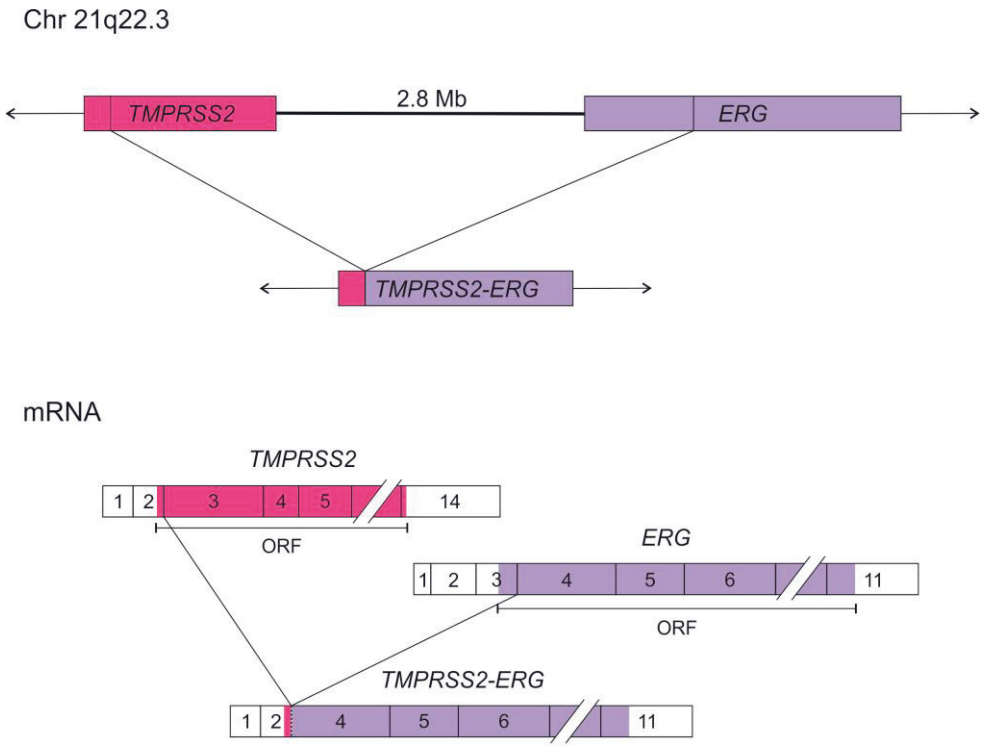


Figure 5. *TMRSS2-ERG* rearrangement. The fusion of the genes is shown at the genomic (upper figure) and mRNA (lower figure) levels. The figures are based on illustrations in Tandefelt et al. (Tandefelt et al., 2014).

As the initiation of cancer typically requires multiple genetic or epigenetic alterations, ETS gene fusion alone is also insufficient to drive PC carcinogenesis, and at least one additional alteration is needed (Aytes et al., 2013; Carver et al., 2009; Higgins et al., 2015; King et al., 2009). The most common genetic alteration associated with ETS rearrangements in PC is loss of *PTEN* (Han et al., 2009; Reid et al., 2010; Taylor et al., 2010). ETS rearrangements occur early in PC development, even in PIN lesions, indicating that they promote prostate carcinogenesis (Carver et al., 2009; Clark et al., 2007; Perner et al., 2007). These findings are supported by experiments in normal immortalized prostate cells, PC cell lines, and mouse xenografts suggesting that ETS proteins have oncogenic properties (Aytes et al., 2013; Hollenhorst, Ferris, et al., 2011; Kedage et al., 2016; Mesquita et al., 2015; Shao et al., 2012; Tomlins et al., 2008). In addition, ETS overexpression is maintained in advanced PC (Cai et al., 2009; Leinonen et al., 2013), although whether ETS contributes to an aggressive phenotype is uncertain.

ETS proteins all share a conserved ETS DNA-binding domain (DBD), but there is very little sequence homology between different ETS subfamilies (Laudet et al., 1999). DBDs of all ETS proteins recognize a core GGA(A/T) motif (Karim et al., 1990; Nye et al., 1992), although the nucleotides surrounding the core motif are also used for binding and can differ between ETS proteins (Wei et al., 2010). Genome-wide occupancy studies indicate that there are two types of binding sites for ETS proteins: 1) redundant binding sites, where multiple ETS proteins are able to bind and are often located near the TSSs of housekeeping genes, and 2) specific binding sites, which are more frequently found in the enhancers of genes regulating ETS-specific biological functions (Hollenhorst et al., 2009). The same has also been found for the PC-associated ETS proteins ERG, ETV1, ETV4, and ETV5 (Hollenhorst, Ferris, et al., 2011). It has been proposed that in cancer, these oncogenic ETS proteins function through specific binding sites. Part of the specificity is attributed to the extended binding motifs that can vary between ETS proteins (Wei et al., 2010). For PC-associated ETS proteins, the consensus binding sequence is the same: ACCGGAAGT (Wei et al., 2010). The cooperation of ETS proteins with other factors in the same regulatory region can also increase the specificity of ETS binding. There are various partners that have been associated with ETS proteins (Hollenhorst, McIntosh, et al., 2011), including AR.

In PC, ETS proteins, especially ERG, co-occupy or cobind to the majority of the same regulatory regions as AR (Massie et al., 2007; Yu et al., 2010). ERG can also physically interact with AR (Mounir et al., 2016; Wasmuth et al., 2020; Yu et al., 2010) and assist AR in binding to new genomic sites with low AR occupancy (Cai et al.,

2013; Chen et al., 2013; Wasmuth et al., 2020). Similar cooperation with ERG has been detected for forkhead box protein A1 (FOXA1) and homeobox protein Hox-B13 (HOXB13) (Kron et al., 2017), which can also modulate the AR binding landscape in PCs (Pomerantz et al., 2015; Sahu et al., 2011). This modified AR binding can lead to aberrant expression of target genes. Whether ERG is activating or repressing the transcriptional activity of AR is postulated to be context dependent. ERG can repress the AR pathway by cooperating with corepressors such as the enhancer of zeste 2 polycomb repressive complex (EZH2) and histone deacetylase 1/2 (HDAC1/2) to restrain epithelial differentiation and promote epithelial–mesenchymal transition (Chng et al., 2012) or by activating androgen-independent oncogenes that induce the invasion and growth of PC cells (Hollenhorst, Ferris, et al., 2011; Yang et al., 2017; Yu et al., 2010). Conversely, ERG can enhance the survival of AR-dependent PC cells, especially in the context of PTEN loss, which, together with ERG overexpression, promotes AR signaling (Chen et al., 2013).

Because of its high specificity and frequency in PC, the potential of *TMPRSS2-ERG* fusion as a biomarker has been widely studied. The most common methods for detecting *TMPRSS2-ERG* gene fusion in tissue specimens are fluorescence *in situ* hybridization (FISH) and polymerase chain reaction (PCR), which determine the occurrence of gene rearrangements and expression level, respectively (Tomlins et al., 2005). Immunohistochemistry (IHC) can also be used to detect ERG protein overexpression in tissue specimens (Tomlins et al., 2009). Additionally, *TMPRSS2-ERG* fusion can be detected noninvasively in urine samples by PCR or similar techniques (Laxman et al., 2006). The use of *TMPRSS2-ERG* fusion as a diagnostic marker for PC has been studied in various sample types using these methods (Yao et al., 2014). According to a meta-analysis, there is indeed a significant association between *TMPRSS2-ERG* detection and PC diagnosis (Yao et al., 2014). While the specificity for PC is high, the overall sensitivity is close to 50% (Yao et al., 2014), which is also near the prevalence of *TMPRSS2-ERG* in PC. Thus, the use of *TMPRSS2-ERG* for first-line screening is not advisable, but when TMPRSS2-ERG is used as a marker from urine, it may serve as a quick and noninvasive method for the diagnosis and confirmation of PC (Yao et al., 2014). Several methods have combined *TMPRSS2-ERG* detection with other PC-specific markers, most notably *PCA3* (Tomlins et al., 2016). There have also been many studies trying to connect *TMPRSS2-ERG* with outcomes in patients with PC, but the results have been variable (Leinonen et al., 2010; Pettersson et al., 2012; Song & Chen, 2018). According to meta-analyses, *TMPRSS2-ERG* is associated with tumor stage at

diagnosis but does not strongly predict either biochemical recurrence or lethal disease (Pettersson et al., 2012; Song & Chen, 2018).

Androgen receptor signaling

The AR signaling pathway is essential for the growth and development of both the normal prostate and PC. AR is a TF and a steroid hormone receptor that uses steroids such as testosterone as ligands for its activation and translocation to the nucleus. The *AR* gene is located on the X chromosome and consists of 8 exons encoding a 110 kDa protein (Figure 6A). The AR protein contains four distinct domains: an NH_2-terminal domain (NTD), a DBD, a short flexible hinge region harboring the nuclear localization signal, and a carboxy-terminal ligand-binding domain (LBD) (Figure 6A). In normal physiological settings, AR binds testosterone or DHT via its LBD. When these ligands are unavailable, AR remains bound to a complex containing heat shock proteins (HSPs) and other chaperone proteins to prevent the degradation of AR (Figure 6B) (Smith & Toft, 2008). These functionally inactive AR complexes are located mainly in the cytoplasm and are anchored to cytoskeletal elements. Testosterone entering prostate cells can be converted to the more potent DHT by 5a-reductase (Figure 6B) (Dai et al., 2017), which then binds to AR (Figure 6B). This binding leads to the release of AR from the HSP complex, the dimerization of two ligand-bound AR molecules, and the nuclear translocation of the AR dimer (Figure 6B) (Dai et al., 2017). In the nucleus, AR dimers bind to androgen response elements (AREs), which are specific DNA sequences recognized by the DBD of the AR (Figure 6B). At these sites, AR, together with coregulators, other TFs (general and specific), and RNA polymerase II, forms a transcription initiation complex that initiates the transcription of AR-regulated target genes (Dai et al., 2017). These genes are involved in cell proliferation, differentiation, and survival.

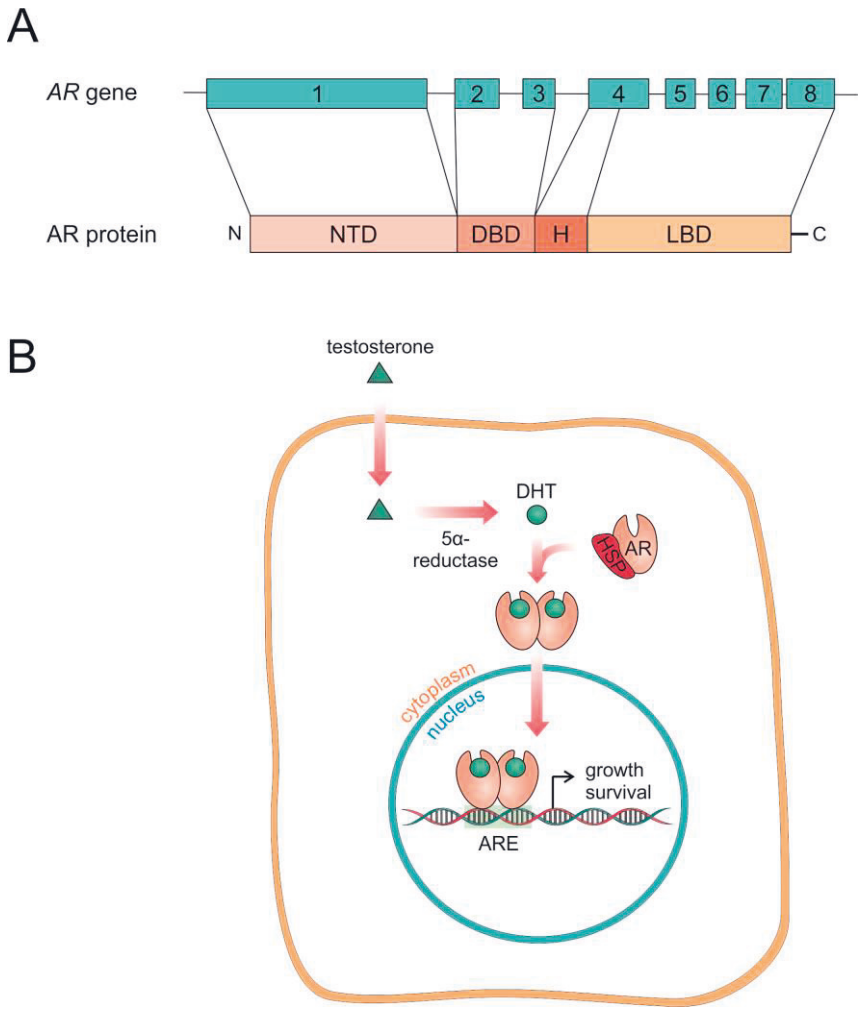


Figure 6. Androgen receptor function. A) *AR* exons and corresponding protein domains. B) AR signaling in cells. The figures are based on illustrations in Tan et al. (Tan et al., 2015).

The binding of AR to AREs is flexible and dependent on the cellular context. The consensus sequence for AR is an inverted repeat of the AGAACA half-site separated by three bases (Claessens et al., 2017). The affinities by which AR monomers bind to these half-sites vary: one AR monomer binds to the upstream half-site with high affinity, while the other AR monomer binds to the adjacent half-site with lower affinity (Claessens et al., 2017). This relaxed binding allows AR to bind to AREs with greater selectivity (Sahu et al., 2014). According to genome-wide studies, the AR binding landscape, also called the cistrome, undergoes rewiring during PC progression. This results in alterations to the target genes to which AR binds and

subsequently regulates (Pomerantz et al., 2015; Sharma et al., 2013; Stelloo et al., 2015). The mechanisms responsible for this altered binding include changes in coregulators and other TFs that interact with AR and participate in the AR transcriptional complex (Mills, 2014). In early PC, ERG, FOXA1, and HOXB13 interact with AR and modulate the AR cistrome to attain a neoplastic phenotype (Kron et al., 2017; Pomerantz et al., 2015; Wasmuth et al., 2020). In addition, aberrant expression of certain AR coregulators, most notably p160 steroid receptor coactivators, such as steroid receptor coactivators 1 to 3 and nuclear receptor coactivators (NCOAs) 1 to 3, and certain chromatin modelers, such as histone acetyltransferase p300, CREB-binding protein, and histone acetylases 1 to 3, have been known to act as coactivators or corepressors for AR regulation in PC development and progression (Foley & Mitsiades, 2016).

The shift from AR to regulate the differentiation of normal prostate cells to drive their malignancy is considered to be an early and central event in prostate carcinogenesis (Copeland et al., 2018). Furthermore, AR continues to play a central role in advanced PC. The AR dependence of advanced PC and ADT, the therapy form utilizing this dependence, was introduced in the 1940s (Huggins & Hodges, 1941). Since then, ADT has been further developed and currently includes various methods to inhibit the transcriptional activity of AR (see Chapter 2.1.4 Treatment). However, advanced PC that is treated with ADT eventually relapses as the disease evolves into CRPC. In these patients, *AR* amplification, which leads to *AR* overexpression, is the most frequent genetic alteration observed after ADT in tumors (Koivisto et al., 1997; Linja et al., 2001; Robinson et al., 2015; Visakorpi et al., 1995). AR overexpression is thought to sensitize ADT-treated PC cells to low androgen levels, which eventually leads to castration resistance (Waltering et al., 2009). In addition to *AR* amplification, other alterations in the *AR* or in the AR pathway are common in CRPC. These include point mutations in AR that enable AR activation by alternative steroid hormones (Jernberg et al., 2017), truncated AR variants (*e.g.*, AR-V7) that are ligand-independent and constitutively active (Dehm & Tindall, 2011), AR upstream enhancer amplification that leads to increased *AR* expression (Quigley et al., 2018; Takeda et al., 2018; Viswanathan et al., 2018), and alterations in AR-associated TFs and coregulators such as FOXA1, NCOAs 1 and 2, and NCORs 1 and 2 (Robinson et al., 2015; Taylor et al., 2010). Modifications in the *AR* and AR signaling pathways play major roles in castration resistance and cancer recurrence but occasionally also occur in hormone-naïve PCs, albeit at a much lower frequency (Taylor et al., 2010; The Cancer Genome Atlas Research Network, 2015).

PI3K/AKT/mTOR signaling

The PI3K/AKT/mTOR pathway is an essential signal transduction network in all eukaryotic cells involved in the regulation of the survival, growth, and proliferation of cells in response to external stimuli, such as nutrients, hormones, and growth factors (Figure 7) (Glaviano et al., 2023). Alterations in the PI3K/AKT/mTOR pathway that lead to elevated signaling are very common in various cancers, including PC (Glaviano et al., 2023; Shorning et al., 2020). Approximately 40% of primary PCs and 70–100% of advanced PCs have a dysregulated PI3K/AKT/mTOR pathway (Abida et al., 2019; Carver et al., 2011; Robinson et al., 2015; Taylor et al., 2010). The most common genetic alterations in the PI3K/AKT/mTOR pathway in PC (both primary PC and mCRPC) are *PTEN* (16.4–32%), *DEPTOR* (5.1–21.4%), *SGK3* (5.6–20.5%), *FOXO1/3* (0–15.2%/4.5–13.4%), *MAP3K7* (5.9–14.8%), *RRAGD* (6.5–14.4%), *SESN1* (5.4–13.6%), *PIK3CA* (5.5–11.5%), *PIK3C2B* (1.4–11.5%), and *PDPK1* (0–8.1%) (Shorning et al., 2020).

The phosphoinositide 3-kinase (PI3K) family consists of three classes of signal transducers, of which only class I kinases are related to cell growth (Jean & Kiger, 2014). Class I PI3Ks are activated by external signals through membrane receptors, usually either receptor tyrosine kinases (RTKs) or G protein-coupled receptors (Figure 7) (Fruman et al., 2017). When activated, class I PI3Ks phosphorylate phosphatidylinositol 4,5-bisphosphates (PIP_2) into phosphatidylinositol (3,4,5)-trisphosphates (PIP_3), which leads to the recruitment of several effector proteins with pleckstrin homology (PH) domains, including phosphoinositide-dependent kinase-1 (PDK-1) and AKT, to membrane signaling complexes (Figure 7) (Fruman et al., 2017). Class I PI3Ks consist of one of four catalytic subunits and one of several regulatory subunits that form heterodimers and can be further divided into type IA and IB PI3Ks (Fruman et al., 2017). Catalytic subunits of type IA PI3Ks, especially *PI3KCA* and, less frequently, *PIK3CB*, are altered by gain-of-function mutations or amplifications in various cancers, including prostate tumors (Glaviano et al., 2023; Shorning et al., 2020). *PIK3CA* gain has been found to correlate with poor PC prognosis and to function as a genetic driver in PC (Pearson et al., 2018). In addition, *PIK3CA* alterations often coincide with *PTEN* loss in patients with PC and can synergistically cooperate *in vivo* to accelerate carcinogenesis and cancer progression via hyperactivation of the PI3K/AKT/mTOR pathway (Pearson et al., 2018).

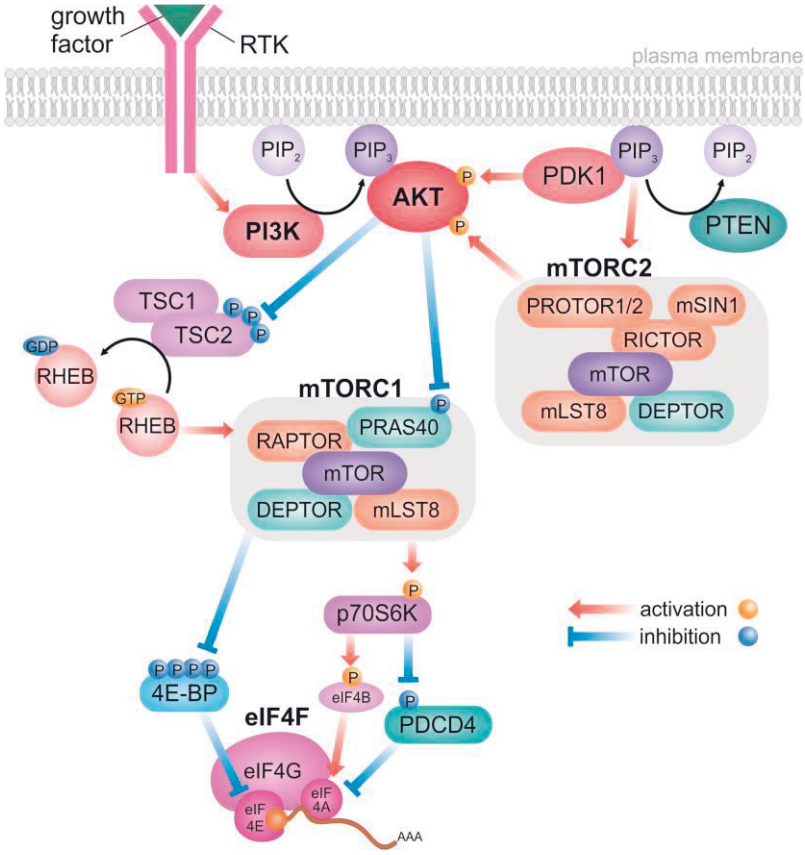


Figure 7. Schematic representation of the PI3K/AKT/mTORC1 signaling network. Signaling molecules related to 5'cap-dependent protein synthesis are shown.

Phosphatase and tensin homolog (PTEN) is a negative regulator of the PI3K/AKT pathway that dephosphorylates PIP_3 back to PIP_2 (Figure 7) (Fruman et al., 2017). Loss-of-function mutations or deletions in *PTEN* lead to sustained signals of cell growth and survival through overactivation of the PI3K/AKT pathway, and these mutations are very common in several cancers (Glaviano et al., 2023). *PTEN* loss is also often observed in both primary and CRPC (Shorning et al., 2020), although it occurs more frequently in advanced PC (McMenamin et al., 1999). *PTEN* loss is associated with poor outcomes in patients with PC (Ahearn et al., 2016; Chaux et al., 2012; Lotan et al., 2015; McMenamin et al., 1999). In mouse studies, *PTEN* loss has been shown to be a genetic driver of invasive PC (Mulholland et al., 2011; Wang et al., 2003).

AKT, also known as protein kinase B, is part of the AGC serine/threonine kinase family, which consists of three isoforms, AKT1, AKT2, and AKT3 (Manning &

Toker, 2017). After its recruitment to the plasma membrane by PIP_3, AKT is phosphorylated by PDK-1 and mechanistic target of rapamycin complex-2 (mTORC2) at different phosphorylation sites (Figure 7). While phosphorylation by PDK-1 at Thr308 in AKT1 is sufficient to mediate specific downstream responses, phosphorylation by mTORC2 at Ser473 in AKT1 is needed for maximal AKT activity (Manning & Toker, 2017). Activated AKT targets several substrates, including the proline-rich AKT substrate of 40 kDa (PRAS40), which is an inhibitory subunit of mTORC1, and tuberous sclerosis complex 2 (TSC2) (Figure 7) (Manning & Toker, 2017). Although the phosphorylation of PRAS40 at Thr246 by AKT reduces its inhibitory effect on mTORC1, the main mechanism of mTORC1 activation by AKT involves the inhibition of TSC2 through its phosphorylation at multiple residues (Manning & Toker, 2017). TSC2 is a part of a larger TSC complex that promotes the conversion of RHEB-GTP (Ras homolog enriched in brain—guanosine triphosphate) to RHEB-GDP (diphosphate) (Figure 7). In the GTP form, RHEB is an important activator of mTORC1 (Manning & Toker, 2017).

In certain cancers, *AKT* is altered by activating mutations or amplifications. In PC, the frequency is usually low: the prevalence of *AKT1* mutations is 0.4–0.9%, and that of AKT1 amplifications is 0.2–4.5% in primary PC and CRPC (Glaviano et al., 2023; Shorning et al., 2020). Similarly, *TSC2* mutations and deep deletions, *i.e.*, putatively homozygous deletions according to RNA-seq data, and *RHEB* mutations or amplifications occur at low frequencies in PC: the prevalence of *TSC2* mutations is 1–1.8%, that of deletions is 0–4.2%, that of *RHEB* mutations is 0–0.1%, and that of amplifications is 0.6–4%. However, activation of AKT or RHEB and inhibition of TSC in cancers occur more often through alterations upstream of AKT (Glaviano et al., 2023).

mTOR is a serine/threonine kinase that acts through mTORC1 and mTORC2 complexes. In addition to mTOR, these complexes have many other structural and regulatory components, some of which are shared between the complexes, while others are distinct. mTORC1 consists of mTOR, RAPTOR (regulatory protein associated with mTOR), and mLST8 (mammalian lethal with SEC13 protein 8, also known as GßL), in addition to two inhibitory subunits, PRAS40 and DEPTOR (DEP domain containing mTOR-interacting protein), mTORC2 consists of mTOR, mLST8, and DEPTOR but has RICTOR (rapamycin-insensitive companion of mTOR) instead of RAPTOR and the additional regulatory subunits mSIN1 and PROTOR1/2 (Saxton & Sabatini, 2017). The functional roles of these complexes also differ. While mTORC2 phosphorylates and fully activates AKT and other AGC kinases, mTORC1 has multiple substrates through which it increases the production

of proteins, lipids, and nucleotides and suppresses catabolic pathways, such as autophagy (Saxton & Sabatini, 2017). Activation of mTORC2 occurs through mSIN1 inhibition at the PH domain by PIP_3, whereas mTORC1 activation is promoted essentially through TSC (Figure 7) (Saxton & Sabatini, 2017). Additionally, AKT, AMPK, and Rag GTPases can directly phosphorylate mTORC1 subunits (Saxton & Sabatini, 2017).

In PC, mutations and deep deletions of mTOR complex components that negatively affect mTOR activation are rare (mutations 0–1.6% and deletions 0–2%), but amplifications of components that promote mTOR activity are more frequent, especially in mCRPC (5–7.7%) (Shorning et al., 2020). Interestingly, the mTOR inhibitor *DEPTOR* is often amplified in PCs (5.1–21.4%), which is also associated with poor prognosis in patients with PC according to The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) data (Shorning et al., 2020). The role of DEPTOR in cancer cells has been shown to be complex, and in certain cancers, it may have oncogenic effects (Catena & Fanciulli, 2017).

mTORC1 promotes messenger RNA (mRNA) translation mostly through two key effectors, ribosomal protein S6 kinase beta-1 (p70S6K) and eIF4E-binding proteins (4E-BPs), both of which regulate translation through the eukaryotic initiation factor 4F (eIF4F) complex (Figure 7) (Roux & Topisirovic, 2018). eIF4F is composed of three subunits, a 5'-cap-binding protein eIF4E, a scaffold protein eIF4G, and an ATP-dependent RNA helicase eIF4A, which together with additional factors recruit mRNA to the ribosome (Roux & Topisirovic, 2018). mTORC1 activates p70S6K by phosphorylation at Thr389, which in turn activates multiple substrates that induce translation initiation, including eIF4B, which is a positive modulator of the eIF4F complex (Figure 7) (Roux & Topisirovic, 2018). p70S6K also phosphorylates programmed cell death 4 (PDCD4) at Ser67, which leads to the ubiquitination and degradation of PDCD4 (Wang & Yang, 2018). PDCD4 is an inhibitor of translation initiation that binds eukaryotic translation initiation factor 4A (eIF4A) and inhibits its RNA helicase activity, thus suppressing the translation of RNAs with structured 5' untranslated regions (UTRs) (Figure 7) (Wang & Yang, 2018). In addition to p70S6K, 4E-BPs also regulate eIF4F by repressing it through eIF4E binding (Figure 7) (Roux & Topisirovic, 2018). mTORC1 inactivates 4E-BPs by hierarchal phosphorylation, which leads to the dissociation of 4E-PBs from eIF4E (Roux & Topisirovic, 2018).

Mutations in p70S6K or 4E-BPs are not detected in cancers, nor are 4E-BP deletions, but in certain cancers, amplification of p70S6K is observed (Artemenko et al., 2022; Wang et al., 2014). However, hyperactivation of p70S6K and

hyperphosphorylation of 4E-BPs due to alterations in the PI3K/AKT/mTOR pathway are common in various cancers (Artemenko et al., 2022; Wang et al., 2014). For example, in patients with PC, high levels of phospho-4E-BP1 and eI4E are associated with increased mortality (Graff et al., 2009). PDCD4 has also been found to function as a tumor suppressor in *in vitro* and mouse studies, although no genetic alterations have been reported in patient tumors (Matsuhashi et al., 2019). Low levels of genetic alterations (deep deletions, amplifications, or mutations, <6%) in *PDCD4* can be detected in pancancer whole-genome sequencing data via cBioPortal (Cerami et al., 2012; ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020). In PC, alterations in *PDCD4* are even less prevalent, as 1% of patients in the TCGA-PRAD cohort had deep deletions of PDCD4 (Cerami et al., 2012; The Cancer Genome Atlas Research Network, 2015). In addition to genetic alterations, PDCD4 levels decrease posttranscriptionally in many cancers (Matsuhashi et al., 2019). A widely studied mechanism of PDCD4 regulation involves miRNAs, especially miR-21, which is upregulated in several cancers, including PC (Matsuhashi et al., 2019). *PDCD4* mRNA has a target sequence for miR-21 at its 3'-UTR that is used for targeted degradation by the miRNA-induced silencing complex (Matsuhashi et al., 2019). In PC, miR-21-induced *PDCD4* degradation is AR-mediated (Zennami et al., 2019). PDCD4 downregulation has also been shown to be a potential prognostic marker for many solid tumors (Li et al., 2016), but there have been no studies on this topic in patients with PC.

In addition to PCDC4, other partners of the PI3K/AKT/mTOR pathway can also interact with AR signaling (Pungsrinont et al., 2021). For example, in PC cells with *PTEN* loss, PI3K and AKT inhibition elevates AR protein expression, which in turn activates AR-mediated gene expression (Carver et al., 2011; Mulholland et al., 2011). AR inhibition also upregulates AKT signaling, which may facilitate resistance to ADT (Pungsrinont et al., 2021). Inhibitors targeting PI3K, AKT, and mTOR have been studied in clinical trials with patients with mCRPC, although in many cases, their efficacy has been limited, at least as monotherapies (Pungsrinont et al., 2021). However, several studies have suggested that combined inhibition of the PI3K/AKT/mTOR and AR pathways has potential for the treatment of advanced PCs (Tortorella et al., 2023).

2.2 Long noncoding RNAs

The Human Genome Project revealed that only approximately 1.2% of the human genome encodes proteins (International Human Genome Sequencing Consortium, 2004). Hence, the function of the majority of the genome is still a mystery. In 2003, the Encyclopedia of DNA Elements (ENCODE) project was established to identify and analyze functional elements in the human genome. This project utilized next-generation sequencing technologies to investigate whole transcriptomes and their regulation. The findings from these studies estimated that nearly 75% of the genome is actively transcribed and that 80% of the genome participates in biochemical events (Djebali et al., 2012; The ENCODE Project Consortium, 2012). These findings suggested that ncRNAs, which are not translated into proteins, are much more abundant than initially thought.

Over the years, dozens of different classes of ncRNAs with various functional roles have been identified. Based on their regulatory roles, ncRNAs can be divided into two main groups: housekeeping ncRNAs and regulatory ncRNAs. Whereas housekeeping ncRNAs regulate generic cellular functions, such as translation (ribosomal RNAs and transfer RNAs) and splicing (small nuclear RNAs), regulatory ncRNAs are modulators of gene expression. Regulatory ncRNAs are often further divided based on their size into small ncRNAs, *e.g.*, miRNAs that silence gene expression by targeting mRNAs and piwi RNAs that silence transposable elements (TEs), and lncRNAs that modulate gene expression via various mechanisms. LncRNAs are classically defined as a group of ncRNAs that are >200 nt long. However, based on the gained knowledge of ncRNA biology over the last decade, an updated recommendation of >500 nt has been suggested (Mattick et al., 2023). LncRNAs are a highly heterogeneous group of RNAs with various structures and functions. In the human genome, the estimated number of different lncRNA genes varies from nearly 18 000 (Gencode v34) to over 90 000 (Fang et al., 2018; Frankish et al., 2019), depending on the annotation method. Automated transcriptome assembly approaches yield greater numbers but more inaccurate annotations. Thus far, lncRNAs have often been grouped according to their genomic position and orientation relative to protein-coding genes (Figure 8), as the biological functions of the majority of lncRNAs are still unknown. Nevertheless, multiple studies have revealed that lncRNAs can function as key regulators in various biological processes, including pathogenesis, such as cancer, through several mechanisms.

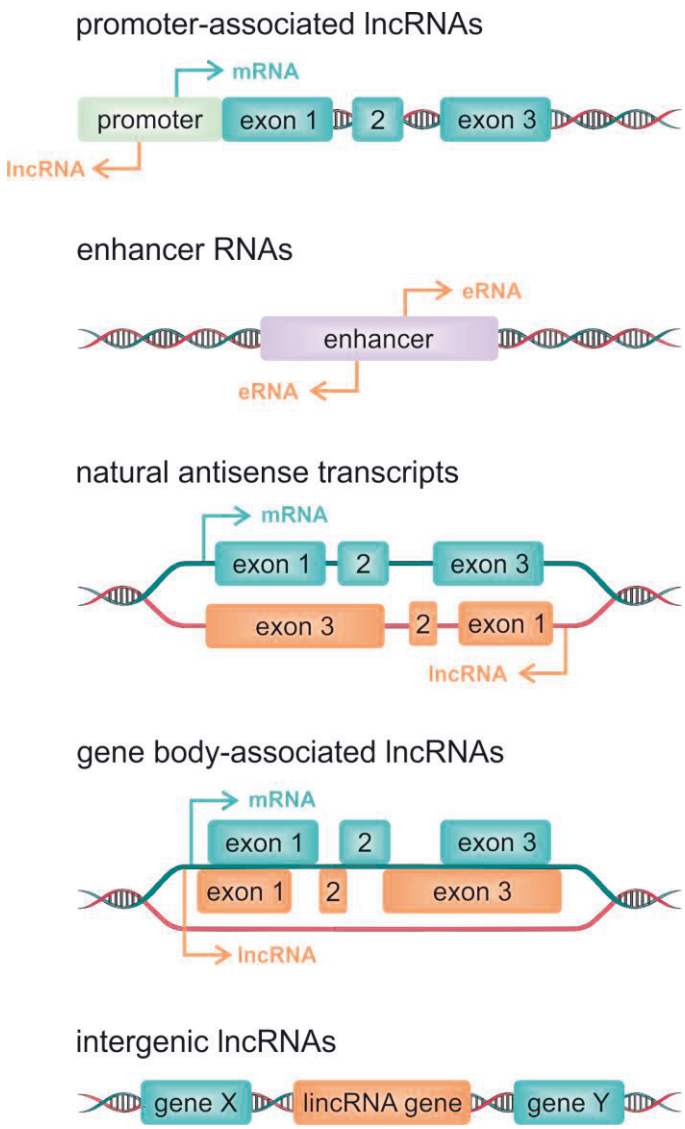


Figure 8. LncRNA classification based on genomic context.

2.2.1 Long intergenic noncoding RNAs

The majority of the characterized lncRNA genes are located in the intergenic region between two protein-coding genes (Figure 8). These ncRNAs are called long (or large) intergenic ncRNAs (lincRNAs). As they are autonomously transcribed and do not overlap with coding sequences, lincRNAs are thought to function more

independently of nearby protein-coding genes than are genic lncRNAs (Ransohoff et al., 2018). In many ways, lincRNAs resemble mRNAs without protein-coding potential, but there are also differences in the abundance and specificity of the transcription of these RNAs compared to mRNAs. Like other lncRNAs, lincRNAs are regulators of gene expression via either the transcript itself or the transcription process.

Characteristics of lincRNAs

LincRNAs resemble mRNAs in many aspects: lincRNAs are transcribed by RNA polymerase II, they can have multiple exons, they have a 5' cap, and they can be polyadenylated (Quinn & Chang, 2016). However, the transcription of lincRNAs is less precise than that of mRNAs, and lincRNAs are more often cotranscriptionally cleaved than spliced and prematurely terminated (Schlackow et al., 2017). In addition, lincRNAs have fewer and longer exons, are expressed at 10-fold lower levels, and are more tissue-specific than mRNAs (Cabili et al., 2011; Derrien et al., 2012; Lagarde et al., 2017; Melé et al., 2017).

LincRNA primary sequences are generally less evolutionarily conserved than mRNA sequences, but their promoters are equally conserved (Guttman et al., 2009; Ponjavic et al., 2007). It has been shown that fewer TFs bind to the promoters of lincRNA genes and that they have fewer histone modifications than mRNA promoters (Melé et al., 2017). In addition, certain TFs (*e.g.*, GATA TFs) are more conservatively bound to lincRNAs than to mRNA promoters (Melé et al., 2017). It has also been discovered that certain histone modifications are more enriched on lincRNAs. For example, histone 3 lysine 9 trimethylation (H3K9me3) often occurs on lincRNA promoters. Although H3K9me3 is a canonical repressive marker, it is associated with tissue specificity in active lincRNA genes (Melé et al., 2017).

The subcellular localization of lincRNAs differs from that of mRNAs; mRNAs are trafficked to the cytoplasm for translation, and lincRNAs are more often located in the nucleus (Djebali et al., 2012). This could be partly explained by differences in processing, as lincRNAs found in the nucleus are less efficiently spliced and polyadenylated, which can also lead to their rapid degradation by exosomes (Melé et al., 2017; Schlackow et al., 2017). Shorter half-lives could indicate a lack of functionality for nuclear lincRNAs, at least in their RNA form. Nevertheless, many nuclear lincRNAs may escape or avoid this nuclear surveillance process and play important functional roles in cells (Statello et al., 2021). Conversely, cytoplasmic lincRNAs are more efficiently spliced than their nuclear counterparts, although not

as efficiently as mRNAs (Melé et al., 2017). Cytoplasmic lincRNAs are also exported from the nucleus through the same pathways as mRNAs but preferentially through a pathway involving transcripts with fewer exons (Zuckerman et al., 2020). Interestingly, when lincRNAs and other lncRNAs are exported into the cytoplasm, 70% of them are bound to ribosomes (Carlevaro-Fita et al., 2016). Some ribosome-associated lncRNAs were found to be degraded at ribosomes, likely due to translation-coupled nonsense-mediated decay (Carlevaro-Fita et al., 2016), but for others, the reason for polysomal localization is still unknown. Cytoplasmic lincRNAs can also be sorted to other parts of the cell, such as specific organelles, or can remain in the cytoplasm by interacting with diverse RNA-binding proteins (Statello et al., 2021). However, the mechanisms by which lincRNAs are transported to their sites of function are still largely unknown.

Molecular mechanisms of lincRNA function

Due to the limited conservation of the primary sequences of lincRNAs and other lncRNAs, it has been suggested that these lncRNAs lack functions and are mainly byproducts of the transcriptome machinery (Ponting et al., 2009). However, lncRNAs contain shorter sequences, promoters, and splice signals that are highly evolutionally conserved, and many lncRNAs reside in the same genomic neighborhood across species (Deniz & Erman, 2017). This indicates that factors other than the primary sequence may play stronger roles in lincRNA function. Indeed, RNA structure mapping experiments in whole transcriptomes (Aw et al., 2016; Lu et al., 2016; Smola et al., 2015; Spitale et al., 2015) and targeted lincRNA studies (Hawkes et al., 2016; McHugh et al., 2015; Somarowthu et al., 2015) have revealed structural motifs in lincRNAs that rely more on secondary and tertiary structures than on primary sequences. For example, *XIST*, one of the most studied lincRNAs that has a critical function in X chromosome inactivation, consists of multiple clusters of tandem repeats that each have a separate function and can interact with different protein partners (Loda & Heard, 2019). Taken together, similar to proteins, lincRNAs appear to consist of domains that have different functional properties.

One source of functional elements in lincRNA domains may be transposable elements (Johnson & Guigó, 2014). This is supported by the discovery that TEs are more enriched in lincRNAs than in other genetic elements, including protein-coding genes (Kelley & Rinn, 2012). There are also a few examples of the functional properties of TEs in lncRNAs. For instance, an Alu element embedded within a

lncRNA was found to form an RNA duplex with another Alu element in an mRNA and induce the degradation of the mRNA (Gong & Maquat, 2011). Additionally, TEs may function as regulatory elements by modulating lincRNA expression (Kapusta et al., 2013) or as nuclear localization signals (Lubelsky & Ulitsky, 2018).

The functional domains of lincRNAs are thought to interact either through sequence complementarity or by relying on the secondary and tertiary structures of their RNA strands (Mattick et al., 2023). LincRNAs can interact with DNA, other RNAs, and proteins and can achieve their mechanisms of action through these interactions (Mattick et al., 2023; Zhang et al., 2019). LincRNAs can be involved in the regulation of gene expression either locally, in *cis*, or distant from the site of their transcription, in *trans* (Kopp & Mendell, 2018). This regulation may occur via several mechanisms in both the nucleus and cytoplasm (Figure 9A–H) (Ransohoff et al., 2018; Yao et al., 2019).

In the nucleus, lincRNAs can regulate nuclear organization, transcription, and splicing (Figure 9A–E) (Ransohoff et al., 2018; Yao et al., 2019). At the chromosomal level, lincRNAs can act as scaffolds that interact with chromatin and/or nuclear proteins to modulate chromatin architecture (*e.g.*, *XIST* and *Firre*; Figure 9A–B) or nuclear bodies (*e.g.*, *NEAT1* and *MALAT1*) to achieve coordinated transcriptional regulation and/or transcript processing (Yao et al., 2019). Moreover, lincRNAs can participate in the regulation of gene expression through the recruitment of chromatin modifiers that change the epigenetic landscape of regulatory elements (*e.g.*, *XIST*, *HOTTIP*, and *HOTAIR*; Figure 9C) (Yao et al., 2019). At the transcriptional level, lincRNAs have been found to regulate the expression of neighboring genes (Liu et al., 2017), although it appears that these lincRNAs often function in a transcript-independent manner (*e.g.*, *Blustr* and *PVT1*), as the act of transcription or the regulatory DNA element inside the lincRNA locus is sometimes the source of the function rather than the transcript itself (Cho et al., 2018; Engreitz et al., 2016; Groff et al., 2016). LincRNAs can also affect TF binding, *e.g.*, by competing with DNA (*e.g.*, *Gas5*) or by assisting TFs in binding to DNA (*e.g.*, *Xist*; Figure 9D) (Kino et al., 2010; Loda & Heard, 2019). At the posttranscriptional level, nuclear lincRNAs have been reported to play a role in the modulation of splicing (*e.g.*, *MALAT1*; Figure 9E) (Romero-Barrios et al., 2018).

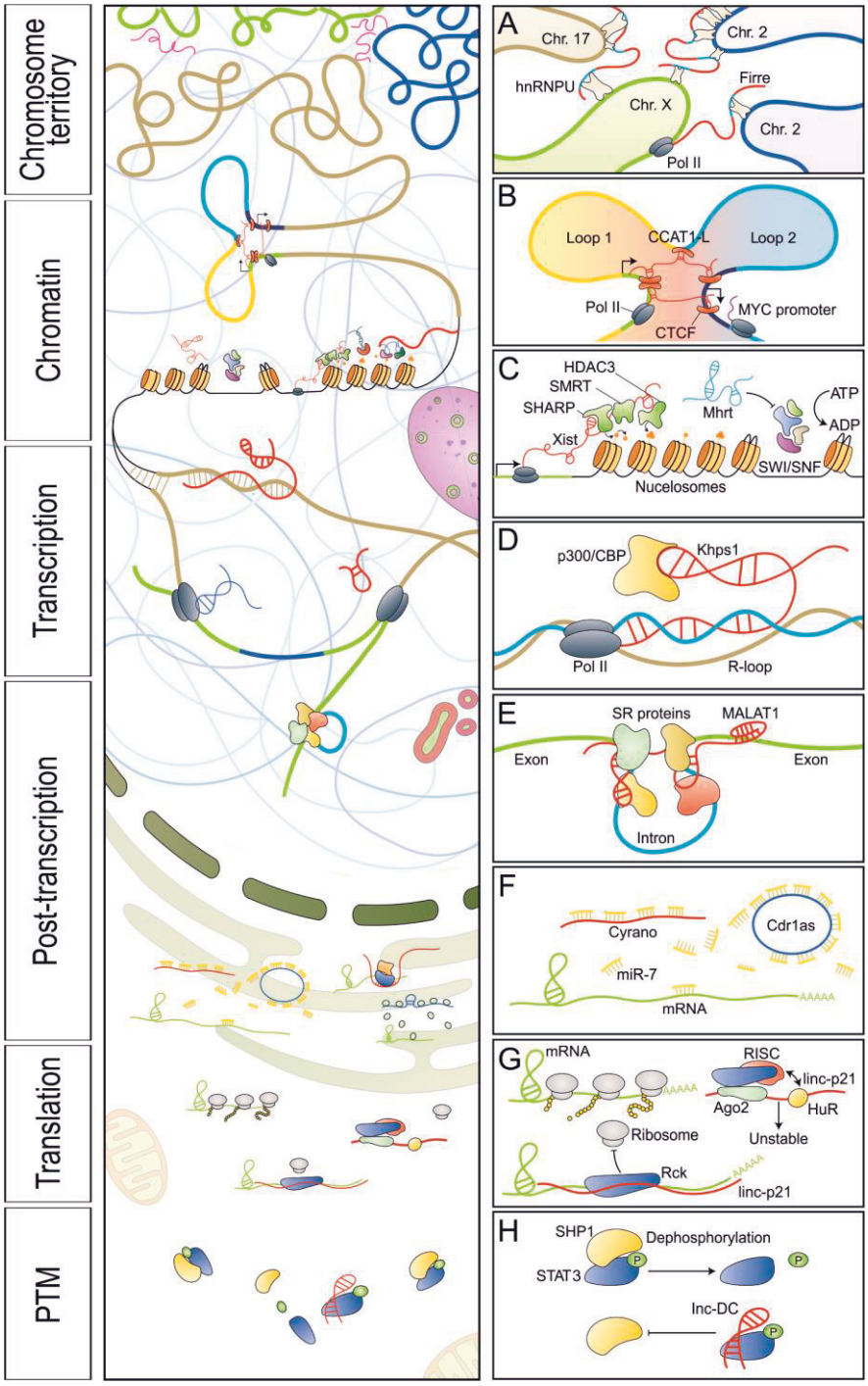


Figure 9. Functional mechanisms of lincRNAs and other lncRNAs in the nucleus and cytoplasm. A) *Firre* transcripts localize to their transcription site and five additional autosomal

chromosomal loci *in trans* to affect interactions between distant genomic regions. B) *CCAT1-L* accumulates *in cis* to modulate chromatin loops between enhancers and the promoter of *MYC*. C) LncRNAs regulate chromatin accessibility. Left, *Xist* recruits HDAC1-associated repressor protein (SHARP), silencing the mediator for retinoid and thyroid hormone receptor (SMART) and HDAC3 to silence Xi. Right, *Mhrt* prevents SWI/SNF binding to corresponding DNA loci. D) *Khps1* enhances Pol II transcription by forming an R-loop that anchors Khps1-interacting p300/CBP to the *SPHK1* promoter. E) *MALAT1* interacts with SR proteins and alters their phosphorylation to impact pre-mRNA splicing in splicing speckles. F) A regulatory network consisting of different types of ncRNAs. *Cyrano*, which harbors *miR-7* binding sites, targets *miR-7* for degradation, and prevents *miR-7* from repressing its target RNAs, including the circRNA *Cdr1as*. G) LncRNAs regulate translation. The association of *lincRNA-p21* (linc-p21) with HuR favors the recruitment of let-7/Ago2, leading to its destabilization. In the absence of HuR, *lincRNA-p21* targets mRNAs to repress their translation by recruiting the translation repressor Rck129. RISC, RNA-induced silencing complex. H) LncRNAs modulate posttranslational modifications (PTMs). *Lnc-DC* directly interacts with STAT3 to prevent its dephosphorylation by SHP1. The figure was modified from Yao et al. (Yao et al., 2019) with permission from Springer Nature.

In the cytoplasm, lincRNAs can modulate mRNA turnover, translation, and posttranslational modifications (Figure 9F–H) (Ransohoff et al., 2018; Yao et al., 2019). Several lincRNAs act as competing endogenous RNAs or "sponges" (*e.g.*, *linc*R*o*R and *TUGI*; Figure 9F); these lincRNAs share sequence similarity with miRNAs and can therefore compete for miRNA binding, which leads to the increased expression of target mRNAs (Ransohoff et al., 2018; Salmena et al., 2011). Similarly, lincRNAs containing *Alu* elements can bind mRNAs containing these elements, which may lead to STAU1-mediated mRNA decay (Gong & Maquat, 2011). Certain lincRNAs can also function as decoys for RNA-binding proteins that mediate mRNA decay (*e.g.*, *NORAD*) (Soghli et al., 2021). Ribosome-associated lincRNAs have also been identified, some of which regulate the translation of their target mRNAs by recruiting translation repressors (*e.g.*, *lincRNA-p21*; Figure 9G) (Yoon et al., 2012). In addition, lincRNAs may contain short ORFs that can be translated into small peptides of less than 100 amino acids (*e.g.*, MRLN and NoBody). These micropeptides can function by inhibiting enzymes or interacting with mRNA decapping complexes (Anderson et al., 2015; D'Lima et al., 2017; Vitorino et al., 2021). A few lncRNAs (*e.g.*, *lnc-DC* and *NKILA*) can modulate posttranslational modification by masking the modification sites or sites bound by posttranslational modification enzymes (Figure 9H) (Yao et al., 2019), a mechanism that lincRNAs could also utilize. Although extensive investigations of individual lincRNAs have revealed many molecular mechanisms during the last decade, we still do not know how the majority of lincRNAs function or if we can apply these principles to all lincRNAs.

2.2.2 Prostate cancer-associated long noncoding RNAs

The first lncRNAs that were associated with PC were initially identified in studies that aimed to identify biomarkers for PC diagnosis and prognosis. These genes include *PCA3* and *PCGEM1*, which are both highly prostate-specific and are overexpressed in PC (Bussemakers et al., 1999; Srikantan et al., 2000). Wider screens to identify PC-associated lncRNAs were possible only after the advancement of whole-transcriptome sequencing technologies and the discovery of thousands of novel lncRNAs. The first systematic study to discover PC-associated lncRNAs was performed by Prensner et al. in 2011 and employed RNA sequencing (RNA-seq) to evaluate samples from patients with PC (Prensner et al., 2011). They reported 121 lincRNAs that they called PC-associated transcripts (PCATs), which were differentially expressed in localized and metastatic PC specimens compared to benign adjacent tissue (Prensner et al., 2011). In addition, various other RNA-seq and microarray-based studies have identified additional PC-associated lncRNAs (Crea et al., 2014; Du et al., 2013; Iyer et al., 2015; Yan et al., 2015). While the differential expression-based approach has been successful in identifying new candidates for biomarkers and players in cancer development, very little is known about the functionality of lncRNAs. Only some of these lncRNAs have been further characterized and found to have functional roles in PC development (Mirzaei et al., 2022; Mitobe et al., 2018). In addition, alternative approaches have been utilized to identify functional PC-associated lncRNAs. For example, certain lncRNAs have been connected with genomic regions and SNPs that have been associated with PC risk (Chung et al., 2011; Gao et al., 2018; Guo et al., 2016; Hua et al., 2018; Jin et al., 2011).

LncRNA biomarkers in prostate cancer

LncRNAs are highly tissue-specific and aberrantly expressed in cancers (Cabili et al., 2011; Yang et al., 2014), which makes them suitable candidates for biomarkers. In fact, one of the few PC-associated biomarkers approved by the FDA is *PCA3* (Groskopf et al., 2006; Sartori & Chan, 2014), a PC-associated lncRNA. *PCA3* is highly overexpressed in PC compared to nonmalignant prostatic tissue, and it is highly PC-specific (Bussemakers et al., 1999). Furthermore, *PCA3* can be detected in urine samples (Hessels et al., 2003), which are noninvasive and thus cause fewer adverse effects than tissue biopsies. The FDA-approved Progensa® PCA3 assay measures the concentration of amplified *PCA3* molecules in post-DRE urine

samples (Groskopf et al., 2006; Hologic, 2019). The assay measures a score that compares the *PCA3* levels to the *KLK3* (gene encoding for PSA) mRNA levels (*PCA3*/*KLK3* * 1000). While *PCA3* is highly PC specific, it is still expressed at low levels in other cells (de Kok et al., 2002) when *KLK3* expression is highly prostate specific. The PCA3 assay can be used to determine whether a biopsy is needed or whether another biopsy is needed after the first negative biopsy, as high PCA3 scores are associated with an increased likelihood of a positive biopsy (Gittelman et al., 2013; Wei et al., 2014).

More recently, potential prognostic lncRNA biomarkers have been discovered in PC (Xu et al., 2018), of which *SChLAP1* has been one of the most extensively studied (Mehra et al., 2016; Prensner et al., 2014). *SChLAP1* is overexpressed in a subset of PC tumors and has been associated with metastasis in two independent datasets (Böttcher et al., 2015; Prensner et al., 2013, 2014). In addition, *SChLAP1* can independently predict lethal PC in formalin-fixed, paraffin-embedded (FFPE) tissue specimens from radical prostatectomies via RNA *in situ* hybridization (RNA-ISH) (Mehra et al., 2014, 2016). *SChLAP1* can also be detected in post-DRE urine, in which *SChLAP1* expression is increased in a subset of samples, especially from patients with intermediate and high PC risk (Prensner et al., 2014).

Functions of lncRNAs in prostate cancer

LncRNAs have been shown to play important roles in carcinogenesis, tumor progression, and drug resistance in several cancers (Liu et al., 2020; Schmitt & Chang, 2016). As with protein-coding genes, this is due to lncRNA dysregulation in cancer cells. Aberrant lncRNA expression, either by genetic or epigenetic alterations, abnormal TF binding in the regulatory elements of lncRNAs, or lncRNA gene body mutations that affect lncRNA structure, are potential mechanisms of lncRNA dysregulation (Minotti et al., 2018; Schmitt & Chang, 2016). These dysregulated lncRNAs can further modulate well-known oncogenic signaling pathways, although a few lncRNAs have also been associated with the tumor microenvironment (Liu et al., 2021). Essentially, lncRNAs have been linked to every hallmark of cancer (Schmitt & Chang, 2016).

In PC, many lncRNAs have been found to promote or repress PC proliferation (*e.g.*, *PCGEM1*, *PCAT1*, *MALAT1*, and *HOTAIR*) or progression (*e.g.*, *SChLAP1*, *HOTTIP*, *MALAT1*, and *DANCR*) in knockdown and overexpression studies *in vitro* or *in vivo* (Mirzaei et al., 2022; Mitobe et al., 2018). In some cases, this phenotypic effect has been linked to dysregulation of key signaling pathways in PC, including

androgen and PI3K/AKT signaling pathways (Mirzaei et al., 2022; Mitobe et al., 2018). Several PC-associated lncRNAs have also been found to regulate or be regulated by these pathways (Mirzaei et al., 2022; Mitobe et al., 2018). For example, *PCGEM1*, *PCA3*, *HOTAIR*, *CTBP1-AS*, and *ARLNC1*, together with many other PC-associated lncRNAs, are androgen-regulated (Mitobe et al., 2018; Takayama et al., 2011; Zhang et al., 2015, 2018), of which *HOTAIR*, *CTBP1-AS*, and *ARLNC* have also been found to modulate AR signaling, resulting in feedback regulation (Takayama et al., 2013; Zhang et al., 2015, 2018). In addition to regulating oncogenes, lncRNAs may also inhibit tumor suppressor genes. For example, certain PI3K/AKT-associated lncRNAs can promote this pathway through PTEN inhibition, including in PC (*e.g.*, *MCM3AP-AS1* and *LINC01296*) (Mirzaei et al., 2022). Mechanistically, PC-associated lncRNAs may function similarly to other lncRNAs (Mirzaei et al., 2022; Mitobe et al., 2018). Several of these lncRNAs have miRNA target sites and are thought to function as miRNA sponges by competing with their target mRNAs (*e.g.*, *PCAT7*, *UCA1*, and *LINC00665*) (Mirzaei et al., 2022; Mitobe et al., 2018). Many nuclear PC-associated lncRNAs interact with transcription factors or epigenetic regulators, such as *SRA*, *MALAT1*, *HOTTIP*, and *CTBP1-AS* with AR, EZH2, TWIST1-WDR5, and PSF, respectively, to modulate the transcription of target genes at the chromosomal level (Mitobe et al., 2018). However, for most PC-associated lncRNAs, the mechanism of action is still unknown.

3 AIMS OF THE STUDY

The aims of this study were to identify clinically significant and biologically relevant novel lincRNAs in PC, characterize them, and study their regulation and function during PC progression. The specific aims were as follows:

1. To discover unannotated lincRNAs that are differentially expressed between benign prostate and untreated PC or untreated PC and locally recurrent CRPC.

2. To identify novel lincRNAs that are specifically expressed in PC and associated with PC progression to discover potential biomarkers for PC.

3. To examine the transcriptional regulation of novel lincRNAs by essential PC-associated signaling pathways.

4. To study the role of novel lncRNAs in PC cell function.

4 MATERIALS AND METHODS

4.1 Research models

4.1.1 Cell lines and xenografts (I, II, III)

The PC cell lines 22Rv1, DU145, LNCaP, and PC3; the normal prostate cell line RWPE-1; the breast cancer cell line CAL-51; and the mouse fibroblast line NIH-3T3 were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). The normal prostate cell line PrEC was purchased from Lonza (Basel, Switzerland). The VCaP and DuCaP PC cell lines were a gift from Prof. J. Schalken (Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands). The LAPC-4 PC cell line was kindly provided by Dr. Charles Sawyers (University of California at Los Angeles, Los Angeles, CA, USA), and the normal prostate cell line EP156T was provided by Dr. Varda Rotter (Weizmann Institute of Science, Rehovot, Israel). Parental LNCaP cells that were transfected with either empty pcDNA3.1(+) (LNCaPpcDNA3.1) or wild-type AR-cDNA (LNCaP-ARhi) were previously established by our group (Waltering et al., 2009). The cells were cultured under the recommended conditions and regularly tested for mycoplasma contamination. The previously established xenografts, LuCaP69 and LuCaP73, were gifts from Dr. Robert L. Vessella (University of Washington, Seattle, WA, USA).

4.1.2 Clinical patient samples and clinical data (I, II, III)

Fresh-frozen tissue samples were obtained from Tampere University Hospital (Tampere, Finland). The samples were snap frozen and stored in liquid nitrogen. In publication I, fresh-frozen tissue samples from 12 BPHs, 28 untreated primary PCs, and 13 locally recurrent CRPCs were used; the samples contained ≥70% cancerous or hyperplastic cells. This sample set is referred to as the Tampere PC cohort. In publication II, 87 untreated PCs were used; the amount of cancer cells in these samples ranged from 30–80%. In publication III, FFPE samples from an untreated

PC and a CRPC tumor and two fresh-frozen tissue samples from untreated PC samples were used for *EPCART* studies. In addition, in publication III, prostate tissue microarray (TMA) samples of 111 untreated PCs and 60 locally recurrent CRPCs were used for IHC analysis of PDCD4.

BPH samples were obtained by radical prostatectomy, cystoprostatectomy, or transurethral resection of the prostate (TURP). Untreated PCs were obtained by radical prostatectomy. Locally recurrent CRPC samples were acquired by TURP, and the patients were treated with orchiectomy (6 patients), LHRH (3 patients), bicalutamide and orchiectomy (1 patient), LHRH and bicalutamide (1 patient), or estrogen (1 patient). Biochemical progression was defined as increasing blood PSA levels and two consecutive samples with PSA $\geq$ 0.5 ng/ml. The use of clinical material was approved by the ethics committee of the Tampere University Hospital (Tampere, Finland; ETL code R03203). Written informed consent was obtained from all subjects.

4.2 Modification of gene expression

4.2.1 RNA knockdown (I, II)

Small interfering RNA (siRNA)-mediated knockdown of protein-coding genes (*ERG*, *ETV4*, and *AR*) was carried out via reverse transfection with Lipofectamine RNAiMAX (Invitrogen) and 25 nM siRNAs (Sigma–Aldrich) according to the manufacturer's instructions (publications I and II). In publication I, *PCAT5* knockdown was performed by forward transfection of PC-3 and 22Rv1 cells with INTERFERin (PolyPlus Transfection) and 50 mM siRNAs (IDT) or by reverse transfection of DuCaP cells with Lipofectamine RNAiMAX (Invitrogen) and 25 nM siRNAs (IDT) according to the manufacturer's instructions. For lncRNA siRNA screening, knockdown was conducted by either forward or reverse transfection with Lipofectamine RNAiMAX (Invitrogen) and 10 nM (22Rv1, LNCaP, PC-3, and RWPE-1 cells) or 20 nM (DuCaP and VCaP cells) siRNA (IDT or Sigma–Aldrich), unless otherwise specified. Commercial nontargeting siRNAs (Sigma–Aldrich, IDT, Qiagen, or Ambion, Thermo Fisher Scientific) were used as negative controls (NCs). A no-template (NT) control was used to monitor the effects of the transfection reagents.

Antisense oligonucleotide (ASO) knockdown of *EPCART* was carried out by reverse transfection with Lipofectamine RNAiMAX (Invitrogen) and 50 nM locked nucleic acid (LNA) oligonucleotides (Exiqon) according to the manufacturer's instructions. Nontargeting LNA oligonucleotides (Exiqon) were used as negative controls. An NT control was used to monitor the effects of the transfection reagents.

4.2.2 CRISPR/Cas9 knockout (II)

In publication II, GenScript's CRISPR Gene Editing Services was used to knock out *EPCART* in LNCaP cells and validate the deletion. Gene editing was performed by two single guide RNAs (sgRNAs) that were designed for the area covering the promoter, 1st exon, and 2nd exon. The sgRNAs were cloned and inserted into the AIO-1.0-Cas9-GGG-2A-EGFP vector and cotransfected by Celetrix electroporation. Single-cell clones were produced, and full deletion of the designated area was confirmed by PCR and Sanger sequencing for two cell clones (called del-4 and del-56); one clone without the deletion was used as a control (called WT). We analyzed the expression of *EPCART* in the cell clones by droplet digital PCR (ddPCR).

Whole-genome sequencing (WGS) was performed for the *EPCART* deletion and WT clones to validate the deletion site. Genomic DNA (gDNA) was extracted with an AllPrep DNA/RNA Kit (Qiagen). Library preparation and sequencing were performed by Novogene Co. (Beijing, China). For library construction, gDNA was randomly fragmented by sonication, and then the DNA fragments were end polished, A-tailed, and ligated with the full-length adapters of Illumina sequencing, followed by further PCR amplification with P5 and indexed P7 oligos. The PCR products used for the final construction of the libraries were purified with the AMPure XP system. Libraries were subsequently checked for size distribution via an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantified via real-time PCR (to meet the criteria of 3 nM). WGS was performed on a NovaSeq 6000 (Illumina) platform for 150 bp paired-end reads. On average, 622 million reads per sample were obtained.

4.2.3 Hormone deprivation and androgen induction (II)

In publication II, the effect of DHT on the expression of novel lncRNAs was examined in hormone-deprived PC cells. The cells were grown in RPMI 1640

medium without phenol red containing charcoal/dextran-treated serum for four days and treated with DHT (0 nM or 10 nM) for 24 hours.

4.2.4 Stable overexpression

To construct the *EPCART* overexpression plasmid, the sequence for the *EPCART* transcript (exons 2–5) was synthesized with additional restriction sites (*Nhe*I and *Xho*I) and then added to the pcDNA3.1(+) plasmid (Invitrogen) by GenScript. Either the pcDNA3.1(+) empty expression vector (Invitrogen) or pcDNA3.1(+) containing *EPCART* was transfected into LNCaP or PC-3 cells with Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's instructions. A stable cell pool was selected with 400 μg/ml geneticin (G418; Invitrogen) for several weeks, after which a lower geneticin concentration (200 μg/ml) was used for stable cell line maintenance. *EPCART* expression was determined by quantitative reverse transcription PCR (qRT–PCR).

4.3 Gene expression analysis

4.3.1 RNA extraction (I, II, III)

Total RNA was extracted from fresh-frozen tissues in publications I and II and from cell lines in publications I, II, and III. Freshly frozen tissue blocks were cut into 10x20-micrometer sections using a cryotome. In publication I, RNA from the tissue sections was extracted with an AllPrep RNA/DNA Mini Kit (Qiagen) according to the manufacturer's instructions. In publication II, RNA from the sections was extracted with TRIzol (Invitrogen) according to the manufacturer's instructions. TRIzol was also used for RNA extraction from cell lines in publications I, II, and III. In publication III, TRIzol LS (Invitrogen) was used for the extraction of polysome profiling fractions. For the siRNA-knockdown and DHT-induction samples in publication II and for the RNA-sequencing (RNA-seq) samples in publication III, TRIzol-extracted RNA was treated with DNase I and purified with RNeasy Mini Spin Columns (Qiagen) according to the manufacturer's instructions.

In publication III, *EPCART* localization was studied in subcellular fractions of PC cell lines. Nuclear and cytoplasmic RNA was extracted with a SurePrep Nuclear or Cytoplasmic RNA Purification Kit (Fisher BioReagents) following the

manufacturer's instructions. The expression levels of *EPCART*, the cytoplasmic control (*GAPDH*), and the nuclear control (*U1*) were analyzed by qRT–PCR.

4.3.2 Quantitative polymerase chain reaction (I, II, III)

The PCR-based gene expression quantification methods used either qRT–PCR or ddPCR.

In publications I and III, qRT–PCR was carried out using Maxima RT (Thermo Scientific) and Maxima SYBR Green qPCR Master Mix (Thermo Scientific) according to the manufacturer's instructions and by using CFX Opus, CFX96, or CFX384 real-time PCR detection systems (Bio-Rad) for detection. In publication II, BioMark HD (Fluidigm), a high-throughput qPCR system, was used for screening the expression levels of lncRNAs. cDNA synthesis (Reverse Transcription Master Mix) and preamplification (Preamp Master Mix) reagents were purchased from Fluidigm and used following the manufacturer's instructions. Quantification was performed with an EvaGreen-based detection system (SsoFast EvaGreen Supermix with Low ROX, Bio-Rad) on 48.48 Dynamic Arrays (Fluidigm) following the manufacturer's instructions. Relative expression was calculated by the ΔCq (quantification cycle) or 2^-ΔΔCq method by using *TBP* as a reference gene in publications I, II, and III. In publication II, unsupervised hierarchical clustering was performed for ΔCq values that were quantified relative to the median expression levels of the genes in the tissue samples. Clustering was performed using the complete-linkage agglomerative clustering method based on the Euclidean distance matrix and visualized using the R package gplots version 3.0.1.

DdPCR was used for absolute quantification of RNA transcripts in publication II. RNA was reverse transcribed by Maxima RT (Thermo Scientific), and ddPCR was conducted with QX200 ddPCR EvaGreen Supermix (Bio-Rad) according to the manufacturers' instructions. PCR was performed in a T100 Thermal Cycler (Bio-Rad) and detected in a QX200 droplet digital PCR system (Bio-Rad). The concentrations of transcripts in the samples were calculated and normalized to that of *TBP*.

In publication III, the RNA content in subcellular fractions was calculated as the percentage of transcript abundance = 2^[Cq(total RNA)+Cq(RNA fraction)]*100, where the total RNA abundance is the sum of the nuclear and cytoplasmic fractions. The percentage of RNA in polysome profiling fractions was calculated for each

fraction as %RNA = $2^{-Cq_x} / (2^{-Cq_1} + 2^{Cq_2} + \cdots + 2^{-Cq_x}) * 100$, where x = the number of the fraction that was calculated.

4.3.3 RNA sequencing (I, III)

RNA-seq was carried out for Tampere PC cohort patient samples in publication I and for *EPCART*-deletion and WT clones in publication III. Library preparation and sequencing were performed by BGI (Beijing Genomics Institute, Beijing, China) in publication I and by Novogene Co. (Beijing, China) in publication III. Total RNA isolated from each sample was used for poly(A) RNA isolation by oligonucleotide (dT) beads. Fragmentation buffer was used for random RNA fragmentation. Fragmented RNA was used for first-strand cDNA synthesis by random hexamer primers, after which reagents were added to initiate second-strand cDNA synthesis. For library preparation in publication III, strand-specific libraries were constructed by generating second-strand cDNA by incorporating dUTP into the reaction mixture. Second-strand cDNA synthesis was followed by end repair, 3' end adenylation, and sequencing adapter ligation. Uracil digestion was included as an additional step in publication III to remove the dUTP-containing strands. The double-stranded cDNA library was completed through size selection and PCR enrichment. Sequencing of 150 bp paired-end reads was performed by using an Illumina HiSeq2000 (publication I) and an Illumina NovaSeq 6000 (publication III). On average, 110 million clean reads per sample were obtained in publication I, and 105 million clean reads were obtained in publication III.

4.3.4 RNA *in situ* hybridization (III)

In publication III, the amount and localization of *EPCART* transcripts were studied by RNA-ISH. FFPE tissue sections were treated using the ViewRNA ISH Tissue 2-Plex Assay (Affymetrix) or RNAscope 2.0 HD Detection Kit (Red) (Advanced Cell Diagnostics) according to the manufacturer's instructions. First, the slides were briefly deparaffinized in xylene and dehydrated in 100% ethanol. Sections were then pretreated and boiled, and a target probe and signal amplifier were hybridized using a ThermoBrite System (Leica Biosystems) or a HybEZ II System (Advanced Cell Diagnostics). Probes for human housekeeping genes (*GAPDH*, *ACTB*, and *PPIB*) were used as positive controls, and a probe for bacterial dihydrodipicolinate reductase (dapB) was used as a negative control in every assay. Signal detection was

performed by using Fast Red substrate. The slides were counterstained with Gill's hematoxylin (Sigma–Aldrich). Finally, the slides were mounted first with aqueous ImmunoHistoMount (Sigma–Aldrich) and then with organic mounting medium. Slides were scanned with an Aperio ScanScope XT scanner (Aperio Technologies) and imaged at a higher resolution under a Zeiss LSM780 laser scanning confocal microscope.

4.3.5 Polysome profiling (III)

In publication III, polysome profiling was performed for *EPCART*-deletion and WT clones in three biological replicates. Each cell lysate was prepared from two 90% confluent 150 mm plates according to McGlincy et al. (McGlincy & Ingolia, 2017) with minor changes. Briefly, the complete cell growth medium was changed 2 h prior to harvesting, and the cells were returned to the incubator. Then, the plates with cells were placed on ice and washed with ice-cold PBS supplemented with 100 μg/mL cycloheximide (CHX). The PBS was removed, and the plates were floated in liquid N_2 to snap-freeze the cells. While still frozen, freshly prepared lysis buffer (20 mM Tris, pH 7.4; 150 mM NaCl, 10 mM $MgCl_2$, 1% Triton X-100, 1 mM dithiothreitol (DTT), 10 U/mL DNase I, 100 μg/mL CHX) was dripped onto each plate. The cells were scraped from the plates and allowed to slowly thaw on ice. The lysates were triturated through a 26G needle and clarified by centrifugation for 10 min at 10 000 × g and +4 °C. The RNA concentration in the lysates was measured with a Qubit Broad Range kit (Thermo Fisher Scientific). Then, 150 μg of lysate was layered onto a 10–50% sucrose gradient prepared in polysome buffer (20 mM Tris, pH 7.4; 150 mM NaCl, 10 mM MgCl2, 1 mM DTT, 100 μg/mL cycloheximide) and centrifuged at 35 000 rpm (209 627.4 × g) for 3 h at +4 °C in a TH641 rotor (Sorvall). The gradients were fractionated into 15× 750 μL fractions using an automated piston fractionator (Biocomp) with a dual-wavelength A260/280 detection flow cell.

4.4 Protein expression analysis

4.4.1 Protein extraction and Western blotting (II, III)

In publication II, siRNA knockdown was analyzed at the protein level by Western blotting. The cells were lysed in Triton-X lysis buffer supplemented with 1 mM DTT

and 1× Halt protease inhibitor cocktail (Thermo Scientific). The lysates were sonicated four times for 30 s at medium power with Bioruptor equipment (Diagenode), and the cellular debris was removed by centrifugation. Proteins were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and transferred to polyvinylidene difluoride membranes (Immobilon-P, Millipore). Primary antibodies against protein targets in Table 3 were used and detected by anti-mouse horseradish peroxidase (HRP)-conjugated antibody produced in rabbits (DAKO) or by anti-rabbit HRP-conjugated antibody produced in swine (DAKO) and Clarity Western ECL Substrate (Bio-Rad) or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) with a ChemiDoc MP Imaging System (Bio-Rad).

Table 3. List of primary antibodies used

Target	Clone	Clonality	Host	Manufacturer
AR	AR-441	Monoclonal	Mouse	NeoMarkers
ERG	EPR3864	Monoclonal	Rabbit	Abcam
PDCD4	EPR3431	Monoclonal	Rabbit	Abcam
phospho-PDCD4 (Ser67)		Polyclonal	Rabbit	Abcam
p70 S6K		Polyclonal	Rabbit	Cell Signaling Technologies
Phospho-p70 S6K (Thr389)	108D2	Monoclonal	Rabbit	Cell Signaling Technologies
4E-BP1	53H11	Monoclonal	Rabbit	Cell Signaling Technologies
Phospho-4E-BP1 (Ser65)	174A9	Monoclonal	Rabbit	Cell Signaling Technologies
mTOR		Polyclonal	Rabbit	Cell Signaling Technologies
Phospho-mTOR (Ser2448)	D9C2	Monoclonal	Rabbit	Cell Signaling Technologies
Raptor	24C12	Monoclonal	Rabbit	Cell Signaling Technologies
mLST8 (/GβL)	86B8	Monoclonal	Rabbit	Cell Signaling Technologies
PRAS40	D23C7	Monoclonal	Rabbit	Cell Signaling Technologies
Phospho-PRAS40 (Thr246)	C77D7	Monoclonal	Rabbit	Cell Signaling Technologies
Deptor	EPR26159-220	Monoclonal	Rabbit	Abcam
Akt		Polyclonal	Rabbit	Cell Signaling Technologies
p-Akt1/2/3 (Ser473/474/472)		Polyclonal	Rabbit	Santa Cruz Biotechnology
LC3B		Polyclonal	Rabbit	Cell Signaling Technologies
β-tubulin I	SAP.4G5	Monoclonal	Mouse	Sigma–Aldrich
Pan-actin	ACTN05	Monoclonal	Mouse	Thermo Scientific
Fibrillarin	C13C3	Monoclonal	Rabbit	Cell Signaling Technologies

4.4.2 Mass spectrometry (III)

Five replicate samples from each *EPCART* deletion and WT clone were prepared for mass spectrometry (MS) analysis. The cell pellets (approximately 1 x 10^6 cells/sample) were mixed with cold RIPA lysis buffer supplemented with 1% Halt protease inhibitor cocktail (Thermo Scientific), lysed via ultrasonication for 5 min and incubated for 25 min on ice. The clarified supernatant of the cell lysate was collected by centrifugation, avoiding the cell debris, and the total protein concentrations were measured with a Bio-Rad DC Protein Assay (Bio-Rad). Fifty micrograms of protein was precipitated with cold acetone, and the precipitate was collected by centrifugation and dissolved in 2% SDS (Sigma–Aldrich) in 50 mM triethylammonium bicarbonate (TEAB) (Honeywell). The protein cysteine disulfide bonds were reduced with a reducing agent to a final concentration of 3 mM tris-(2-carboxyethyl)-phosphine (Sigma–Aldrich), and the mixture was incubated for 1 h at +60 °C. The samples were transferred to 30 kDa molecular weight cutoff filters (Pall Laboratory) and flushed with 8 M urea in 50 mM Tris-HCl (Sigma–Aldrich), and subsequent alkylation of the free reduced cysteine thiols was performed by incubation in the dark for 20 min to a final concentration of 50 mM iodoacetamide (Sigma–Aldrich). The protein samples were rinsed multiple times with aliquots of 8 M urea buffer and 50 mM TEAB, after which TPCK-treated trypsin (Sciex; trypsin to protein ratio 1:25) was used to digest the proteins for 16 h at +37 °C. After multiple rinses with aliquots of 50 mM TEAB, peptides were eluted from the filter with 0.5 M sodium chloride (Sigma–Aldrich) and dried in a vacuum centrifuge. The peptide samples were reconstituted in 0.1% trifluoroacetic acid (TFA) and cleaned and desalted with C18 tips (Thermo Scientific). The tips were washed with 2.5% acetonitrile (ACN) and 0.1% trifluoroacetic acid, and the peptides were eluted from the tips with 80% ACN and 0.1% formic acid (FA) and dried in a vacuum centrifuge for future use. For the MS analysis, the peptide samples were resuspended in 2% ACN or 0.1% FA to a 1.5 μg/μL concentration.

A Quadrupole Time-of-Flight TripleTOF5600 mass spectrometer (AB Sciex) coupled to an Eksigent 425 Nano LC system and an Eksigent Flex cHiPLC system with a Nanospray III electrospray interface (AB Sciex) was used for analysis. Three micrograms of the peptide sample was loaded onto a trap column (cHiPLC® ChromXP C18-CL, 3 μm particle size, 120 Å, 75 μm i.d. × 5 mm), and loading and desalting were carried out with 2% ACN and 0.1% FA at a 2 μl/min flow rate for 10 minutes. Consecutively, the trap column was switched to an in-line column with a reversed-phase analytical nano cHiPLC column (cHiPLC® ChromXP C18-CL, 3

μm particle size, 120 Å, 75 μm i.d. × 15 cm). Peptide separation was performed using a 120-minute gradient of mobile phases A and B, where A was 0.1% FA and 1% ACN in water and B was 0.1% FA in ACN at a 300 nl/min flow rate. The eluted peptides were electrosprayed onto the mass spectrometer via a fused silica emitter (New Objective).

The data-dependent acquisition (DDA) method was used to generate MS data that were used to create a spectral library. All 15 samples were used to generate a spectral library containing 247 249 spectra and 25 144 peptides from 2519 proteins (at a false discovery rate of 1%) by searching against the Swiss-Prot human database (canonical 20 370 genes) using Protein Pilot software 4.5 (AB Sciex). All 15 samples were then rerun again on the same instrument using the same LC conditions, with a data-independent (SWATH) acquisition mode to acquire protein quantification data. Retention time normalization was carried out using 6 peptides for each of the two proteins with the highest scores (HSPD1 and HSPA8). A total of 2083 proteins were quantified (at a false discovery rate of 1%) for each sample after processing against the spectral library using PeakView (AB Sciex) and MarkerView software programs.

Additionally, publicly available MS data were used in publication III. The quantitative proteome data for untreated PCa and localized CRPC from Latonen et al. (Latonen et al., 2018) and for untreated PCa and mCRPC from Iglesias-Gato et al. (Iglesias-Gato et al., 2018) were retrieved from the supplementary publications.

4.4.3 Immunohistochemical analysis (III)

In publication III, the protein expression levels of PDCD4 in primary PC and CRPC tissues were validated by IHC analysis of the TMA samples. IHC staining was performed with a Ventana BenchMark GX IHC/ISH system (Ventana Medical Systems, Roche), an UltraView Universal Dab Detection Kit (Roche), and an anti-PDCD4 antibody (EPR3431, Abcam) at a 1:4000 dilution. Slides were scanned with a NanoZoomer S60 Digital slide scanner (Hamamatsu Photonics) with a 20× objective. Nuclear and cytoplasmic staining intensities of PDCD4 were classified on a scale from 0 to 3, with negative (0), weak (1), moderate (2), or strong (3) staining within cancerous areas. If possible, a minimum of 200 cells was calculated for each sample. The histoscore (H-score) was calculated as H-score = (0 x percentage of cells with no cytoplasmic staining) + (1 x percentage of "1+" cells) + (2 x percentage of "2+" cells) + (3 x percentage of "3+" cells).

4.5 Cell function assays

4.5.1 Validation of transcript structure (I, III)

In publications I and III, the 5' and 3' ends of *PCAT5* and *EPCART* were validated by rapid amplification of cDNA ends (RACE) together with Sanger sequencing. RACE was performed with the GeneRacer Kit (Invitrogen) in publication I and with the SMARTer RACE 5'/3' Kit (Takara Bio) in publication III according to the manufacturers' instructions. RACE PCR products were obtained using the supplied primers and the appropriate gene-specific primers. The products were separated on agarose gels and extracted, and in publication III, the 5' RACE products were cloned and inserted into a pRACE plasmid following the manufacturer's instructions. The RACE products were sequenced bidirectionally by Sanger sequencing using both the kit's universal primers and gene-specific primers. Sequencing was performed using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI 3500xL Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions.

4.5.2 Cell proliferation assay (I, II)

Cell proliferation was measured either by imaging (publication I) or with a cell viability reagent (publication I and II). Brightfield imaging was performed by scanning the wells with an Olympus IX71 (Olympus) microscope by using Surveyor software (Objective Imaging). The area of the attached cells in each well was calculated by ImageJ software (Wayne Rasband, National Institutes of Health, Bethesda, MD) and divided by the mean area of Day 1 for each following day. Alternatively, the cell viability reagent alamarBlue (Thermo Scientific) was used for proliferation measurements according to the manufacturer's instructions. The fluorescence was measured (excitation 570 nm, emission 585 nm) by an EnVision 2104 multilabel reader (Wallac, PerkinElmer). The relative viability was calculated relative to Day 1.

For analysis of proliferation in stable *EPCART*-overexpressing cell models, brightfield images were acquired with a Cell-IQ Automated Imaging and Analysis System (CM Technologies). The area of the attached cells in each well was calculated

by ImageJ software (Wayne Rasband, National Institutes of Health, Bethesda, MD) and divided by the mean area of Day 1 for each following day.

4.5.3 Wound healing assay (I, II)

In publications I and II, cell mobility was assessed by a wound healing assay. Before imaging, a sterile pipette tip was used to scratch a wound on a confluent cell layer. In publication I, imaging was performed with an Olympus IX71 (Olympus) microscope by using Surveyor software (Objective Imaging). Wound closure was analyzed by ImageJ software (Wayne Rasband, National Institutes of Health, Bethesda, MD). In publication II, time-lapse imaging was carried out with a Cell-IQ Automated Imaging and Analysis System (CM Technologies). The Cell-IQ Analyzer was used to analyze wound closure.

4.5.4 Invasion assay (I)

Cell invasion was evaluated in BioCoat Matrigel Invasion Chambers (BD Biosciences) coated with basement membrane matrix. The Matrigel was rehydrated in growth medium for 2 h at 37 °C. After transfection with siRNAs, the cells were harvested, resuspended in 1% FBS, and placed in the upper chamber. The tumor cells in the upper chamber were allowed to migrate into the lower chamber, which contained attractant medium supplemented with 10% FBS and 5 μg/ml fibronectin, for 22 h at 37 °C. Uninvaded cells were removed by wiping them off the top membrane with cotton swabs. The membranes were then fixed with 3.7% formaldehyde for 2 min, permeabilized with 100% methanol for 10 min, and stained with 1% toluidine for 15 min at room temperature for visualization. The cells that invaded the lower surface were imaged under a bright-field microscope and counted.

4.5.5 Colony formation assay (I)

In publication I, adhesion-independent cell proliferation was assessed by a colony formation assay. The base agar containing 1% agar was dissolved in growth medium supplemented with 20% FBS, 2% L-glutamine, and 2% penicillin–streptomycin. A top layer containing 0.7% agar supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin–streptomycin was prepared. Cells transfected with siRNAs were

harvested and suspended in normal growth medium. The top agar was mixed with the suspended cells and transferred to base agar. The base and top agars were covered with normal growth medium and incubated at 37 °C for 15 days. After incubation, colonies were fixed with 3.7% formaldehyde and stained with 0.1% toluidine blue. Excess dye was removed by washing with 10 mM phosphate buffer (pH 7.4).

4.6 Next-generation sequencing data processing and analysis

4.6.1 RNA-seq data (I, II, III)

In publication I, to discover novel transcripts, RNA-seq reads from Tampere PC cohort samples were first aligned against the hg19 human reference genome using TopHat-1.4.0. Aligned RNA-seq reads were combined into pools according to sample histology. Transcriptomes were then built individually for each pool using Cufflinks-1.3.0 and Illumina iGenome transcriptome annotations as a reference to guide transcript assembly. Cufflinks assembles mapped reads into transcribed fragments (transfrags) based on their genomic coordinates. A merged transcriptome was generated by combining all overlapping and proximal sequences. The transfrags were classified as either previously annotated or unannotated transcripts based on reference annotations (UCSC hg19), NCBI build 37.2, Ensembl GRCh37, and Gencode-12e. The resulting novel transcripts were separated into intergenic and intragenic transcripts. Intragenic transcripts were defined as those that overlapped with previously annotated introns but did not match with previously annotated exons. The novel intergenic transcripts were further filtered to exclude sequencing artifacts and transcriptional noise. The differences in the read counts of novel transcripts between the sample groups (BPH, primary PC, and CRPC) were computed using DESeq-1.2.1. The outliers of differentially expressed transcripts were assessed by the Mann–Whitney *U* test. Only transcripts with an adjusted $p<0.001$ from both DESeq and *U* tests and a normalized read count exceeding 500 reads were included in the subsequent analyses. Additionally, previous PCATs (Prensner et al., 2011) were excluded. The transcripts were manually curated to merge the remaining transcripts into transcripts and to identify likely false-positives (*e.g.*, unannotated 3'-UTRs or simple DNA repeats) that had not yet been filtered out. Exon structures, putative isoforms, and strandedness were manually inferred

based on the recurrence of junctions in the paired-end read data coinciding with canonical intron splice site motifs. The read counts were recomputed for the curated transcripts, and the medians-of-ratios were normalized to the transcripts from the RefSeq 37 human transcriptome. Reads per kilobase expression levels were computed for the transcripts.

In publication I, to evaluate the expression of protein-coding genes, RNA-seq reads were first aligned against RefSeq 38 human transcript sequences using Bowtie-2.0.0-beta6. Expression values were normalized across all samples using median-of-ratios normalization. The read counts for a given gene were divided by the total length of the gene's exons (in kilobases) to correct for gene size bias. For hierarchical clustering, log2 expression ratios were calculated for each gene relative to the gene's median expression. To reduce the effect of noise, genes with low expression (<1000 reads) were omitted, similar to stably expressed genes (standard deviation (SD) of log ratios < 1.0). Hierarchical clustering across columns and rows was performed using the L1 distance metric.

In publication I, the expression of novel lncRNAs was assessed in publicly available RNA-seq datasets. RNA-seq data were obtained from 24 normal human tissues (Cabili et al., 2011), 21 normal prostate and PC cell lines (Prensner et al., 2011), H1 human embryonic stem cells (ENCODE, 2012, Nature), 10 normal prostate tissues (Kannan et al., 2011), and 54 PC samples (Kannan et al., 2011; Prensner et al., 2011). These data were re-aligned to hg19 with TopHat as described above, and the read counts for all the novel transcripts were computed.

In publication II, to analyze the expression of novel lncRNAs in TCGA-PRAD samples (The Cancer Genome Atlas Research Network, 2015), RNA-seq data for 400 samples were downloaded from the Genomic Data Commons Data Portal and aligned against the hg19 human reference genome using TopHat-2.1.1. A catalog of gene exons was built by combining the Ensembl 75 splice variants and adding the novel lncRNA 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. The TCGA-PRAD data of novel lncRNAs and more than 3,000 human genes linked to transcriptional regulation from the TFcheckpoint database (Chawla et al., 2013) were compared. The expression values were log2 transformed, and the Pearson correlation coefficient was calculated for each novel lncRNA and TF in a pairwise manner. In publication III, expression data for proteins of interest and clinical patient data in TCGA-PRAD were retrieved from cBioPortal (Cerami et al., 2012). For correlation analysis in the Tampere PC and TCGA-PRAD

cohorts, Spearman's rank correlation coefficient was calculated for *EPCART* and PDCD4 expression in a pairwise manner.

For comparative analysis of novel RNA expression in the Tampere PC cohort and TCGA-PRAD data, unsupervised hierarchal clustering was performed. The differential expression of the sample groups was calculated as the log2-fold change (log2FC) of the median + SD of the normalized RNA-seq expression. Average linkage clustering and Euclidean distance measurement methods were used for the analysis of log2FC values on the Heatmapper web server (http://heatmapper.ca/) (Babicki et al., 2016), which was also used for visualization.

In publication III, the read quality of strand-specific RNA-seq data from *EPCART*-deletion and WT samples was assessed with FastQC v. 0.11.8, and reads were aligned to GRCh38 using STAR v2.71a (Dobin et al., 2013) followed by indexing with SAMtools v1.8 (Li et al., 2009). Read counts of protein-coding transcripts were calculated using BEDTools v. 2.27.1 subcommand multicov (Quinlan & Hall, 2010), and GENCODE v.38 annotation was used for gene calls. Differential expression analysis between *EPCART*-deleted and WT clones was performed using the DESeq2 R package v. 1.22.2 (Love et al., 2014). Differentially expressed protein-coding genes (p value <0.05) were analyzed by Ingenuity Pathway Analysis (IPA, Qiagen). Canonical pathways were filtered only to show signaling pathways for the further analysis of each sample pair (del-4 vs. WT or del-56 vs. WT). A comparative analysis of the Functions and Diseases of the del-4 mutant vs. the WT and the del-56 mutant vs. the WT was performed; only the Molecular and Cellular Functions were filtered for further analysis. For IPA, the p value of overlap, calculated using the right-tailed Fisher's exact test, was used to identify significant pathways. The overall activation/inhibition states of canonical pathways are predicted based on a z score algorithm. Z scores that are greater than or equal to 2 represent predictions of activation, while predictions of inhibition are made for z scores less than or equal to -2. Log p values >1.3 (= p <0.05) were considered significant.

4.6.2 Chromatin immunoprecipitation sequencing data (I, II)

In publication I, chromatin immunoprecipitation sequencing (ChIP-seq) data from the untreated or vehicle-treated cell lines 22Rv1, LNCaP, and VCaP, as well as from the doxycycline-treated C4-2B cell line and metastatic PC tissue samples, were used to confirm transcription factor binding events in the promoters of novel lncRNAs.

The raw data from two published studies (Massie et al., 2011; Yu et al., 2010) were downloaded from the Gene Expression Omnibus (GEO) and aligned to GRCh37 using Bowtie (version 0.12.8.) and the nondefault parameters –v2 and –m20. TF binding sites were inferred with the MACS peak detection algorithm (Zhang et al., 2008). Processed data from two additional studies (Chng et al., 2012; Little et al., 2012) were downloaded from GEO and converted from hg18 to hg19 using LiftOver (Hinrichs et al., 2006) with default parameters.

In publication II, to investigate the binding sites of TFs, ChIP-seq peaks were retrieved from the 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 novel lncRNAs (−15 kb/+2 kb from the 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 histone modification detection in publication I, H3K4me1-3, H3K9me3, H3K36me3, H3K37me3, and RNA polymerase II in VCaP cells under three different conditions were retrieved from GEO (GSE14097). Any MACS-inferred peak region (central coordinate) located at a maximum of 1 kb upstream or 500 bp downstream from the inferred TSSs was considered a binding event to the proximal promoter of the novel lncRNAs.

For the analysis of *EPCART* loci, CTCF ChIP-seq signal peaks from LNCaP and 22Rv1 cells and a prostate gland sample were retrieved from the ENCODE portal (Davis et al., 2018) with the following identifiers: ENCFF805BDU, ENCFF539QXW, and ENCFF406TZC. In addition, histone modifications in LNCaP and PC tissues were retrieved from GEO (GSE148935 and GSE96652, respectively).

4.6.3 Other next-generation sequencing data

The read quality of the WGS data from the *EPCART*-deletion and WT samples was assessed with FastQC v. 0.11.8, and the reads were aligned to GRCh38 using STAR v2.71a (Dobin et al., 2013) followed by indexing with SAMtools v1.8 (Li et al., 2009). The mapped reads were visualized with Integrative Genomics Viewer (Robinson et al., 2011).

For the analysis of *EPCART* loci, DNase-seq data from LNCaP cells were retrieved from the ENCODE portal (Davis et al., 2018) with the following identifier: ENCSR000EPF. In addition, Hi-C data from prostate cell lines and PC tumors were retrieved from the GEO (GSE118629 and GSE164347, respectively). TE repeats were visualized with RepeatMasker (Smith et al., 2013) in Integrative Genomics Viewer (Robinson et al., 2011).

4.7 Statistical methods (I, II, III)

Mann–Whitney *U* tests were used to analyze the associations between two sample groups that were not normally distributed (publications II and III). Unpaired two-tailed Student's t tests were used to calculate the significance between the control and experimental conditions in terms of PCR, cell viability, wound healing experiments, and immunoblot quantification (publications I, II, and III). P values <0.05 were considered to indicate statistical significance.

Kaplan–Meier survival analysis and log-rank tests were used to compare biochemical progression-free survival between sample groups. 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, PSA levels, and target gene transcript expression. Age at diagnosis was incorporated into the regression model as a continuous covariate. In publication II, the expression level of each novel lncRNA transcript was divided into either low or high groups using the gene's median ΔCt expression value as a baseline. Similarly, pathologic T status was categorized as either low (1 to 2) or high (3 to 4). Gleason scores were divided into three groups: low (< 7), intermediate (= 7) and high (8 to 10). The diagnostic PSA values were also divided into three groups: low (< 10), intermediate (10 to 19.9) and high (> 20). Cox regression analysis was performed using the coxph function from the survival package version 2.41-3 in R. P values <0.05 were considered to indicate statistical significance in both analyses. In publication III, samples were divided by the first quartile expression of the target gene.

In publication III, for the differential analysis of MS data from *EPCART*-deletion and WT clones, protein quantification data were log2-transformed, and the replicate MS analyses were combined. The coefficient of variation (CV) was calculated for samples originating from the same sample type and passage to identify and exclude quantified proteins with poor repeatability (CV ≥ 30%). Due to the small number of

samples processed, only descriptive analysis and results are reported for individual proteins. These include means by sample types and the associated log2FC between different sample types. Proteins with log2FC > log2(1.5) and < log2(0.67) were included in the pathway analyses. R software (v4.1.2, R Core Team) was used to process the data and perform the descriptive analyses.

5 RESULTS

5.1 Identification of novel prostate cancer-associated transcripts (I)

By the early 2010s, only a few genome-wide transcriptomics studies had been carried out to identify PC-associated lncRNAs (Du et al., 2013; Prensner et al., 2011; Ren et al., 2012). To explore the transcriptomes of primary PCs and CRPCs more profoundly, we performed deep RNA-seq on Tampere PC cohort specimens (12 BPHs, 28 untreated PCs, and 13 locally recurrent CRPCs). In particular, we aimed to identify unannotated lncRNA transcripts that were differentially expressed between benign prostate and primary PC or between primary PC and CPRC. In total, we discovered 145 novel intergenic PC-associated transcripts or isoforms that were named TPCATs (publication I). Of these, 12 had more than one isoform, resulting in 128 TPCATs with one or more isoforms. The majority of the TPCATs had only one exon (78 out of 145; publication I). In addition, there were more TPCATs that were overexpressed than downregulated in primary PC or in CRPC than in BPH, and on average, the expression of TPCATs was more abundant in CRPC than in primary PC (Figure 10). Among the 145 TPCATs, 69% and 86% were overexpressed (log2FC of median + SD >1) in 69% and 86% of the TPCATs, respectively, in primary PC and CRPC (Figure 10). The expression of TPCATs was also examined in TCGA-PRAD data, where a similar majority of TPCATs (78%) were overexpressed in primary PC vs. adjacent normal samples (Figure 10). The differential expression of TPCATs between primary PC and CRPC in the Tampere PC cohort was also significantly altered (log2FC >1 or <-1) for several TPCATs; over 50% of TPCATs were more abundantly expressed in CRPC than in PC, and approximately half of those TPCATs were overexpressed only in CRPC (Figure 10). The expression of TPCATs was validated in RNA-seq datasets of normal tissues and PC cell lines and tumors (publication I). In normal tissues, the expression of TPCATs was most commonly detected in the testes, albeit generally at low levels (publication I). In PC tissue samples of the validation cohort, the expression profiles of TPCATs were similar to those in the Tampere PC cohort (publication I).

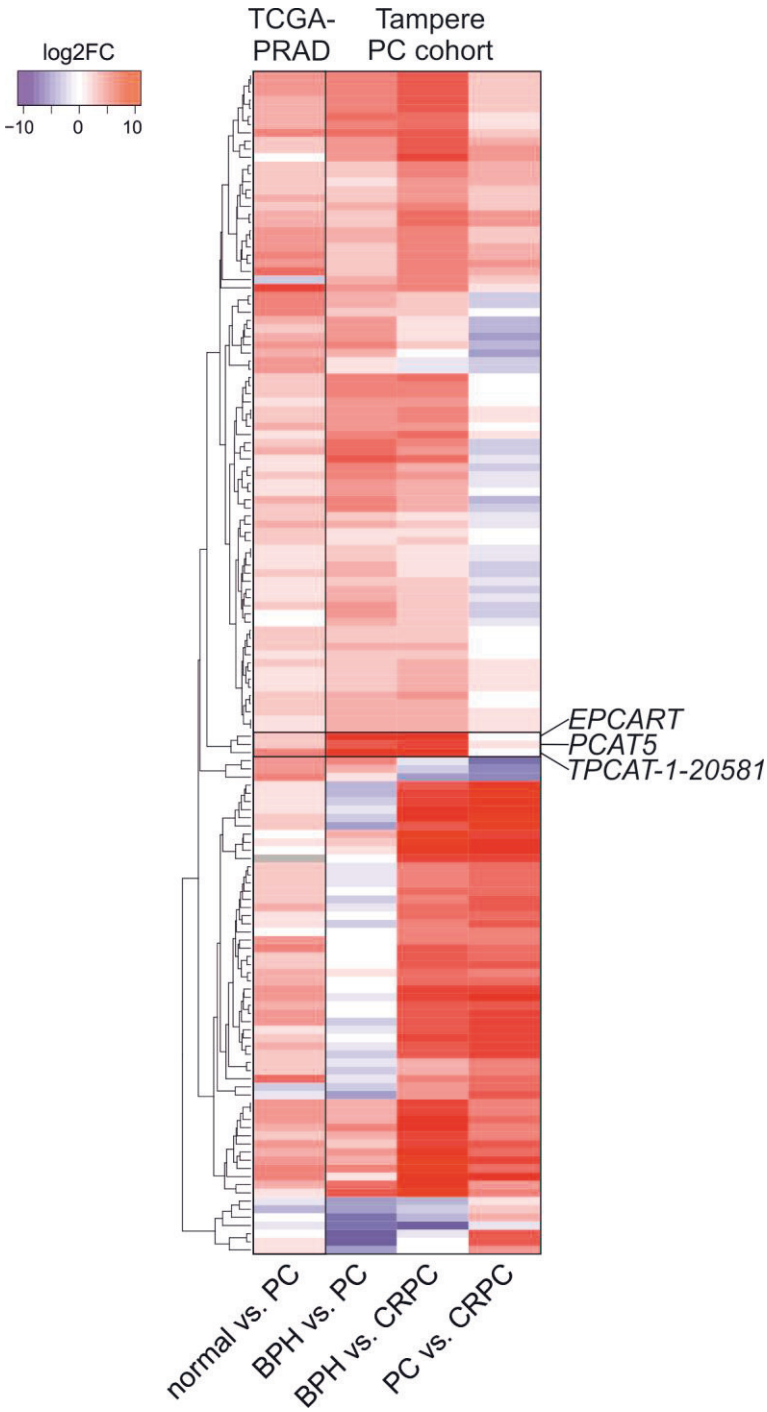


Figure 10. Differential expression of TPCATs. For each TPCAT (n=145), the median ± SD of the normalized RNA-seq expression was used for the calculation of the log2FC values.

5.2 TPCATs as biomarkers in prostate cancer (II, III)

As *PCA3* is an approved lncRNA marker used for assisting in the diagnosis of PC (Hologic, 2019) and a few lncRNAs have been suggested as PC-specific prognostic markers (Prensner et al., 2013; Shukla et al., 2016), we wanted to assess the potential of TPCATs as biomarkers in PC. Robust and tissue-specific expression data are good qualities for diagnostic biomarkers. The three most differentially overexpressed TPCATs in the Tampere PC cohort (BPH vs. primary PC) were *TPCAT-1-20581*, *TPCAT-2-180961*, later renamed ERG-positive PC-associated androgen responsive transcript (*EPCART*), and *TPCAT-10-36067*, officially named *PCAT5* (Figure 10). Among these TPCATs, *EPCART* had the highest expression (Figure 11). Both *PCAT5* and *EPCART* were also highly PC-specific in other cohorts, regardless of whether their expression was compared between PC and benign prostate samples or between prostate and tissue samples from other sources, either malignant or benign (Figures 12-13).

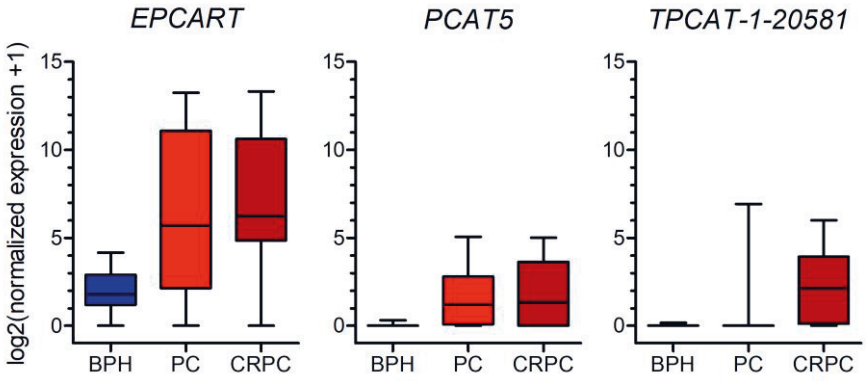


Figure 11. Expression of the most differentially overexpressed TPCATs in the Tampere PC cohort.

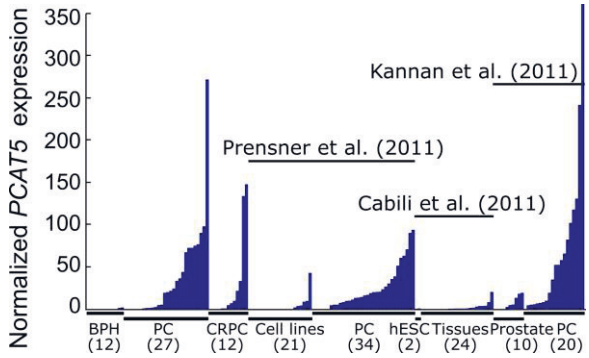


Figure 12. *PCAT5* expression in various PC cancer cohorts and in normal tissues. Retrieved from publication I.

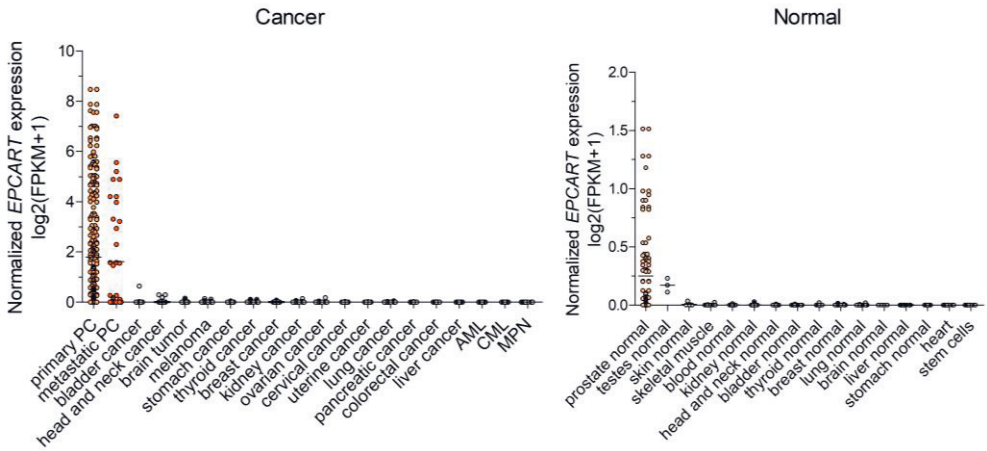


Figure 13. *EPCART* expression in various cancer and normal tissues. The data were retrieved from the MiTranscriptome catalog. The arithmetic mean expression of each group is marked with a black line. Modified from publication III.

For proteins, IHC staining is commonly used to detect protein expression in tissue samples, *e.g.,* in biopsies. For RNAs, antibody-based detection methods are not feasible, but RNA-ISH methods have been successfully used to detect the expression of certain lncRNAs in tissue sections (Mehra et al., 2014; Shukla et al., 2016). To test whether RNA-ISH could be used to detect TPCATs in FFPE sections of PC tissues, we performed RNA-ISH experiments on *EPCART* (publication III). We tested two commercially available RNA-ISH assays, RNAscope (Advanced Cell Diagnostics) and ViewRNA (Affymetrix), both of which are based on the use of several adjacent double-Z probe pairs that bind to the target RNA and multiple signal amplifiers that

bind to the probes or other amplifiers. Both methods were able to detect *EPCART* expression in FFPE samples of PC tissues (Figure 14), but the detected signal was not strong enough for the quantitative detection of *EPCART* in FFPE samples. Nevertheless, we found that *EPCART* was specifically expressed in prostate adenocarcinoma and not in adjacent benign prostate or stroma (Figure 14).

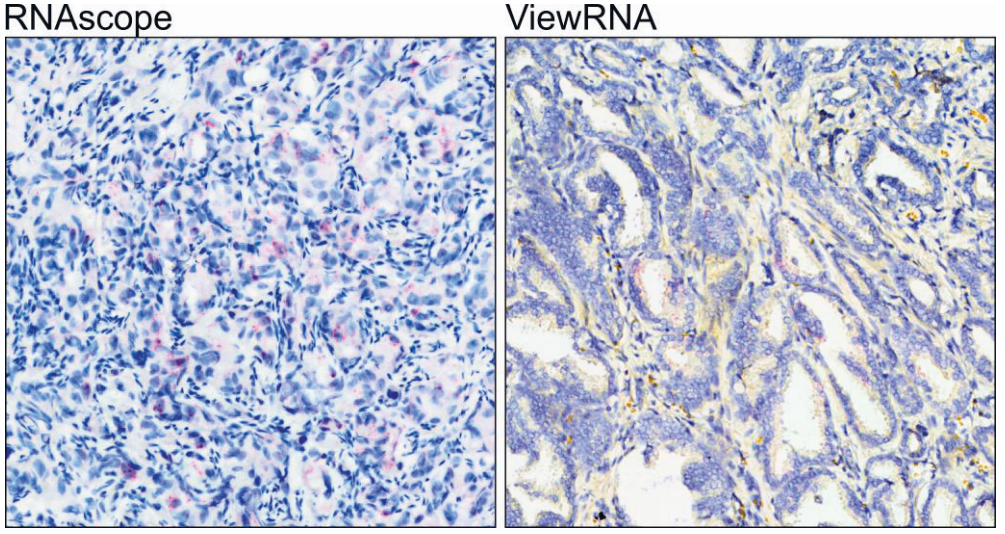


Figure 14. RNA *in situ* hybridization of *EPCART* in PC FFPE tissue specimens. Staining by two ISH kits (RNAscope and ViewRNA) is shown. *EPCART* was labeled with red dye, and the slides were counterstained with hematoxylin. The right image was modified from publication III.

To investigate whether TPCATs have the potential to be used as prognostic markers, we compared high-throughput qRT–PCR data from 87 untreated PC samples and 34 TPCATs to clinicopathological data collected from the same samples over 10 years (publication II). We used biochemical recurrence time to calculate progression-free survival in samples divided into high and low TPCAT expression groups based on the median expression of each TPCAT (publication II). Kaplan–Meier analysis revealed three TPCATs that were significantly different between the assessed groups, namely, *EPCART*, *TPCAT-3-174133*, and *TPCAT-18-31849* (Figure 15). To test whether these TPCATs had independent prognostic value, we compared their expression levels to well-known prognostic parameters: age, Gleason score, diagnostic PSA, and pT stage. Multivariate Cox regression analysis revealed that all three TPCATs were independently associated with PC progression (publication II).

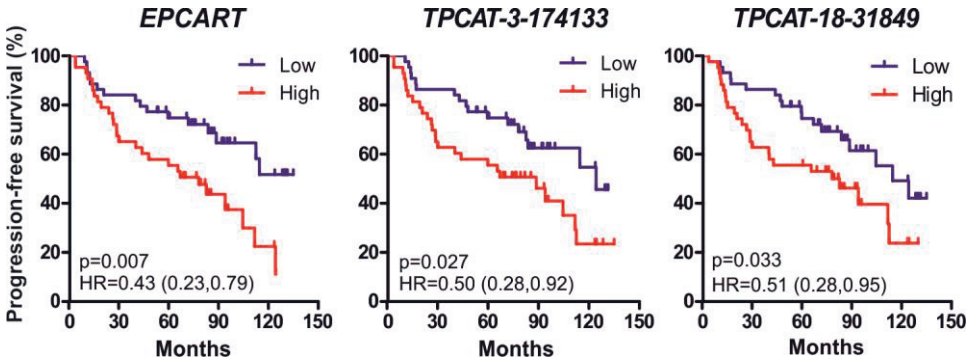


Figure 15. Prognostic TPCATs. Kaplan–Meier analysis was performed for the progression-free survival of patients with PC grouped based on the median expression of each TPCAT. Modified from publication II.

Interestingly, *EPCART* was also found to be strongly associated with the expression of ERG (publication II), even though previous studies have shown that ERG expression or *TMRSS2-ERG* fusion does not have prognostic value in PC (Pettersson et al., 2012; Song & Chen, 2018). To unravel this contradiction, we divided the 87 untreated PC samples into four groups according to their ERG status and median *EPCART* expression: ERG-negative + *EPCART*-low, ERG-positive + *EPCART*-low, ERG-negative + *EPCART*-high, and ERG-positive + *EPCART*-high. Kaplan–Meier analysis revealed that the worst prognosis was in patients in the ERG-negative + *EPCART*-high subgroup, although patients with high *EPCART* expression in ERG-positive samples still had a worse prognosis than those with low *EPCART* expression (Figure 16). These findings indicated that high *EPCART* expression in any PC sample leads to earlier progression, especially in ERG-negative tumors.

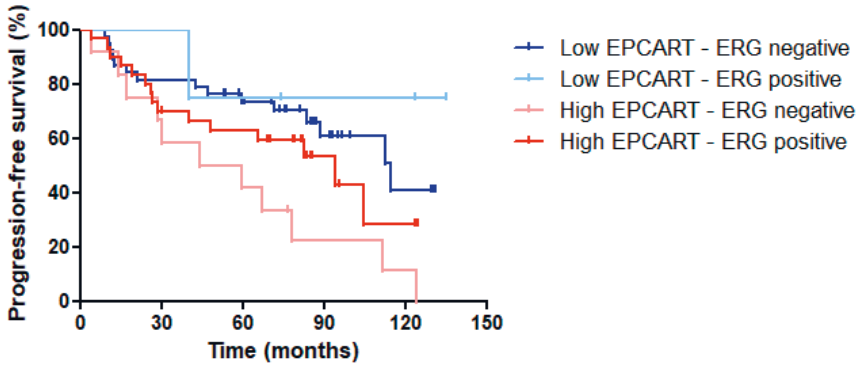


Figure 16. Association of *EPCART* expression and ERG status with PC progression. Kaplan–Meier analysis was performed for progression-free survival of PC patients grouped based on the median expression of *EPCART* and/or ERG status.

5.3 Regulation of TPCATs in prostate cancer (I, II, III)

Very little information exists on what causes the aberrant expression of lncRNAs in PC. To investigate whether copy number alterations or changes in DNA methylation could explain the differences in the expression levels of TPCATs, we integrated RNA-seq expression data with DNA sequencing and methylated DNA immunoprecipitation sequencing data from Tampere PC cohort samples (publication I). Correlation analysis between TPCAT expression and copy number or DNA methylation of the nearby region did not reveal any significant correlation. In addition, there was no significant difference in the differential expression of TPCATs between samples with normal and abnormal copy numbers. These findings suggest that the aberrant expression of TPCATs cannot be explained by genetic alterations or changes in DNA methylation.

We hypothesized that aberrant transcriptional regulation could explain the differential expression of TPCATs. To determine which TFs are responsible for the regulation of TPCAT expression, we calculated the correlations in the Tampere PC cohort between the expression of TPCATs and eight TFs (*ERG*, *AR*, *FOXA1*, *EZH2*, *HDAC1*, *HDAC2*, *HDAC3*, and *RUNX2*), for which public ChIP-seq data in PC cell lines were available for validation at the time (publication I). In addition, we performed a correlation analysis of TCGA-PRAD data on the expression levels of more than 3,000 genes linked with transcription and TPCATs (publication II). In both datasets, the strongest positive correlations were observed between *ERG* and TPCATs, and in the TCGA-PRAD data, 10 out of 34 overexpressed TPCATs had

a moderate or strong positive correlation (r >0.4) with *ERG* expression, the strongest being *PCAT5* and *EPCART*.

The association between TFs and TPCATs was further investigated in publication II, in which we studied the regulation of 34 TPCATs in greater depth. These TPCATs were all overexpressed in untreated PCs and had more than one exon. We analyzed the expression levels of these TPCATs and selected PC-associated TFs (*ERG*, *ETV1*, *AR*, and *FOXA1*) in 87 untreated PC samples by high-throughput qRT–PCR. Hierarchical clustering of these samples revealed that TPCAT expression was grouped into *ERG*-high and *ERG*-low clusters, indicating that several TPCATs were associated with the overexpression of ERG. Next, we divided the samples into *ERG*-positive and *ERG*-negative groups and compared the expression of each TPCAT between these groups. Moreover, we repeated the analysis of the TCGA-PRAD data. In total, we found that the majority of TPCATs (22 out of 34) were associated with the expression of *ERG* (Figure 17). To investigate whether ERG itself can take part in the regulation of TPCATs, we examined the ERG binding sites in the putative promoter (−15 kb/+2 kb from TSS) of TPCATs in publicly available ChIP-seq data from ERG-positive VCaP cells (publication II). Indeed, we found that ERG bound to these regions in the majority of TPCATs (20 out of 34), especially among ERG-associated TPCATs (16 out of 22) (Figure 17). Finally, we validated ERG dependency by ERG siRNA knockdown studies in ERG-positive cells (Figure 17, publication II). The expression of almost half of the TPCATs (16 out of 34) was ERG dependent (log2 FC <-1 or >1), ten of which were ERG-associated TPCATs. Additionally, we performed ETV1 siRNA-mediated knockdown studies in LNCaP cells and detected similar regulatory effects (Figure 17). Taken together, these findings indicate that ERG and other ETS TFs are likely to regulate several TPCATs in PC.

In PC, abnormal AR binding is considered to play a role in cancer initiation (Pomerantz et al., 2015). To assess whether AR also participates in the regulation of TPCATs, we utilized publicly available AR ChIP-seq data from untreated PCs and normal adjacent tissues. Similar to ERG, we counted the AR binding sites (ARBSs) in the putative promoters (−15 kb/+2 kb from the TSS) of TPCATs (publication II). We detected more than six times more ARBSs in PC than in normal tissue in the promoter regions of TPCATs, and nearly 70% of the TPCATs had ARBSs in PC (publication II). The effect of AR knockdown on the AR dependency of TPCAT expression was further investigated in AR-positive cells (Figure 17). In addition, whether TPCAT expression was stimulated by androgens was tested by DHT stimulation in the same cell lines (Figure 17). Over half of the TPCATs were strongly

affected (log2FC <−1 or >1) by either AR knockdown (21 out of 34) or DHT stimulation (19 out of 34). Only seven TPCATs were affected by both agents in opposite ways, although weaker effects were observed for several additional TPCATs (Figure 17).

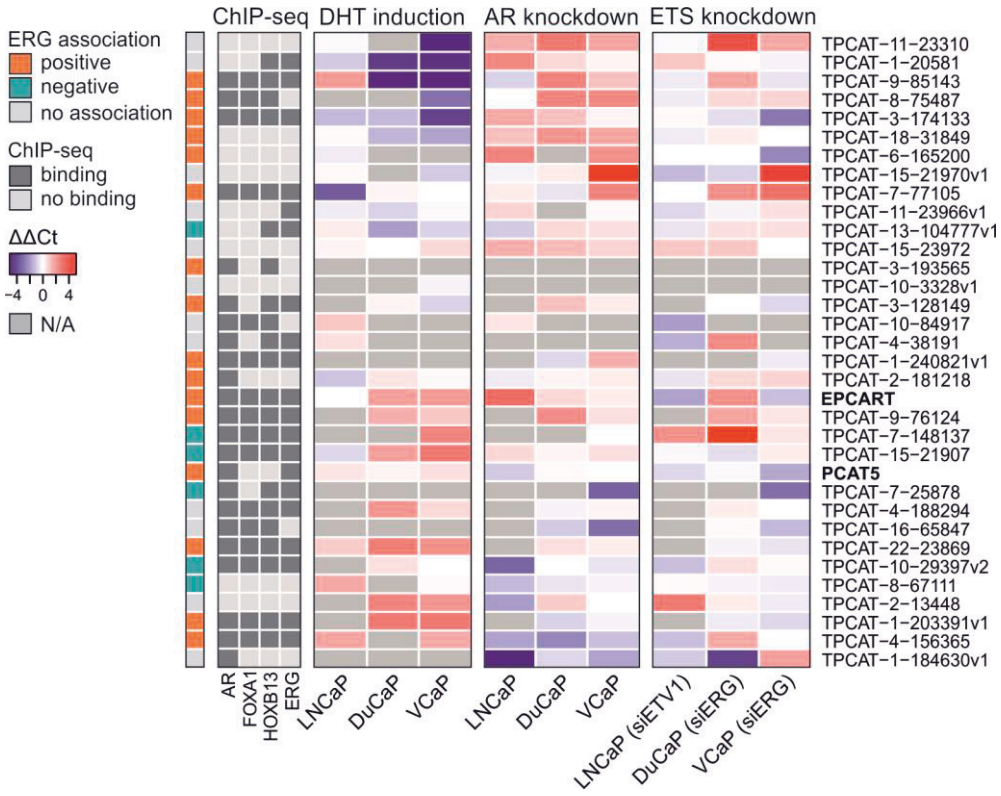


Figure 17. TPCAT regulation in PC cells. Modified from publication II.

FOXA1 and HOXB13 have also been found to facilitate AR binding to novel sites in PC (Pomerantz et al., 2015). They also colocalize to the same sites as AR and ERG (Kron et al., 2017; Pomerantz et al., 2015). To explore whether FOXA1 and HOXB13 also bind to the same sites on the putative TPCAT promoters as AR and ERG, we analyzed the colocalization of these TFs in the promoter regions (−15 kb/+2 kb from the TSS) of the TPCATs (Figure 17). In total, 25% of the identified binding sites (15 out of 61) were cooccupied by all four TFs, which was significantly greater than the percentage detected globally (7%) (publication II). These co-occupied binding sites were found in 38% of the TPCATs (13 out of 34). In addition, 56% of TPCATs had at least one binding site cooccupied by three or more TFs and

68% by two or more TFs; only 24% had no binding sites detected for any of the four TFs. These findings indicated that all four TSs are involved in the regulation of most TPCATs.

5.3.1 Regulation of *PCAT5* (I, II)

PCAT5 is located on chromosome 10p11.21. Sequence analysis revealed a TATA box at close proximity to the *PCAT5* TSS and a polyadenylation signal at the 3' end of the transcript (Figure 18). Publicly available ChIP-seq data revealed that open chromatin histone markers (H3K4me3) and RNA polymerase II (PolII) bind to the putative promoter of *PCAT5* in VCaP cells, indicating active transcription (Figure 18). Additionally, ChIP-seq data revealed ERG and AR binding sites (Figures 17 and 18) and H3K27 acetylation, a marker for active transcription, exclusively in *TMPRSS2-ERG*-positive PC tumors (Figure 19) at the proximal promoter of *PCAT5*.

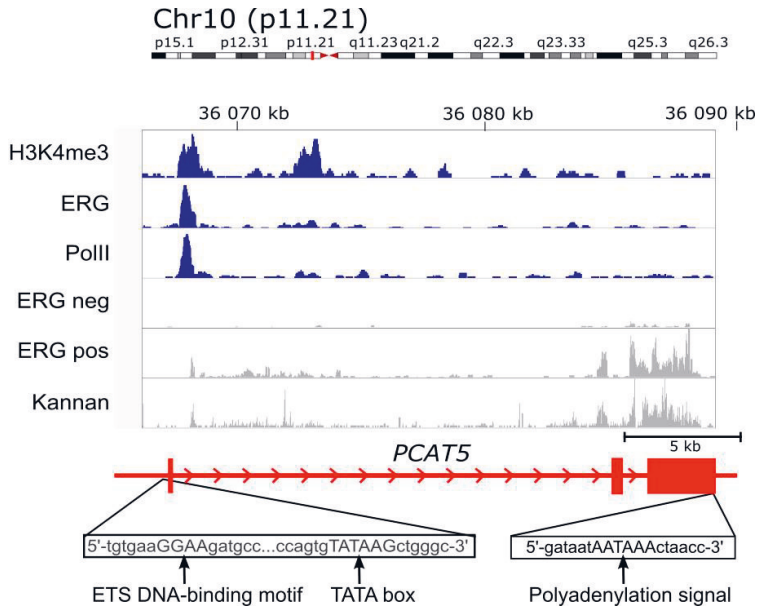


Figure 18. Regulation of the *PCAT5* locus. Retrieved from publication I.

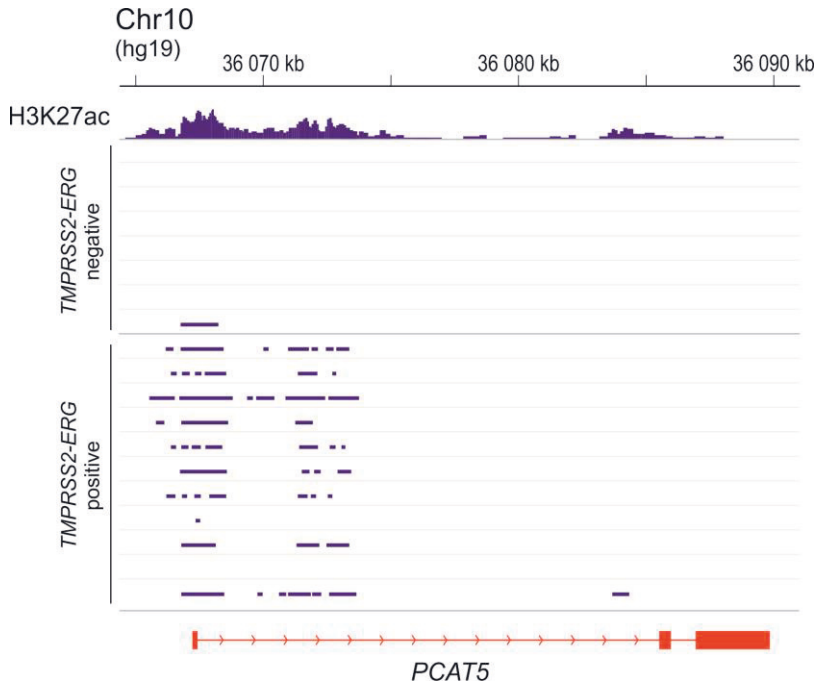


Figure 19. *TMPRSS2-ERG*-positive regulation of *PCAT5*. ChIP-seq peaks of H3K27ac in *TMPRSS2-ERG*-negative and TMPRSS2-ERG-positive PC tumors. The H3K27ac signal in one *TMPRSS2-ERG*-positive tumor is shown at the top.

In publication I, we found that *PCAT5* was substantially regulated by ERG. Its correlation with *ERG* expression was high in the Tampere PC cohort (r = 0.69). In publication II, we also detected a strong correlation between *PCAT5* and *ERG* expression in TCGA-PRAD data (r = 0.84) and significant associations with ERG-positive samples in both the Tampere PC cohort and TCGA-PRAD data. The ERG dependency of *PCAT5* was further analyzed in siRNA knockdown studies (publication I). *PCAT5* is abundantly expressed not only in the VCaP cell line, which endogenously overexpresses ERG but also in the PC-3 cell line, which overexpresses ETV4, which belongs to the same ETS family of TFs as ERG. Knockdown of ETV4 in PC-3 cells and ERG knockdown in VCaP cells both led to statistically significant (p <0.05) reductions in *PCAT5* expression (publication I). These findings indicate that in addition to being *PCAT5* ERG-regulated, other ETS family TFs can also participate in this regulation. In publication II, we detected AR binding to the promoter region of *PCAT5* in a fraction of the PC samples but not in any of the normal prostate samples. In addition, AR knockdown caused a reduction in *PCAT5*

in LNCaP cells, and DHT stimulation caused an increase in *PCAT5* expression, although the effects were not strong (log2 FC <1 or >-1).

5.3.2 Regulation of *EPCART* (II, III)

EPCART is located on chromosome 2q31.3. It has very few protein-coding genes in its close vicinity, but there are other PCATs, including *SChLAP1* (Figure 20). Several publicly available datasets were used to perform a detailed investigation of the *EPCART* locus. Interestingly, according to Hi-C data from prostate cell lines (RWPE-1 model normal prostate cells and LNCaP C4-2B and 22Rv1 PC cells), *EPCART* and *SChLAP1* also reside in the same topologically associated domain (TAD) (Figure 20). DNase-seq in PC cells revealed open chromatin sites at the proximal promoter of *EPCART* and at the intron between *EPCART* exons 2 and 3 (Figure 20). By using ChIP-seq data from PC tissues, we found that the *EPCART* promoter region was cooccupied by several TFs, including AR, ERG, FOXA1, and HOXB13 (Figure 20). The intronic open chromatin site in *EPCART* was found to be bound by CTCF (CCCTC-binding factor) (Figure 20), which could serve as a TAD boundary. This TAD boundary was further validated by Hi-C chromatin interaction data from PC tumors (Figure 20). In the proximal promoter area, sequence analysis did not reveal a canonical TATA box motif, but we discovered a long interspersed nuclear element type 2 (LINE-2) retrotransposon that overlapped with exon 1 of *EPCART*, in addition to overlapping with multiple TF binding sites (Figure 20). Furthermore, other TEs were also found inside the exonic regions of *EPCART* (Figure 20). We also examined histone modifications around the *EPCART* genomic region from publicly available ChIP-seq data and did not detect many modified histone markers, except H3K27me3, in LNCaP cells (Figure 21). However, H3K27ac data from PC tumor samples (Kron et al., 2017) revealed significant histone acetylation in the promoter of *EPCART*, which was not evident in LNCaP cells, and a *TMPRSS2-ERG*-positive CRE close to the *EPCART* promoter (Figure 21).

In publication II, we found *EPCART* to be strongly associated with ERG. According to the TCGA-PRAD data, the correlation between *ERG* and *EPCART* expression was 0.63 (publication II). Furthermore, *EPCART* expression levels were significantly greater in ERG-positive samples than in ERG-negative samples in both the Tampere PC cohort and the TCGA-PRAD cohort (publication II). We also detected an ERG binding site at the proximal promoter of *EPCART* in publicly

available ChIP-seq data from VCaP cells (publication II). However, siRNA-mediated knockdown of ERG in VCaP or DuCaP cells did not significantly reduce the expression of *EPCART* (Figure 17, indicating that other regulatory factors might play a more significant role in the regulation of *EPCART* expression. Indeed, publicly available ChIP-seq data revealed the colocalization of AR, ERG, FOXA1, and/or HOXB13 at three sites in the regulatory region of *EPCART*. AR binding was especially significant, as AR binding to ARBSs was much stronger in PC samples than in normal prostate samples (Figure 21). Moreover, knockdown of AR caused a statistically significant reduction in *EPCART* expression in DuCaP cells, and DHT induction had the opposite effect on the same cell line (publication II). To assess AR binding to the closest ARBS of the *EPCART* TSS, we conducted ChIP–PCR experiments in LNCaP cells overexpressing AR and in PC xenografts (publication II). We detected increased binding of AR to the *EPCART* promoter in LNCaP cells overexpressing AR and in xenografts with AR amplification. In addition, DHT treatment of LNCaP cells significantly increased AR binding to the *EPCART* promoter.

In publication III, high *EPCART* expression was also found to be associated with both heterozygous (hetloss) and homozygous (homdel) *PTEN* deletions in TCGA-PRAD data. As both *PTEN* loss and high *EPCART* expression are associated with *TMPRSS2-ERG* fusion, we investigated the expression of *EPCART* in samples with both *PTEN* deletion and *TMPRSS2-ERG* fusion (Figure 22A). We detected a significant (p <0.05) association between *PTEN* loss and high *EPCART* expression, indicating that *ERG* overexpression is not the cause of the observed association. Additionally, the association of *EPCART* with other common genetic aberrations was also studied in the TCGA-PRAD cohort. The only other genetic aberration that was significantly (p <0.05) associated with *EPCART* expression was *TP53* loss; PC samples with high *EPCART* expression frequently exhibited *TP53* deletion or inactivating mutations (Figure 22B).

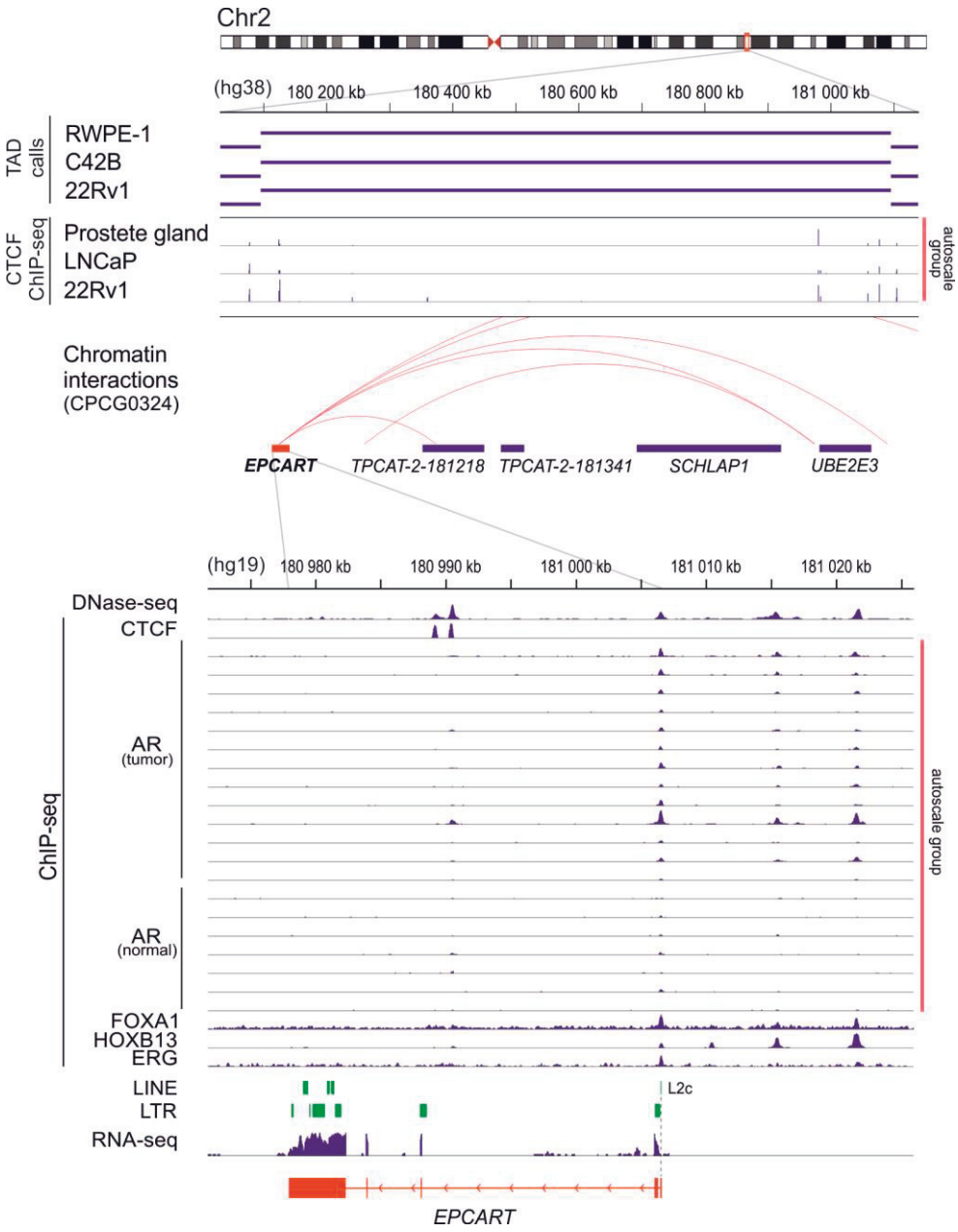


Figure 20. *EPCART* locus. ChIP-seq (CTCF, TFs, and Pol II), Hi-C (TAD and chromatin interaction calls), DNase-seq, and RNA-seq data from normal prostate or PC cells or tissues are shown. Only TEs detected inside *EPCART* exons are shown in green. Modified from publication II.

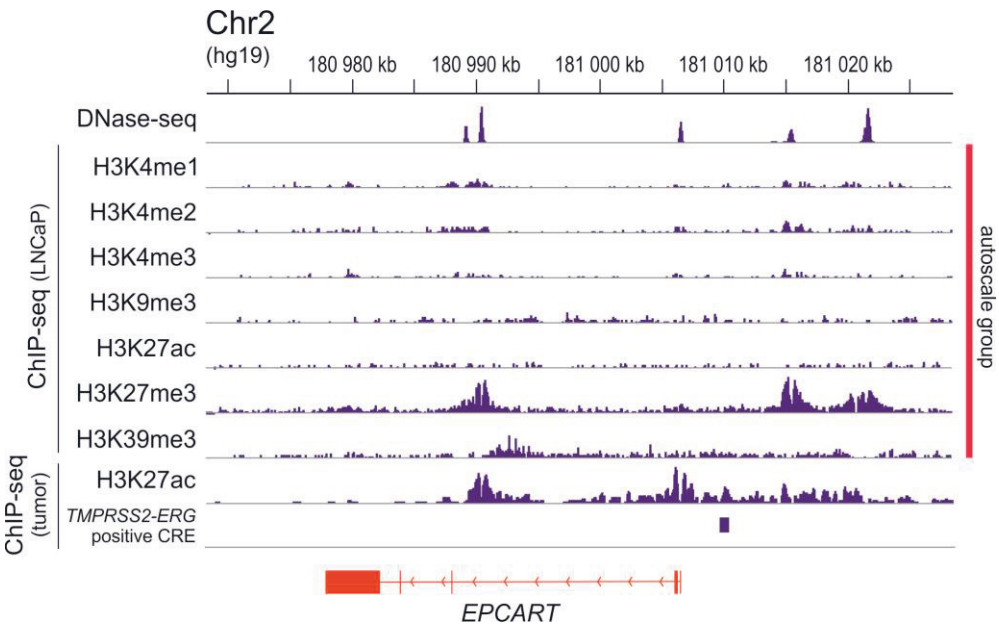


Figure 21. Histon modifications at the *EPCART* locus. DNase-seq (from LNCaP) and different histone modifications (LNCaP under DHT stimulation and one tumor sample) are shown. Based on H3K27ac data from PC tumors, one *TMPRSS2-ERG*-positive CRE was found near the *EPCART* promoter.

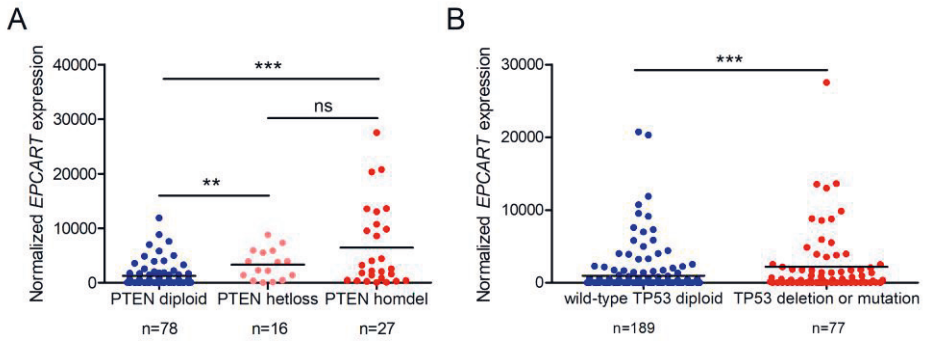


Figure 22. Association of *EPCART* with common genetic aberrations in PC. A) *EPCART* expression in TCGA-PRAD primary PC samples with *PTEN* deletion (heterozygous or homozygous) and the *TMPRSS2-ERG* fusion. B) *EPCART* expression in TCGA-PRAD primary PC samples with *TP53* loss by deletion or inactivating mutation. The arithmetic mean expression of each group is marked with a black line. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; the data were assessed with the Mann–Whitney *U* test.

5.4 Functional characterization of TPCATs (I, II, III)

Because *PCAT5* and *EPCART* are both highly expressed in PC tumors and regulated by PC-specific TFs and because *EPCART* is also independently associated with PC progression, we selected these lncRNAs for further studies to better understand their function and relevance in PC cells.

5.4.1 Transcript structures of *PCAT5* and *EPCART* (I, III)

The alignment of RNA-seq reads from Tampere PC cohort samples to the human genome revealed novel splice junctions of TPCATs (publication I). We validated these splice junctions for *PCAT5* and *EPCART* transcripts by RT–PCR and Sanger sequencing in clinical samples for *PCAT5* (3 exons) in publication I and for *EPCART* (5 exons) in publication III. In addition, we were able to validate the 5' and 3' ends of both *PCAT5* and *EPCART* by RACE (publications I and III, respectively). In the case of *EPCART*, we detected prominent variation at the 5' end, including an alternative exon (publication III). ORF analyses suggested that both TPCATs lack protein-coding potential (publications I and III). In the case of *EPCART*, a short ORF was found, but it had low coding potential, and a BLAST homology search against its hypothetical amino acid sequence did not reveal any homologous proteins (publication III).

5.4.2 TPCAT knockdown in prostate cancer cells (I)

To investigate the functional role of TPCATs in PC, we first wanted to assess whether a reduction in TPCAT transcript levels had any effect on cell proliferation or other cellular properties. The most commonly used method for the reduction of mRNAs is siRNA knockdown, which has also been widely used for the knockdown of lncRNAs (Böttcher et al., 2015; Prensner et al., 2011). In publication I, we used three siRNA oligonucleotides (siRNAs 1–3) to successfully knockdown *PCAT5* in PC-3 cells that highly express *PCAT5* (Figure 23A). *PCAT5* knockdown caused drastic reductions in cell growth, migration, invasion, and colony formation compared to the negative control siRNA (publication I). In addition, *PCAT5* knockdown in DuCaP cells via an siRNA pool (siRNAs 2 and 3) also reduced cell growth (publication I). However, we did not observe a significant growth effect in *PCAT5*-negative 22Rv1 cells (Figure 23A) after *PCAT5* knockdown (publication I).

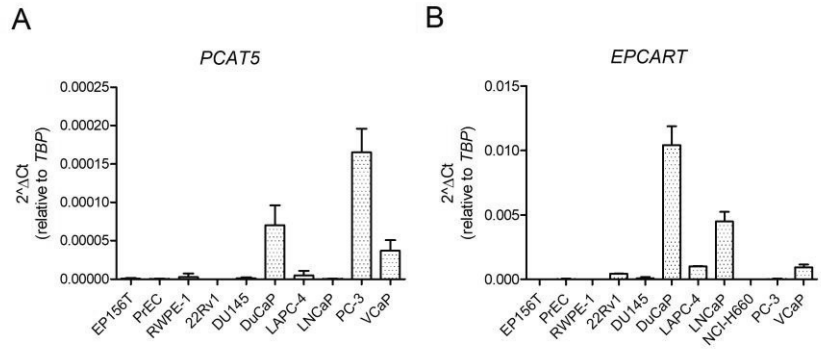


Figure 23. *PCAT5* and *EPCART* expression in normal prostate and PC cell lines. The expression levels of *PCAT5* (A) and *EPCART* (B) were measured by qRT–PCR. The expression values were calculated relative to those of *TBP*. Error bars, SD.

Next, we aimed to perform an siRNA-mediated knockdown screen of multiple TPCATs in PC cell lines to systematically assess the effect of TPCAT knockdown on cell growth. We chose *EPCART* for the initial experiments to test and optimize the siRNA knockdown conditions. *EPCART* was targeted by three siRNAs (siRNAs 1–3) in two *EPCART*-positive cell lines (LNCaP and DuCaP) and in two *EPCART*-negative cell lines (RWPE-1 and PC-3) (Figure 23B). Knockdown of all three siRNAs was confirmed by qRT–PCR, although siRNA 1 was less effective in LNCaP cells than were siRNAs 1 and 2 in DuCaP cells (Figure 24A). Interestingly, growth experiments indicated that siRNA 3 had no effect on the growth of *EPCART*-positive cells but had a major effect on the growth of *EPCART*-negative cells. In addition, both siRNAs 1 and 2 had strong effects on all the tested cell lines. These results indicated that there was a strong off-target effect on the cells.

To investigate how broad this effect was, we used three additional siRNAs for *EPCART* (siRNAs 4–6) and three siRNAs for *TPCAT-18-31849* (siRNAs 1–3). The silencing of *EPCART* by siRNAs 4–6 was equally efficient in LNCaP cells (Figure 24B), but only siRNA 4 had an effect on growth (Figure 24C). In RWPE-1 cells, siRNA 4 had a growth effect similar to that in LNCaP cells; in PC-3 cells, all new *EPCART* siRNAs had equally significant effects on growth (Figure 24D). In the case of *TPCAT-18-31849*, siRNA 3 had the greatest silencing effect on *TPCAT-18-31849*-positive 22Rv1 cells (Figure 25A–B) but had no effect on growth (Figure 25C). In contrast, siRNA 2 had the greatest effect on growth (Figure 25C) but the smallest effect on *TPCAT-18-31849* silencing (Figure 25B). Moreover, knockdown of *TPCAT-18-31849* had similar growth effects on both *TPCAT-18-31849*-positive (22Rv1) and *TPCAT-18-31849*-negative (RWPE-1) cells (Figure 25C).

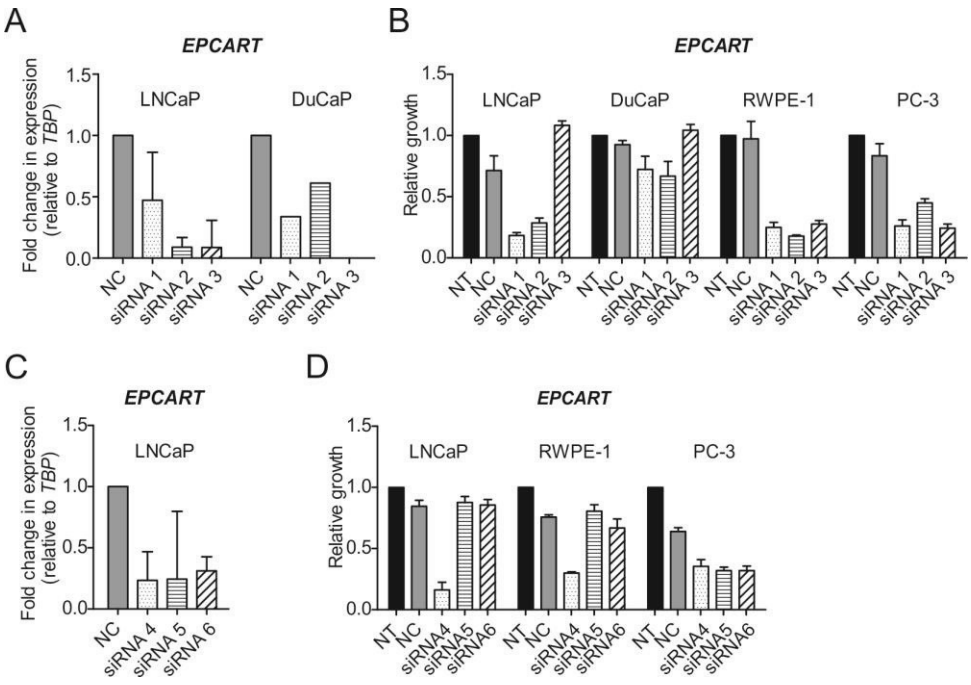


Figure 24. *EPCART* siRNA-mediated knockdown. A and C) siRNA knockdown was validated in *EPCART*-positive cells by qRT–PCR. B and D) Cell proliferation was measured by alamarBlue, and the values on Day 4 were compared to the values on Day 1. NT, untransfected siRNAs; NC, negative control siRNA; siRNAs 1–6, *EPCART* siRNAs 1–6. Error bars, SD.

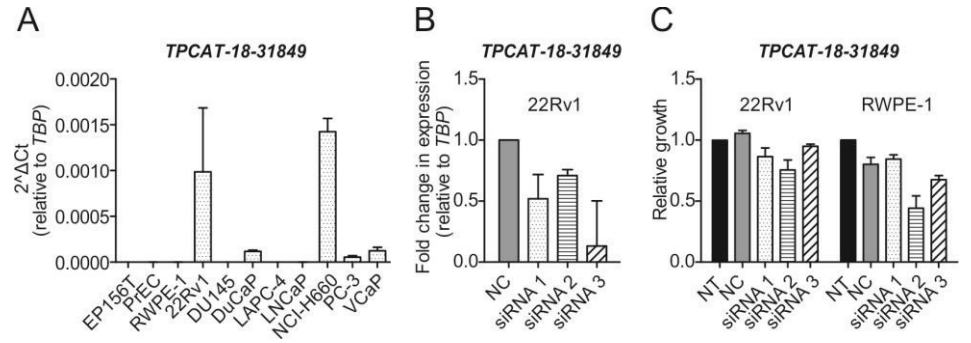


Figure 25. *TPCAT-18-31849* siRNA-mediated knockdown. A) Expression of *TPCAT-18-31849* in prostate cell lines was measured by qRT–PCR. The expression values were calculated relative to those of *TBP.* B) Knockdown was validated in *TPCAT-18-31849*-positive cells by qRT–PCR. C) Cell proliferation was measured by alamarBlue, and the values on Day 4 were compared to the values on Day 1. NT, not transfected; NT, not transfected; NC, negative control; siRNAs 1–3, *TPCAT-18-31849;* siRNAs 1–3. Error bars, SD.

As the off-target effects were likely limited to *EPCART*, we repeated the *PCAT5* knockdown experiments (publication I) with additional cell lines and optimized transfection protocols. Knockdown was performed in *PCAT5*-positive VCaP cells with the same transfection protocol as in publication I. In 22Rv1 and RWPE-1 cells, in which *PCAT5* expression was undetected (Figure 23A), transfection was performed similarly to that of *EPCART* and *TPCAT-18-31849* by using reverse instead of forward transfection. In publication I, the expression of *PCAT5* was successfully silenced in *PCAT5*-positive PC-3 cells with all three siRNAs (siRNAs 1–3), of which siRNAs 2 and 3 were the most effective. Similar results were obtained in VCaP cells (Figure 26A). However, where all siRNAs had great growth effects on PC-3 cells, siRNA 3 had only a minor effect on the growth of VCaP cells (Figure 26A). Furthermore, *PCAT5* siRNAs reduced cell growth in both *PCAT5*-negative cell lines in similar manners (Figure 26B). Interestingly, by using a different transfection protocol, we detected a difference in 22Rv1 cells that was not affected by siRNAs in publication I. Together, these results imply that off-target effects are common, especially when the expression of TPCATs is low or undetectable.

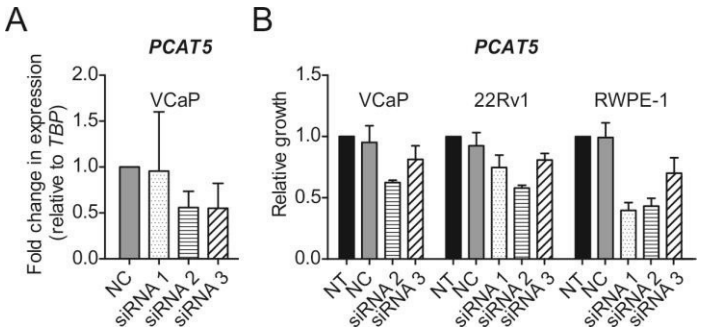


Figure 26. *PCAT5* siRNA-mediated knockdown. A) PCAT5 knockdown was validated in *PCAT5*-positive cells by qRT–PCR. B) Cell proliferation was measured by alamarBlue, and the values on Day 4 were compared to the values on Day 1. NT, untransfected siRNAs; NC, negative control siRNA; siRNAs 1–3, *PCAT5* siRNAs 1–3.

To test whether the off-target effect on growth could be reduced by adjusting the concentration of siRNAs, we repeated the *EPCART* knockdown experiments in LNCaP and RWPE-1 cells with various concentrations (10, 1, and 0.1 nM) of siRNAs 4–6. The expression of *EPCART* decreased at all concentrations tested (Figure 27A). However, even the lowest concentration of siRNA4 affected the growth of both *EPCART*-positive (Figure 27B) and *EPCART*-negative cells (Figure 27C). This finding suggested that off-target effects could be sustained even at very

low siRNA concentrations. We concluded that because the off-target effects we encountered were strong and general among TPCATs and independent of oligonucleotide design, we could not perform siRNA screening experiments.

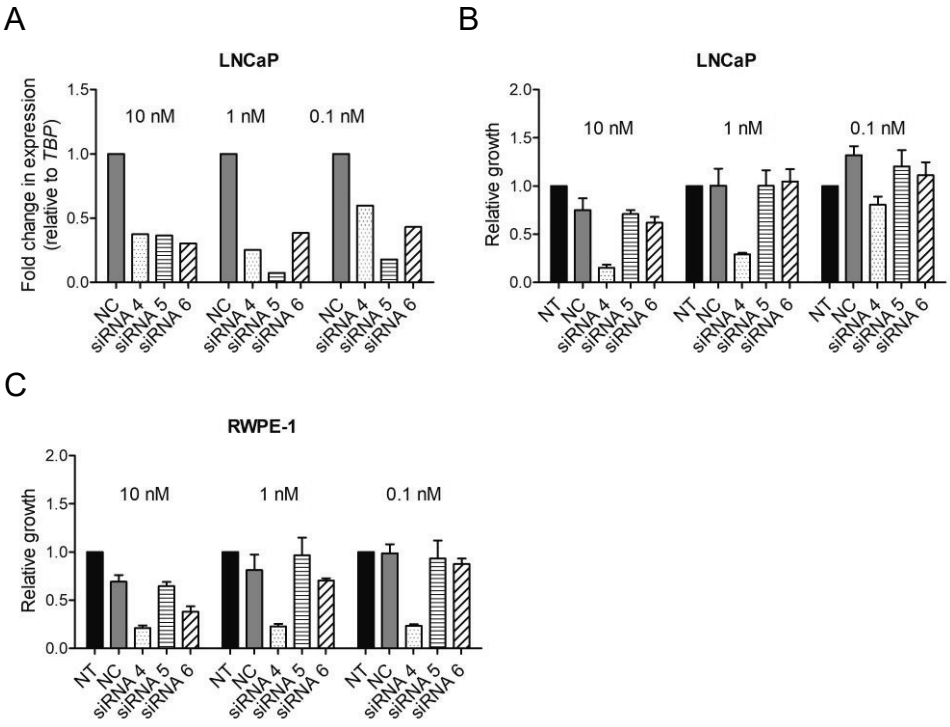


Figure 27. *EPCART* siRNA-mediated knockdown experiments with different siRNA concentrations. A) Knockdown of *EPCART* was validated in *EPCART*-positive cells by qRT–PCR. B-C) Cell proliferation was measured in *EPCART*-positive LNCaP (B) and *EPCART*-negative RWPE-1 cells (C) by alamarBlue, and the values on Day 4 were compared to the values on Day 1. NT, untransfected siRNAs; NC, negative control siRNA; siRNAs 4–6, *EPCART* siRNAs 4–6.

In recent years, LNA-modified antisense oligonucleotides have become widely popular among lncRNA studies because of their ability to knock down RNA expression in the nucleus, where the majority of lncRNAs reside. To investigate whether ASOs could also be used for the knockdown of TPCATs, three LNA-modified ASOs (ASOs 1–3) were designed against *EPCART* transcripts and transfected into *EPCART*-positive (LNCaP and DuCaP) and *EPCART*-negative (DU145, CAL-51, and NIH-3T3) cells. *EPCART* knockdown was validated in LNCaP and DuCaP cells for ASOs 2 and 3 but not for ASO 1 (Figure 28A). In both *EPCART*-positive cell lines, both ASOs 2 and 3 drastically reduced cell growth

(Figure 28B). Unfortunately, we also observed the same growth effects in *EPCART*-negative PC cells (DU145), breast cancer cells (CAL-51) and mouse fibroblasts (3T3; Figure 28B). These results indicate that TPCAT-targeting ASOs could also cause off-target effects in cells. Taken together, both siRNA and ASO knockdown of TPCATs induced off-target effects that led to unreliable results.

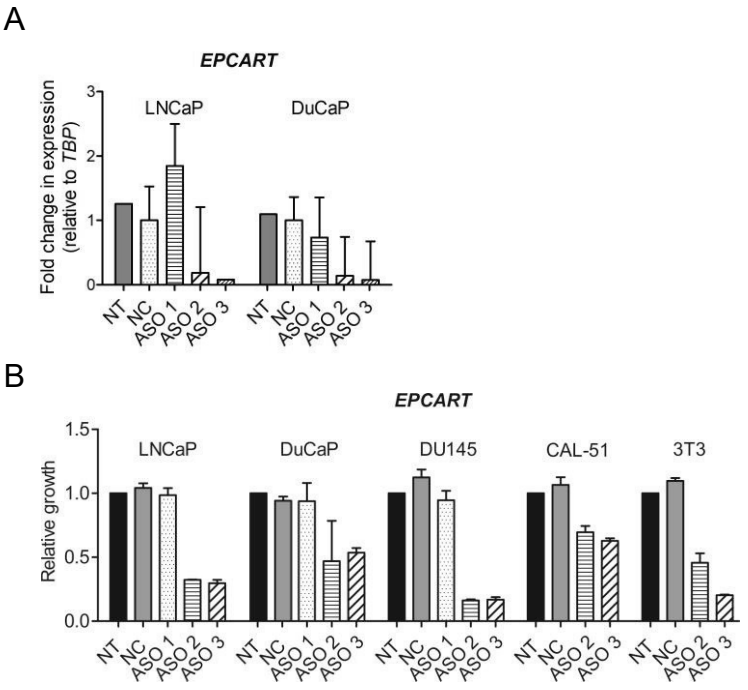


Figure 28. *EPCART* ASO-mediated knockdown. A) Knockdown of *EPCART* was validated in *EPCART*-positive cells by qRT–PCR. B) Cell proliferation was measured by alamarBlue, and the values on Day 4 were compared to the values on Day 1. NT, not transfected ASOs; NC, negative control siRNA; ASOs 1–3, *EPCART* ASOs 1–3.

5.4.3 Functional role of *EPCART* in prostate cancer (II, III)

EPCART knockout cell model (II, III)

As the results of the siRNA and ASO knockdown experiments were unreliable due to off-target effects, we utilized the CRISPR-Cas9 system to knock out the expression of *EPCART*. With protein-coding genes, the use of a single sgRNA to delete a few nucleotides in the gene body is typically enough to knock out protein

expression. However, deletion of such a minor portion of a lncRNA gene is generally not enough to knock out lncRNA expression. To overcome this, two sgRNAs were designed to delete a genomic region (approximately 2.2 kb) containing the first and second exons and the proximal promoter of *EPCART* (Figure 29). We chose LNCaP cells as a model for the KO cells because they are *EPCART*-positive (Figure 23B) and relatively easy to manipulate. The *EPCART*-KO cells were generated by GenScript's CRISPR Gene Editing Services, and they successfully produced two *EPCART-KO* cell clones (*EPCART*-del; del-4 and del-56) and one *EPCART* wild-type clone (WT), for which the CRISPR-Cas9 system failed to delete the targeted region (publication II).

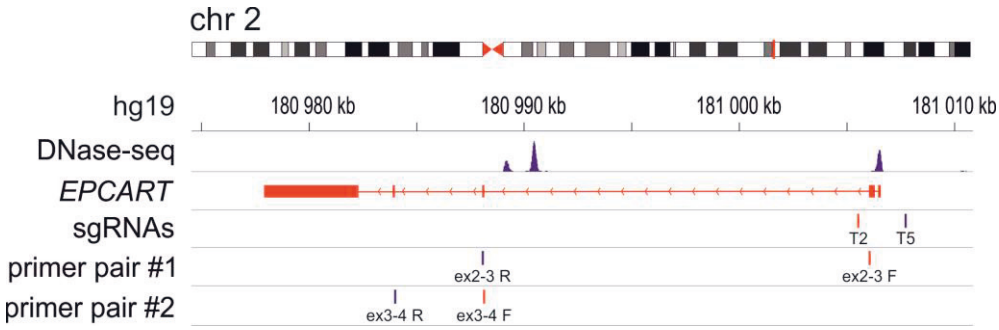


Figure 29. *EPCART-KO* oligos. Single guide RNAs T2 and T5 mark the CRISPR-Cas9 cut sites, and ex2-3_F/R and ex3-4_F/R mark the primers used for *EPCART* expression validation. Retrieved from publication II.

We performed WGS to validate the *EPCART* deletion in *EPCART*-del clones. Interestingly, we found that while the del-56 clone had a distinct deletion between the cut sites, the del-4 clone did not (Figure 30). Instead, del-4 had an inversion at the deletion site. Because LNCaP cells are tetraploid, we wanted to determine whether the inversion was found in all chromosomes or only some chromosomes. For this purpose, we designed primers spanning the deletion site (with amplicon sizes of 2613 bp for the full-length product and 392 bp for the deleted product) for RT–PCR analysis. We found that the del-4 clone had chromosomes with both deletions and undeleted sequences (Figure 31). The bands were then gel-extracted and subjected to Sanger sequencing. As expected, while the WT clone produced a normal sequence from the full-length product, we only observed inverted sequences from the full-length product from del-4 (data not shown). In both clones, the deletion significantly reduced, although did not fully eliminate, the amount of *EPCART* transcripts in the cells (Figure 32).

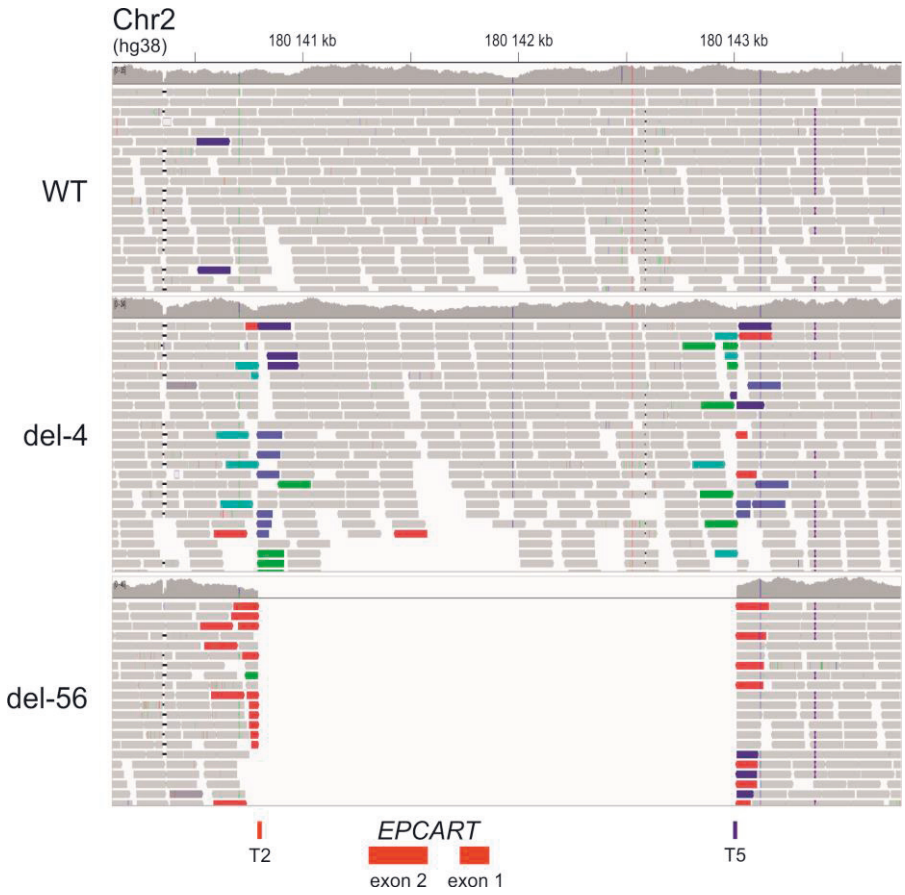


Figure 30. WGS results of the *EPCART* deletion sites in *EPCART*-del and WT clones. T2 and T5 mark the CRISPR-Cas9 cut sites.

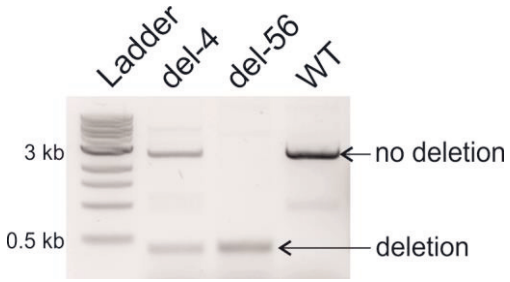


Figure 31. AGE analysis of *EPCART*-del and WT clones with primers spanning the deletion site. The estimated sizes of the full-length and deleted PCR products were 2613 bp and 392 bp, respectively.

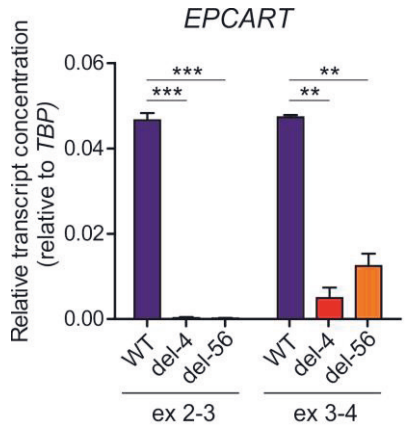


Figure 32. Amounts of *EPCART* transcripts in *EPCART*-del and WT clones. Two primer pairs were used to detect the amounts of *EPCART* transcripts via ddPCR. Retrieved from publication II.

The effect of *EPCART* on cell phenotype (II)

To investigate whether *EPCART* deletion affected the phenotypes of the cells, we performed cell proliferation and migration assays with the *EPCART*-del and WT clones. Both proliferation and migration were reduced in both deletion clones compared to those in the WT clone (Figure 33A–B). The invasion of *EPCART*-del clones was also studied, but as LNCaP cells are inherently poorly invasive, we could not detect many invasive cells in either the *EPCART*-del or WT clones (data not shown). We also studied whether *EPCART* overexpression would enhance PC cell growth. For this purpose, we cloned the *EPCART* transcript sequence (without exon 1) into the pcDNA3.1 mammalian expression plasmid, which has a constitutive CMV promoter. PC cells (LNCaP and PC-3) were transfected with the *EPCART* plasmid or with the empty pcDNA3.1 plasmid as a control, and a stable cell pool was created through antibiotic selection. Clear overexpression of *EPCART* was detected in the cells transfected with the *EPCART* plasmid (>10-fold expression compared with the LNCaP empty control; data not shown), and the overexpression of *EPCART* significantly (p <0.05, Student's t test) increased the growth of LNCaP cells but not PC-3 cells (Figure 33C–D).

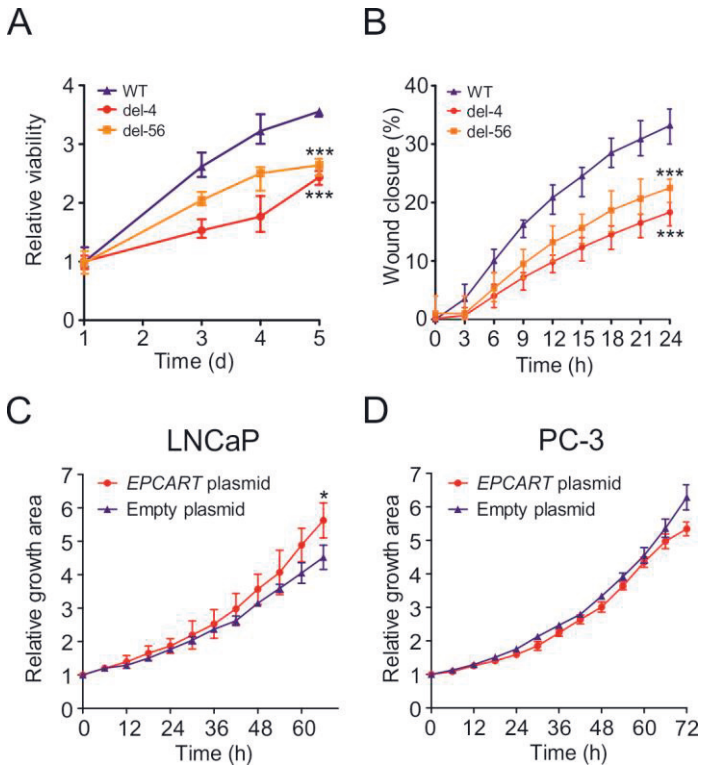


Figure 33. Effect of *EPCART* on PC cells. A) The proliferation of *EPCART*-del and WT clones was measured by alamarBlue. B) The migration of *EPCART*-del and WT clones was measured by a wound healing assay with a Cell-IQ time-lapse imaging system. C–D) The proliferation of LNCaP (C) and PC-3 (D) cells stably overexpressing *EPCART* or the control plasmid was measured with a Cell-IQ time-lapse imaging system. Error bars, ranges in A–B and SD in C–D; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; data were assessed with an unpaired two-tailed t test. . A-B Retrieved from publication II.

EPCART localization in prostate cancer (III)

According to the results of the *EPCART* ISH studies, *EPCART* was mostly localized in the cytoplasm of PC cells (Figure 14). As the Fast Red dye used in the ViewRNA ISH protocol is observable via both brightfield and fluorescence microscopy, we also imaged the prepared tissue slides with a confocal microscope for higher resolution (Figure 34). Indeed, *EPCART* was more clearly detected in the cytoplasm. To validate this observation, we extracted RNA from the cytoplasmic and nuclear cell fractions of LNCaP and DuCaP cells and measured the expression of *EPCART* in these fractions via qRT–PCR (publication III). As expected,

EPCART was primarily located in the cytoplasmic fractions of both cell lines (publication III).

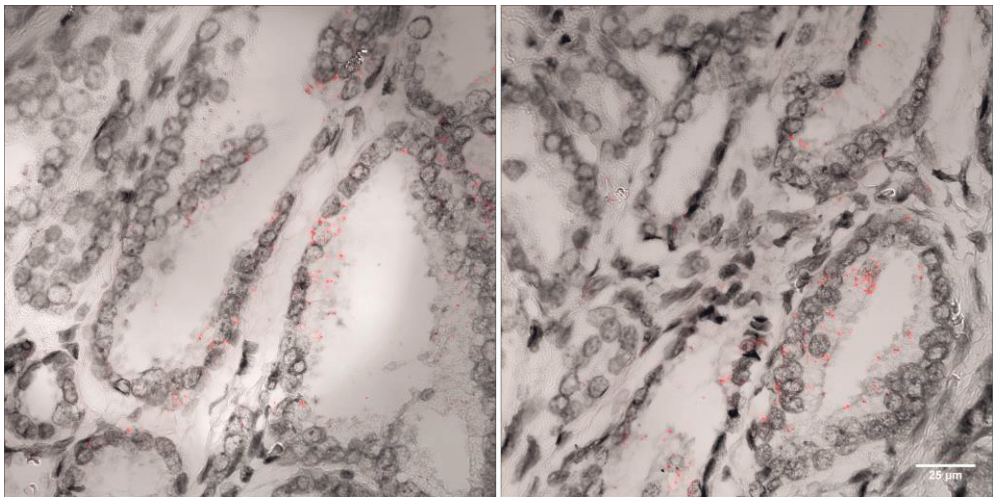


Figure 34. *EPCART* localization in PC cells. Signals from *EPCART* transcripts were detected with Fast Red under a confocal microscope. The images were taken from the same slide with the same settings. The right image was modified from publication III.

EPCART target gene studies (III)

To better understand the mechanisms of *EPCART* function, we studied the molecular changes occurring in the *EPCART*-KO model. For this purpose, we first performed whole transcriptomic studies on the *EPCART*-del and WT clones (publication III). Pathway analysis of differentially expressed protein-coding mRNAs revealed that many molecular and cellular functions related to cell proliferation and growth were significantly affected by *EPCART* silencing (Figure 35). In particular, RNA translation was inhibited, and many canonical pathways associated with translation, including eIF2 signaling, protein ubiquitination, eIF4 and p70S6K signaling, and mTOR signaling, were significantly affected (publication III).

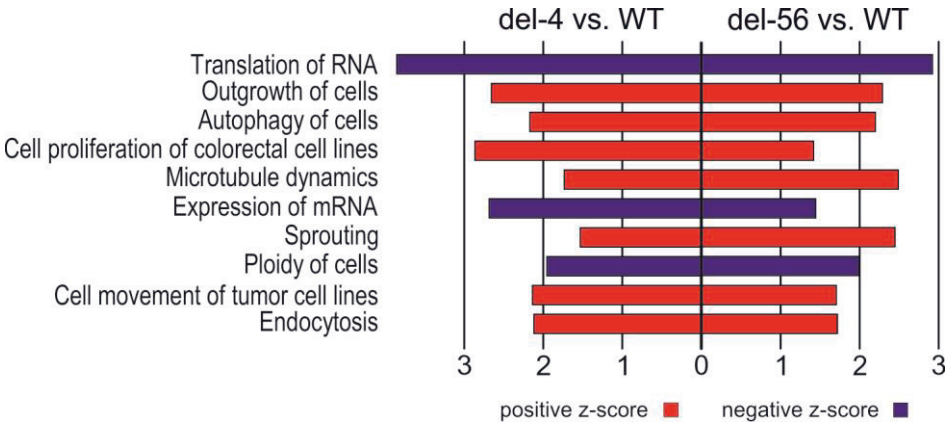


Figure 35. Identification of the molecular and cellular functions of *EPCART*-del clones. The top 10 most different (z score >2) functions between the *EPCART*-del (del-4 or del-56) and WT clones that were shared by both clones are shown. Modified from publication III.

To determine whether translation was globally inhibited in *EPCART*-knockout cells, we performed polysome profiling on the *EPCART*-del and WT clones (publication III). While we did not detect significant differences in the polysome profiles between the *EPCART*-del and WT clones, we detected *EPCART* in the polysome fractions of the WT cells (Figure 36). Interestingly, two lncRNAs, *SChLAP1* and *MALAT1,* which were previously found to be mostly nuclear (Prensner et al., 2013; Tripathi et al., 2010), were also detected in the polysomal fractions of the WT cells in a manner similar to that of the *EPCART* transcripts (Figure 36). As polysome-associated lncRNAs are sometimes degraded at ribosomes (Carlevaro-Fita et al., 2016), we assessed the stability of *EPCART* in WT cells by measuring translation elongation with CHX and transcription with ActD (publication III). *EPCART* transcription was not substantially affected by the treatments, indicating that *EPCART* degradation may not be the main mechanism by which *EPCART* binds ribosomes. These results suggest that the effect of *EPCART* on translation likely depends more on regulating translation-associated pathways than on direct ribosome associations of *EPCART* itself.

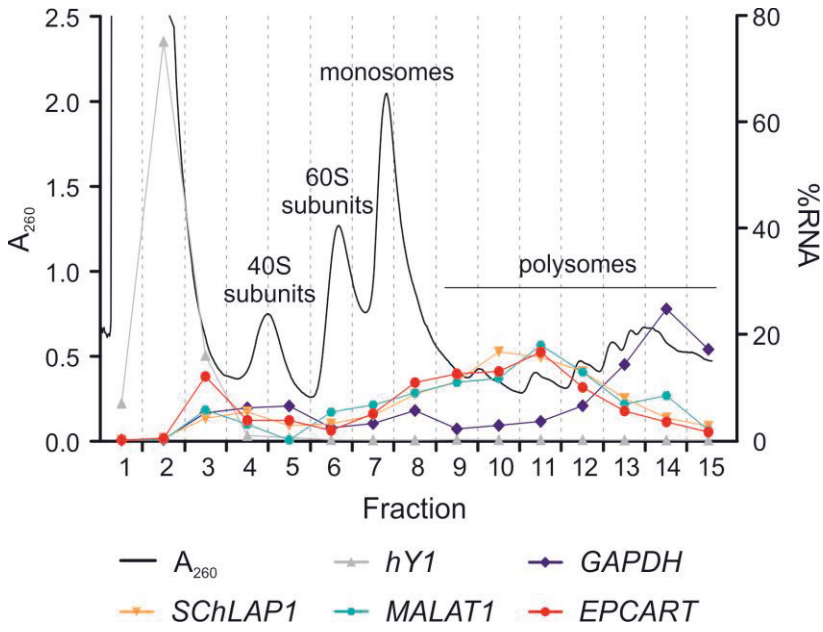


Figure 36. RNA percentages of *EPCAR-T cells* in polysome profiling fractions. The percentage of RNA in each fraction was calculated for *EPCART*, *GAPDH* (positive control), *hY1* (negative control), and two nuclear lncRNAs (*SChLAP1* and *MALAT1*) and plotted in the same graph as the polysome profile (A_{260}). Profile peaks representing ribosomal subunits, monosomes, and polysomes are marked in the graph. Modified from publication III.

To gain insight into the effect of *EPCART* KO on translated proteins, we performed quantitative proteomics by SWATH-MS for the *EPCART*-del and WT clones (publication III). Differential protein expression analysis revealed that PDCD4, a translation inhibitor and a tumor suppressor, was the most significantly overexpressed protein in both *EPCART*-del clones (publication III). As PDCD4 is downstream of the mTORC1 signaling pathway, we next studied whether this pathway is dysregulated in *EPCART*-knockout cells by immunoblotting. We found that PDCD4 levels were increased in *EPCART*-del clones due to decreased phosphorylation of PDCD4 at Ser67, which led to decreased degradation of the PDCD4 protein (Figure 37). Additionally, p70S6K, an inhibitor of PDCD4, was less active in *EPCART*-del clones due to decreased phosphorylation at Thr389 (Figure 37). Phosphorylation of 4E-BP1, a partner of another mTORC1-controlled translation regulatory pathway, was not found to be affected (publication III). Phosphorylation of mTORC1 subunits and their protein expression levels were also studied (publication III). We found that PRAS40 phosphorylation at Thr246, which is catalyzed by AKT and leads to the dissociation of PRAS40 from mTORC1 (Roux

& Topisirovic, 2018), was decreased and DEPTOR protein expression levels were increased in *EPCART*-del clones (Figure 37). Additionally, we detected less AKT phosphorylation at Ser473 and thus less activation of AKT in *EPCART*-del clones (Figure 37). Together, these results suggest that the translation inhibition detected in *EPCART*-del clones is caused, at least in part, by inhibition of the AKT/mTORC1/p70S6K/PDCD4 pathway.

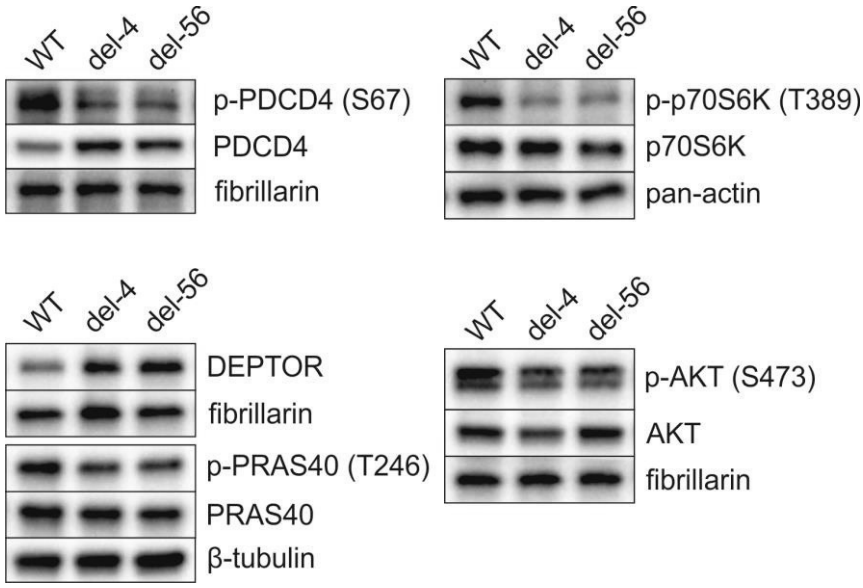


Figure 37. Immunoblot analysis of AKT/mTORC1/p70S6K/PDCD4 signaling in *EPCART*-knockout cells. The target proteins and their phosphorylation levels are shown for the *EPCART*-del and WT clones. Fibrillarin, pan-actin, or β-tubulin were used as loading controls. Modified from publication III.

To study the effect of mTOR inhibitors on *EPCART*-del clones, we treated cells with rapamycin (an mTORC1 inhibitor) and Torin1 (an inhibitor of both mTORC1 and mTORC2). Based on the p70S6K Thr389 phosphorylation results, the mTOR pathway was successfully inactivated in all clones after treatment (Figure 38A). Both rapamycin and Torin1 induce autophagy, so we measured the LC3-II/I ratio, which is used as a marker for autophagy (Mizushima & Yoshimori, 2007). We detected a positive LC3-II/I ratio in the WT clone but not in the *EPCART*-del clones (Figure 38A), indicating that autophagy was affected in the *EPCART*-del clones. *EPCART* expression was also assessed in WT clone samples, and we detected significantly greater expression of *EPCART* in Torin1-treated samples but very little effect in rapamycin-treated samples (Figure 38B).

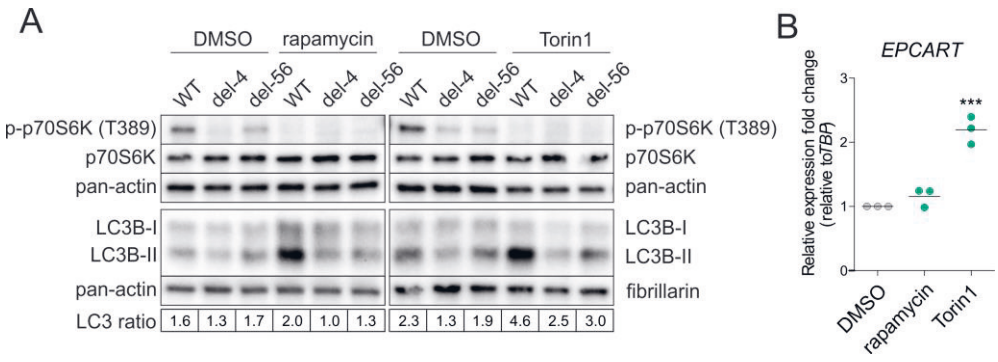


Figure 38. mTOR inhibitor studies in *EPCART*-knockdown cells. A) Immunoblot analysis of markers of mTOR activity (p70S6K) and autophagy (LC3B). DMSO was used as a treatment control. Pan-actin and fibrillarin were used as loading controls. The LC3 ratio was calculated from quantified LC3B intensities (LC3B-II/I). B) *EPCART* expression analysis in three biological replicates by qRT–PCR. ***, $p < 0.001$; the data were assessed with an unpaired two-tailed t test.

PDCD4 has been reported to be a tumor suppressor, and its low expression has been associated with poor prognosis in many cancers (Li et al., 2016). We also found that low PDCD4 mRNA and protein levels were correlated with aggressive PC in clinical PC datasets (Figure 39A–B). This finding was further validated by our IHC studies, which revealed that decreased nuclear and increased cytoplasmic PDCD4 levels were associated with PC aggressiveness (Figure 39C–D). Additionally, we found negative correlations between *EPCART* expression and PDCD4 protein expression in Tampere primary PC specimens (Spearman r = −0.38; n = 17) and TCGA-PRAD data (Spearman r = −0.21; n = 283), supporting the idea that *EPCART* expression and PDCD4 protein regulation are linked.

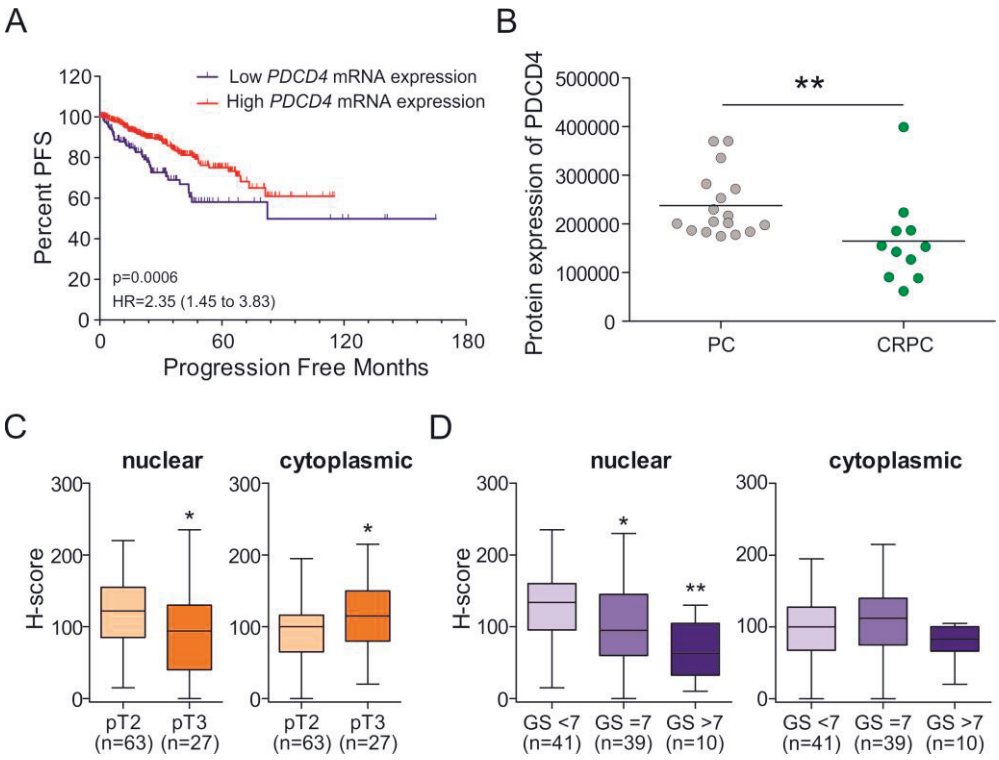


Figure 39. *PDCD4* expression is associated with PCa progression. A) Kaplan–Meier estimates of progression–free survival (PFS) in PCa patients divided by the *PDCD4* expression level in the TCGA-PRAD cohort. The first quartile was used as a cutoff between low and high expression. *P* values were calculated by the log-rank test. HR, hazard ratio. B) PDCD4 expression in primary PCa and in CRPC based on MS results in the Tampere cohort. C–D) IHC analysis of PDCD4 in untreated primary PC samples that were divided into groups according to pT stage (C) or Gleason score (D). H-scores for nuclear and cytoplasmic staining are shown separately. The error bars display the minimum and maximum values. In B–D, the mean expression of each group is marked with a black line, and the significance of expression differences was assessed with the Mann–Whitney test. ; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Modified from publication III.

6 DISCUSSION

6.1 TPCATs are dysregulated in prostate cancer

Only a handful of papers have been published describing the systematic search for PC-associated lncRNAs (Crea et al., 2014; Du et al., 2013; Iyer et al., 2015; Prensner et al., 2011; Ren et al., 2012; Yan et al., 2015), although several individual lncRNAs have been investigated in PC over the years. In publication I, our own RNA-seq data from the Tampere PC cohort samples were particularly deep compared to those of other studies. This allowed us to discover PCATs that had not been previously identified, as many TPCATs were expressed at low or moderate levels on average, and the expression levels of certain TPCATs were high in only a few samples. TPCATs were also found to be highly PC- and/or CRPC-specific and not highly expressed in other tissues. Surprisingly, most of the TPCATs were overexpressed rather than downregulated in PC. Over half of the overexpressed TPCATs were more highly expressed in CRPC than in primary PC, and one-fourth of the TPCATs were CRPC specific. One explanation for this could be a process of continued chromatin relaxation that has been observed during PC development (Braadland & Urbanucci, 2019; Urbanucci et al., 2017; Uusi-Mäkelä et al., 2020). Altered chromatin structure due to genetic and epigenetic alterations can lead to increased DNA accessibility during PC development (Braadland & Urbanucci, 2019). This can then provide transcriptional regulators with better access to previously closed parts of the genome and thus lead to aberrant regulation of genes, including those encoding lncRNAs, in these areas. We also observed that more than half of the TPCATs had only one exon, and one-third came from the X chromosome. The AR locus is located on the X chromosome and is often amplified in CRPC, which could explain the excess X chromosome-derived TPCATs. Whether these TPCATs are functional or products of dysregulated, overactive transcriptional machinery remains to be determined.

6.1.1 Dysregulation of prostate cancer-specific transcription factors leads to aberrant expression of many TPCATs

The regulation of several individual lncRNAs has been studied in PC, in addition to a few studies concentrating on identifying androgen-regulated lncRNAs (Takayama et al., 2011; Zhang et al., 2018). However, no studies have systematically investigated the regulation of PC-associated lncRNAs. In publications I and II, we examined the correlations between TPCAT and TF expression and found that many TPCATs were associated with ERG, which we further validated in *in vitro* experiments for selected TPCATs (which were overexpressed in primary PCs and had >1 exon). We also studied the binding of ERG and other well-known PC regulators, AR, FOXA1, and HOXB13, to putative promoters of these TPCATs. The majority of TPCATs can have these TFs, especially ARs, bound to their promoters, often in collaboration with other TFs. Many of the selected TPCATs were also found to be androgen regulated in *in vitro* studies. Taken together, our results revealed that dysregulated AR binding modulates not only the expression of protein-coding genes but also that of many PC-associated lncRNAs. Other PC-specific TFs, especially ERG, also cooperate with AR in the regulation of PC-associated lncRNAs. As aberrant chromatin binding of these TFs, often at the same sites, can promote PC development and progression (Hankey et al., 2020; Kron et al., 2017), the regulation of lncRNA expression through these TFs could also be a way to impact them. Although genetic alterations can also lead to the dysregulation of cancer lncRNAs (Schmitt & Chang, 2016), we did not find this to be the case with TPCATs.

6.1.2 Role of ERG in the regulation of TPCATs

Surprisingly, we found that many TPCATs are associated with *TMPRSS2-ERG* fusions and ERG overexpression, which has not been previously studied; only *SChLAP1* has been found to be associated with ETS gene fusions (Prensner et al., 2013). Nearly 60% of the 34 overexpressed TPCATs were associated with *ERG* expression and had ERG binding sites at their promoters, and for half of these TPCATs, we were able to confirm that they were regulated by ERG in PC cells. The strongest correlation was detected between *ERG* and *PCAT5* expression, and *PCAT5* had prominent ERG binding at its promoter. According to H3K27ac data, the regulation of *PCAT5* was also clearly activated only when ERG was present in tumors. Although *EPCART* expression was strongly correlated with ERG expression and there was an ERG binding site at its promoter, we did not find clear

evidence of its ERG regulation via *in vitro* experiments. In addition, there was no significant difference in H3K27 acetylation at the *EPCART* promoter between *TMPRSS2-ERG*-positive and TMPRSS2-ERG-negative tumors, and the CRE that was detected was outside the promoter sites, where binding of ERG and other TFs was observed. These results indicated that ERG is not highly significantly involved in the regulation of *EPCART* expression. This finding was further supported by the PC progression analysis, which revealed a population of *EPCART*-high ERG-negative tumors associated with aggressive PCs. Taken together, while we showed that ERG is involved in the regulation of a significant group of PC-associated lncRNAs, it might not be the most essential modulator of the expression of many lncRNAs. This finding goes hand in hand with the theory that ERG is one of the master TFs in PC that induces chromatin modeling and co-opts other PC-specific master regulators, most importantly AR, FOXA1, and HOXB13, by physically interacting with them (Kron et al., 2017). Hence, while ERG promotes the expression of many PC-associated lncRNAs, other PC-specific TFs might play more significant roles in their regulation.

6.2 *EPCART* promotes prostate cancer development and progression

6.2.1 *EPCART* is a highly expressed lncRNA in prostate cancer and is regulated by many cancer-associated transcription factors

Over the last two decades, several lncRNAs have been found to play roles in the development and progression of PC (Mirzaei et al., 2022; Mitobe et al., 2018). To determine whether TPCATs also play functional roles in PC, we first narrowed our list of potentially relevant TPCAT candidates based on their expression and regulation profiles. One TPCAT, *EPCART*, attracted our attention because it was found to be very abundantly and specifically expressed in PCs, and it was also regulated by ARs in PC cells overexpressing ERG. *EPCART* expression was, in fact, one of the highest among all PC-associated lncRNAs found in the MiTranscriptome catalog (Iyer et al., 2015). *EPCART* was uniformly expressed in primary PC and localized CRPC, but we observed a slight reduction in *EPCART* expression in mCRPC samples compared to that in primary PC samples. This difference indicates that *EPCART* is mostly expressed in early-stage PC, which is supported by its

association with *TMPRSS2-ERG* fusion and *PTEN* loss, both of which occur early in PC development (Bismar et al., 2011; Perner et al., 2007). Additionally, after metastasis, the expression of *EPCART* might not be essential for the cells, which could lead to reduced *EPCART* expression through clonal evolution. In normal tissues, there was essentially no *EPCART* expression except in the normal prostate and, to some extent, in the testes. This finding implies that *EPCART* might also play a role in the normal prostate, although its expression is tightly controlled under normal conditions. Moreover, the normal prostate samples in which *EPCART* expression was detected were adjacent normal tissue from PC tumor samples. Hence, it might be possible that, *e.g.,* inadequate grading of the tissue sections or PIN lesions with early-detected mutations such as *TMPRSS2-ERG* already present in the samples could cause the expression of *EPCART* to be greater in these samples than in healthy normal prostates.

In addition to *EPCART*, we found that many other TPCATs were expressed in testes, which is where the majority of all lncRNAs are expressed (Lee et al., 2019). This could be due to the absence of heterochromatin in testes, which allows the binding of transcription regulators to sequences that are inaccessible in other tissues (Lee et al., 2019). It has also been suggested that lncRNAs and many new genes evolved from the testes (Lee et al., 2019). Additionally, TEs become accessible during spermatogenesis, and although repressed in testes by ncRNAs such as piwi-RNAs, they are believed to be the origin of new genetic elements for many genes, including lncRNAs (Lee et al., 2019; Mattick et al., 2023). *EPCART* also has many TEs (LINEs and LTRs) embedded in its exons, including a LINE-2 at its 5' end. This LINE-2 is also the whole exon 1 of *EPCART* and aligns with the binding sites for many TFs. As exon 1 was detected in only a fraction of *EPCART* transcripts and because we found prominent variation at the start site of the 5' end, it is possible that exon 1 is a byproduct of inefficient transcription. This phenomenon is common among lincRNAs, as the transcription of lincRNAs has been found to be less precise than that of mRNAs (Schlackow et al., 2017).

Histone modifications can play a role in the regulation of lncRNAs. The *EPCART* promoter did not contain many histone markers in LNCaP cells, except for H3K27 methylation. H3K37me3 is also a repressive marker for lncRNAs (Wu et al., 2010). As *EPCART* was expressed in LNCaP cells, this histone methylation did not seem to completely decrease *EPCART* expression. In PC tumor samples, there was clear acetylation of H3K37, a known activation marker for expression, at the promoter of *EPCART*, which was not present in LNCaP cells. These differences in H3K27 modifications between samples could explain, at least in part, the distinct

disparity observed between the expression of *EPCART* in tumors vs. LNCaP (>7-fold change between the median expression and >200-fold change between the arithmetic mean expression in MiTranscriptome data) or other PC cell lines.

6.2.2 *EPCART* is located in the same topologically associating domain as other prostate cancer-associated genes

TADs and the chromatin loops inside them play significant roles in genome regulation, and changes in TAD boundaries can contribute to cancer development, including in PC (Tettey et al., 2023; Yeo et al., 2023). We found a TAD boundary inside one of the introns of *EPCART*. We did not detect any genetic alterations in the area or changes in the TAD boundaries between normal and PC cell lines, suggesting that the boundary is stable throughout PC development. Interestingly, *SChLAP1* and two other TPCATs were located in the same TAD as *EPCART*. High *SChLAP1* expression has previously been associated with aggressive PC (Prensner et al., 2013, 2014). One protein-coding gene, *UBE2E3,* was also located in the same TAD. To date, there have been no studies published on the functional role of *UBE2E3* in PC, although a gene fusion between *SCHLAP1* and *UBE2E3* has been reported to be the most frequent gene fusion in the Chinese PC atlas (i.e., present in 29% of cases) (Li et al., 2020). Correlation analysis revealed a strong association between the expression of these genes and *EPCART* in primary PC tumors (Spearman ϱ: 0.79 for *SChLAP1* vs. *EPCART* and 0.71 for *UBE2E3* vs. *EPCART* in the Tampere PC cohort), indicating that the regulation of these genes is linked. Many lincRNAs function in *cis* by regulating the transcription of nearby genes (Ransohoff et al., 2018). Certain ncRNAs can also recruit CTCF and cohesin, *i.e.*, proteins needed for TAD formation, to TAD boundaries and regulate gene expression (Yeo et al., 2023). Although we did not detect any significant changes (log2FC >1 or <−1, RNA-seq data) in the expression of either *SChLAP1* or *UBE2E3* in *EPCART*-KO cells, the expression of *UBE2E3* was somewhat decreased in these cells (log2FC: −0.35 in del-4 and −0.62 in del-56). Together, these findings imply that *EPCART* and other genes in the same TAD are coregulated and may interact with each other to promote PC.

6.2.3 *EPCART* promotes prostate cancer cell growth and migration

Although *EPCART* could play a potential role in the regulation of nearby genes in the nucleus, we detected that the majority of *EPCART* transcripts were localized in the cytoplasm, suggesting a possible function. *EPCART* was also found to interact with translationally active ribosomes. This is a common occurrence among cytoplasmic lncRNAs, and one explanation for this is that ribosome recruitment results in the degradation of lncRNAs (Carlevaro-Fita et al., 2016). Alternatively, certain lncRNAs can structurally associate with the translational machinery, or lncRNAs can be translated into micropeptides that can have a functional role in cancers (Vitorino et al., 2021). Although we found no evidence that the association of *EPCART* with ribosomes was caused by any of these mechanisms, we also could not fully rule them out. Moreover, we observed similar expression profiles to those of *EPCART* in the polysomal fraction for *SChLAP1* and *MALAT1*, which are considered nuclear lncRNAs (Prensner et al., 2013; Tripathi et al., 2010). This could either be an indication of specific or unspecific transport of nuclear lncRNAs into the cytoplasm or a specificity issue in the polysome profiling protocol. In the latter case, the polysome association of *EPCART* could also be due to nonspecific binding, which is more plausible, as *EPCART* was found to be cytoplasmic in the RNA-ISH samples.

For functional studies, we generated an *EPCART*-KO model in LNCaP cells via CRISPR-Cas9 technology. Two deletion clones were created, and their *EPCART* expression was successfully diminished, although low levels of *EPCART* expression were detected from exons 3 and 4 (exon 5 was not examined), which are outside the deletion site. This suggests that there is an alternative but less efficient promoter close to exon 3; this putative promoter could reside in the open chromatin region, which also contains the TAD boundary. Functional experiments for studying the phenotypic changes in *EPCART*-KO cells revealed that *EPCART* deletion decreased both cell proliferation and migration. To determine whether the opposite is true, we stably overexpressed *EPCART* in LNCaP and PC-3 cells and measured their proliferation. Indeed, LNCaP cells overexpressing PC-3 cells exhibited increased proliferation, although this effect was not detected in PC-3 cells. The main difference between these cell lines is that PC-3 cells are AR-negative, while LNCaP cells are AR-positive. As *EPCART* is also AR-regulated, it is possible that the pathways required for *EPCART* function are not active in PC-3 cells.

6.2.4 *EPCART* modulates translation via the PI3K/AKT/mTORC1/PDCD4 pathway

To investigate the pathways connected to *EPCART* and what could explain the phenotypic changes observed in the knockout and overexpressing cells, we carried out RNA-seq on *EPCART*-KO cells and pathway analysis of the DEGs. Analysis of both molecular and cellular functions and individual canonical pathways provided strong evidence for the inhibition of protein translation after *EPCART* downregulation. In particular, *EPCART* deletion caused major changes in eIF2 signaling, protein ubiquitination, the eIF4 and p70S6K pathways, and the mTOR pathway. To validate the inhibitory effect on translation, we performed polysome profiling on *EPCART*-KO cells. We did not detect any significant differences in the polysome profiles between *EPCART*-deleted and WT clones, implying that the effect of *EPCART* on translation is not strong enough to cause drastic changes in the formation of translationally active ribosomes. To determine whether *EPCART* deletion influences the protein expression of individual genes, we carried out quantitative proteomics studies on *EPCART*-KO cells. When protein expression in *EPCART*-KO cells was compared to that in wild-type cells, the most highly overexpressed protein was PDCD4, which is known to be a tumor suppressor in several cancers and an inhibitor of translation (Wang & Yang, 2018). In PC, PDCD4 has been found to be repressed by AR signaling via miR-21 (Zennami et al., 2019). PDCD4 is a downstream target of mTORC1 (Matsuhashi et al., 2019), which plays key roles in the regulation of general translation and other pathways promoting cell growth and proliferation (Roux & Topisirovic, 2018). When we studied the phosphorylation of mTORC1 downstream targets, we noticed that the phosphorylation of PDCD4, which leads to the degradation of PDCD4, was reduced in *EPCART*-KO cells, and the phosphorylation of p70S6K, which phosphorylates PDCD4, was similarly reduced. However, another mTORC1-regulated pathway that inhibits translation, 4E-BP1 (Roux & Topisirovic, 2018), was not affected by *EPCART* deletion. These results suggest that mTORC1 is inhibited in *EPCART*-KO cells specifically through the p70S6K/PDCD4 pathway.

mTORC1 is targeted by several upstream regulators, of which AKT is the most notable (Saxton & Sabatini, 2017). We studied the protein levels and phosphorylation status of the mTORC1 subunits in *EPCART*-KO cells and found that the phosphorylation of PRAS40, which dissociates PRAS40 from mTORC1 (Roux & Topisirovic, 2018), decreased and that of the DEPTOR protein increased. Both of these proteins are inhibitors of mTORC1 (Roux & Topisirovic, 2018), which

consequently supports the observation of mTORC1 inactivation in *EPCART*-KO cells. PRAS40 is phosphorylated by AKT, so we analyzed AKT phosphorylation status and found that AKT was inhibited in *EPCART*-KO cells via dephosphorylation at Ser473. The addition of this phosphate group to mTORC2 (Roux & Topisirovic, 2018) could indicate mTORC2 dysregulation in *EPCART*-KO cells. Furthermore, because DEPTOR is an inhibitor of not only mTORC1 but also mTORC2 (Catena & Fanciulli, 2017), its increased levels could lead to enhanced mTORC2 inhibition, which could explain the decreased AKT activation in *EPCART*-KO cells. Further evidence of the co-operative functions of mTORC2 and *EPCART* was provided by mTOR inhibitor studies in the WT clone (LNCaP), which showed that *EPCART* expression was doubled by the inhibition of both mTORC1 and mTORC2, while mTORC1 inhibition alone did not cause a significant change in *EPCART* expression. Upstream modulators of mTORC2 and AKT, either directly or indirectly, include PI3K and PTEN, both of which are frequently mutated in PC (Shorning et al., 2020; Smith et al., 2020). In PC tumor samples, high expression of *EPCART* was associated with *PTEN* loss, providing further evidence for the interaction of *EPCART* with the PI3K/AKT/mTORC1 pathway.

Interestingly, mTOR inhibitor studies also revealed that the expression of a marker for autophagosomes (LC3-II) was not affected by treatment in *EPCART*-KO cells, which was contrary to what was expected because the inhibition of the mTOR pathway should activate autophagy (Saxton & Sabatini, 2017). This finding implies that the phenotypic changes observed in *EPCART*-del cells might be caused by changes in multiple pathways.

Taken together, our results imply that *EPCART* promotes protein translation in PC cells through activation of the PI3K/AKT/mTORC1/PDCD4 pathway. *EPCART* might also have a role in the modulation of autophagic pathways. However, the exact molecular mechanism of *EPCART* and the partners with which it interacts remain unknown and should be investigated further.

6.2.5 *EPCART* is associated with prostate cancer progression and other prognostic markers

A few PC-associated lncRNAs, including *SChLAP1*, are associated with PC progression, and high expression levels of these lncRNAs have been found to correlate with biochemical recurrence and lethal PC (PC-specific mortality or distant metastases) (Mehra et al., 2014, 2016). In the Tampere PC cohort, we discovered

that high *EPCART* expression was correlated with biochemical recurrence in patients with PC. A previously published study by Böttcher et al. also revealed that high expression of *EPCART* (called EPCAT2R709 in their paper) was associated with PC-related death and the development of clinical metastases in an independent PC cohort from Erasmus Medical Center (Böttcher et al., 2015). Taken together, these findings indicate that high expression levels of two PCATs from the same TAD correlate with aggressive PC. Additionally, low PDCD4 expression at both the mRNA and protein levels was associated with aggressive PC. PDCD4 has previously been found to be a progressive marker in several other cancer types (Li et al., 2016) but has not been studied extensively in PC (Zennami et al., 2019). We also found a negative correlation between *EPCART* expression and PDCD4 protein expression in PC tumor samples, providing further evidence for the relationship between *EPCART* and PDCD4.

6.3 Can TPCATs be utilized for clinical applications?

One of the main themes in cancer-associated lncRNA research has been the prospect of their use as biomarkers. The tissue specificity and aberrant expression of lncRNAs in cancers make certain lncRNAs ideal candidates for biomarkers. Thus far, only one approved lncRNA biomarker, *PCA3*, is in use for diagnostic purposes. Although no approved prognostic markers are available, many have been studied, including *SChLAP1* in PC. In the Tampere PC cohort, three TPCATs, *EPCART*, *TPCAT-3-174133*, and *TPCAT-18-31849*, were found to be associated with biochemical recurrence in PC independent of other prognostic markers. Among these TPCATs, *EPCART* was also highly expressed among PC-associated lncRNAs and highly PC specific, which are also good attributes for a diagnostic marker. Based on the RNA ISH studies, *EPCART* was found specifically in PC lesions, but the detected signal was relatively weak, and more extensive ISH studies could be more informative. Nevertheless, we showed that it is possible to detect *EPCART* in tissue specimens. Tissue biopsies generally cause adverse effects for patients, and finding alternative ways to detect biomarkers has been widely studied in recent years, mostly in blood and urine samples (Connal et al., 2023). These so-called liquid biopsies can contain RNAs, although their levels of degradation can vary markedly; "naked" mRNAs in plasma degrade in 15 seconds, while miRNAs are generally more stable (Connal et al., 2023). Circulating lncRNAs, including *PCA3*, may also be more stable in the bloodstream (Badowski et al., 2022). Additionally, *PCA3* and *SChLAP1* have been

successfully detected in post-DRE urine samples (Hessels et al., 2003; Prensner et al., 2014). As *EPCART* is expressed at similar levels as these lncRNAs, it would be intriguing to study whether *EPCART* is also expressed in blood or post-DRE urine samples. Certain ncRNAs, including lncRNAs, can be selectively packed into exosomes and exported outside the cell and thus can be found in liquid biopsies (Tellez-Gabriel & Heymann, 2019). As *EPCART* is cytoplasmic, it may also be found in exosomes, which remains to be tested.

The use of lncRNAs in cancer therapies has been less studied, but similar strategies used for other ncRNAs have been tested for lncRNAs in certain cancers (Mercer et al., 2022). The most studied therapeutic approach utilizing lncRNAs is targeted silencing by siRNAs or ASOs, although off-target effects and inefficient delivery of oligos inside cells are issues (Jackson & Linsley, 2010; Mercer et al., 2022; Qian et al., 2020; Yoshida et al., 2019). In our TPCAT knockdown studies, we found considerable off-target effects for TPCATs; thus, the selection of sufficiently specific oligonucleotides that would not cause adverse effects would be challenging. Alternatively, small molecules that bind lncRNA secondary structures and disrupt their interactions with proteins have been investigated; for example, the small molecule NP-C86 binds to lncRNA *GAS5* and the protein UPF1 (Mercer et al., 2022; Shi et al., 2019). For these types of treatments to be effective and safe, a deeper understanding of the structure and function of lncRNAs is needed, which is still lacking for most lncRNAs.

6.4 Challenges of studying novel lncRNAs (*i.e.*, why it took a decade to get here)

LncRNAs are a relatively new group of ncRNAs, as most of them have been discovered by means of whole-transcriptome sequencing studies in the last 15 years. LncRNAs are also unique among ncRNAs because of their diverse size, quantity, and function. Most other ncRNAs are small, under 200 nt long (except ribosomal RNAs, which are 1.9 and 5 kb long in humans) and have distinct mechanisms of function; lncRNAs are relatively long (typically 1–10 kb long) and can have many functions in cells. Various lncRNAs are also expressed at low levels and only in certain cell or tissue types. These qualities, combined with the fact that much of the research in the past has concentrated on studying proteins as functional players in cells and confined RNAs as molecules serving the central dogma of molecular

biology (DNA → RNA → protein), have created many challenges for studying lncRNAs.

The most apparent difficulty when studying lncRNAs is the lack of suitable methodologies. Most of the techniques used in cell and molecular biology are optimized for protein-coding genes, such as expression vectors, which include sequences that are not translated but are present in the transcripts; expression vectors, which include sequences that are translated and used as tags; loss-of-function studies, which only mutate single bases in ORFs; and the use of antibodies for the detection of proteins in various types of samples. As RNA transcripts are often the functional units of lncRNAs, any extra sequences added to them remain in the transcripts and are not translated, ensuring that they are nonfunctional rather than lacking a few nucleotides, and the use of antibody-based technologies is not a feasible option. Due to their large molecular size, the structure of lncRNAs is also incredibly laborious to study with conventional techniques, such as X-ray crystallography. These limitations are why there is very little information about the secondary or tertiary structures of lncRNAs and how they interact with other molecules, which makes it very difficult to know what part of the lncRNA sequence is necessary for its function and how it is altered. This also influences how we usually assess the homology of sequences; conserved sequences in mRNAs are not fully comparable to conserved sequences in lncRNAs, as evolutionary pressure is markedly different between these sequences. Essentially, when you do not know what you do not know, studying novel subjects can become difficult.

In our own studies, we needed to address many of these obstacles. The study of unannotated lncRNAs has led to additional challenges. For example, all the easily accessible online tools for studying genomes and transcripts, such as cBioPortal, were mostly useless because TPCATs were not found in their systems, which meant that all the computational analyses regarding TPCATs needed to be performed by us. Even after annotating the novel genes to the HUGO Gene Nomenclature Committee, it takes years before the information is updated into those tools. However, most of the difficulties were encountered while executing functional studies. Many TPCATs are expressed at much lower levels in cell lines than in many tissue samples, which demands the selection of an accurate *in vitro* model. Creating overexpression models was problematic due to prominent variation at the 5´ end, which is why we synthesized the full *EPCART* transcript (excluding exon 1) based on the sequence available from RNA-seq. In addition, the transcript was moved to an expression vector (pcDNA3.1), which is optimized for mRNAs and thus contains additional sequences that were added to the *EPCART* transcript. Therefore, the total

sequence of the produced *EPCART* transcript might not be fully functional because its structural properties and determinants are currently unknown. However, the greatest challenge was caused by TPCAT knockdown. We detected severe off-target effects for all studied TPCATs, including *PCAT5* and *EPCART*, with several siRNAs, even at low (0.1 nM) siRNA concentrations. The same effects were also detected in studies in which *EPCART* was knocked down with ASOs. The off-target effects of siRNAs and ASOs have been studied by several groups, and the severity of these effects seems to depend on the degree of complementarity between off-target sites and siRNA/ASO sequences (Jackson & Linsley, 2010; Yoshida et al., 2019). The siRNAs and ASOs used in our experiments all targeted different parts of the transcripts, and the sequences were also cross-checked by BLAST homology searches against RefSeq RNA sequences. By comparing this information against the cell viability data, we could not find an explanation for this phenomenon. One possibility is that these siRNAs bind TE sequences that are repeated throughout the genome and are very common in lncRNA sequences (Kapusta et al., 2013), as well as in TPCATs. As a BLAST search was performed against RefSeq RNA sequences, it excluded many unannotated ncRNAs that could have a functional role in cells. While we studied only a few TPCATs, our results raise the question of how common these off-target effects are among lncRNAs. In the literature, a great majority of the functional lncRNA gene silencing studies have been performed by siRNA knockdown experiments (Mitobe et al., 2018), including our *PCAT5* studies in publication I, and few of them have required control experiments in cells with no lncRNA expression. Hence, to validate the results of these studies, experiments with better controls and/or alternative methods for silencing are needed.

An alternative to knockdown experiments is the use of techniques such as CRISPR-Cas9, which alter the DNA of the gene to generate gene knockouts. In the case of lncRNAs, a single nucleotide mutation is most likely not sufficient to cause significant changes in lncRNA function, meaning that alterations in a larger area are needed. The most commonly used approaches for creating functional lncRNA-Kos are removing the promoter and the first exon of lncRNAs, introducing pretermination poly(A) signals, and deleting the entire lncRNA locus (Lyu et al., 2023). For the *EPCART*-KO model, we chose the first option (and deleted the nearby second exon), which decreased the expression of *EPCART* to very low levels but still allowed the expression of a potentially shorter transcript. In addition, one of the two KO clones (del-56) had an inversion in some of its chromosomes. Whether these events have a significant effect on the function of *EPCART* is unclear, as we

do not know which sequences, at either the DNA or the RNA level, have functional significance.

For the past 15 years, a great deal of methodological improvement has taken place in techniques for studying RNA functions. Among other methods, RNA ISH (fluorescent and chromogenic) for RNA localization studies, RNA immunoprecipitation techniques for finding specific RNA targets (DNA, RNA or protein), CRISPR interference methods for more specific knockdown experiments, and techniques for studying RNA secondary structures have been developed (Wang & Chekanova, 2019). Although these methods have provided much-needed information about the functional mechanisms of lncRNAs, the majority of these methods are laborious and expensive, which is why only a small percentage of the functional lncRNA studies published to date have used these methods. In other words, the lack of easily accessible technologies has shaped our knowledge about lncRNA functions. For example, in many PC-related lncRNA reviews, a great proportion of lncRNAs function through miRNAs (Mirzaei et al., 2022; Mitobe et al., 2018). However, finding miRNA-associated lncRNAs is relatively easy. As with any technology, as time passes and progress progresses, these technologies will eventually become more affordable, and our knowledge of lncRNAs will increase. Moreover, there is still a great need for new RNA-based technologies. It is especially important to study the tertiary structures of lncRNAs to gain a better mechanistic understanding of their functions. In the protein world, artificial intelligence has provided much-needed help for solving previously unsolved protein structures (Tunyasuvunakool et al., 2021). This has been enabled by the thousands of presolved structures available for proteins that can be used to train computational models, while such information sources for RNA do not exist. Therefore, although artificial intelligence may be incredibly useful in the long run, experimental research is still needed to elucidate the mysteries of lncRNA structures.

7 SUMMARY AND CONCLUSIONS

The studies that comprise this dissertation revealed over one hundred novel lncRNAs that were aberrantly expressed in prostate cancer (TPCATs) and assessed their regulatory potential, biomarker potential, and function in PC for selected lncRNAs (primarily *EPCART*). The following conclusions can be drawn from their findings:

- The majority of TPCATs are overexpressed in PC. Additionally, one-fourth of TPCATs are specifically expressed in CRPC.

- High expression levels of *EPCART*, *TPCAT-3-174133*, and *TPCAT-18-31849* are associated with the biochemical recurrence of PC. All three TPCATs also had independent prognostic value. In addition, *EPCART* is expressed abundantly and specifically in PC and thus has the potential to be used in both PC diagnostics and prognostics.

- The aberrant expression of several TPCATs, including *EPCART*, can be explained by the dysregulation of PC-specific TFs, especially AR and ERG. These TFs often bind to the proximal promoters of TPCATs.

- *EPCART* promotes the growth and migration of PC cells, accomplished, at least in part, via the modulation of protein translation through the PI3K/AKT/mTORC1/PDCD4 pathway.

These findings improve our knowledge about how lncRNAs are regulated and function in PC. These studies have also identified novel prospective markers and molecular targets for the detection and treatment of advanced PC, respectively. Future research is required to reveal further molecular mechanisms behind their functions and to study their biomarker potential in larger cohorts or in liquid biopsies.

REFERENCES

Abate-Shen, C., & Shen, M. M. (2000). Molecular genetics of prostate cancer. *Genes & Development*, *14*(19), 2410–2434. https://doi.org/10.1101/GAD.819500

Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., Cieslik, M., Benelli, M., Robinson, D., Van Allen, E. M., Sboner, A., Fedrizzi, T., Mosquera, J. M., Robinson, B. D., De Sarkar, N., Kunju, L. P., Tomlins, S., Wu, Y. M., Nava Rodrigues, D., … Sawyers, C. L. (2019). Genomic correlates of clinical outcome in advanced prostate cancer. *Proceedings of the National Academy of Sciences*, *116*(23), 11428–11436. https://doi.org/10.1073/pnas.1902651116

Ahearn, T. U., Pettersson, A., Ebot, E. M., Gerke, T., Graff, R. E., Morais, C. L., Hicks, J. L., Wilson, K. M., Rider, J. R., Sesso, H. D., Fiorentino, M., Flavin, R., Finn, S., Giovannucci, E. L., Loda, M., Stampfer, M. J., De Marzo, A. M., Mucci, L. A., & Lotan, T. L. (2016). A Prospective Investigation of PTEN Loss and ERG Expression in Lethal Prostate Cancer. *JNCI: Journal of the National Cancer Institute*, *108*(2), djv346. https://doi.org/10.1093/jnci/djv346

Amling, C. L., Bergstralh, E. J., Blute, M. L., Slezak, J. M., & Zincke, H. (2001). Defining prostate specific antigen progression after radical prostatectomy: what is the most appropriate cut point? *The Journal of Urology*, *165*(4), 1146–1151. https://doi.org/10.1016/S0022-5347(05)66452-X

Anderson, D. M., Anderson, K. M., Chang, C.-L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., & Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, *160*(4), 595–606. https://doi.org/10.1016/j.cell.2015.01.009

Artemenko, M., Zhong, S. S. W., To, S. K. Y., & Wong, A. S. T. (2022). p70 S6 kinase as a therapeutic target in cancers: More than just an mTOR effector. *Cancer Letters*, *535*, 215593. https://doi.org/10.1016/j.canlet.2022.215593

Aw, J. G. A., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K.-L., Tapsin, S., Chan, Y.-S., Tan, C.-P., Sim, A. Y. L., Zhang, T., Susanto, T. T., Fu, Z., Nagarajan, N., & Wan, Y. (2016). In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell*, *62*(4), 603–617. https://doi.org/10.1016/j.molcel.2016.04.028

Aytes, A., Mitrofanova, A., Kinkade, C. W., Lefebvre, C., Lei, M., Phelan, V., LeKaye, H. C., Koutcher, J. A., Cardiff, R. D., Califano, A., Shen, M. M., & Abate-Shen, C. (2013). ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(37), 3506. https://doi.org/10.1073/pnas.1303558110

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., & Wishart, D. S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Research*, *44*(W1), W147-153. https://doi.org/10.1093/nar/gkw419

Baca, S. C., Prandi, D., Lawrence, M. S., Mosquera, J. M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T. Y., Ghandi, M., Allen, E. V., Kryukov, G. V., Sboner, A., Theurillat, J.-P., Soong, T. D., Nickerson, E., Auclair, D., Tewari, A., Beltran, H., … Garraway, L. A. (2013). Punctuated evolution of prostate cancer genomes. *Cell*, *153*(3), 666–677. https://doi.org/10.1016/j.cell.2013.03.021

Badowski, C., He, B., & Garmire, L. X. (2022). Blood-derived lncRNAs as biomarkers for cancer diagnosis: the Good, the Bad and the Beauty. *Npj Precision Oncology*, *6*(1), Article 1. https://doi.org/10.1038/s41698-022-00283-7

Barbieri, C. E., & Rubin, M. A. (2015). Genomic rearrangements in prostate cancer. *Current Opinion in Urology*, *25*(1), 71–76. https://doi.org/10.1097/MOU.0000000000000129

Barbieri, I., & Kouzarides, T. (2020). Role of RNA modifications in cancer. *Nature Reviews Cancer 2020 20:6*, *20*(6), 303–322. https://doi.org/10.1038/s41568-020-0253-2

Barron, D. A., Rowley, D. R., Barron, A., & Rowley, D. R. (2012). The reactive stroma microenvironment and prostate cancer progression. *Endocrine-Related Cancer*, *19*(6), R187–R204. https://doi.org/10.1530/ERC-12-0085

Baylin, S. B., & Jones, P. A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harbor Perspectives in Biology*, *8*(9), a019505. https://doi.org/10.1101/cshperspect.a019505

Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & Development*, *23*(7), 781–783. https://doi.org/10.1101/gad.1787609

Bhartiya, D., Talwar, J., Hasija, Y., & Scaria, V. (2012). Systematic curation and analysis of genomic variations and their potential functional consequences in snoRNA loci. *Human Mutation*, *33*(10), E2367-2374. https://doi.org/10.1002/humu.22158

Bismar, T. A., Yoshimoto, M., Vollmer, R. T., Duan, Q., Firszt, M., Corcos, J., & Squire, J. A. (2011). PTEN genomic deletion is an early event associated with ERG gene rearrangements in prostate cancer. *BJU International*, *107*(3), 477–485. https://doi.org/10.1111/j.1464-410X.2010.09470.x

Blakely, E. L., Yarham, J. W., Alston, C. L., Craig, K., Poulton, J., Brierley, C., Park, S.-M., Dean, A., Xuereb, J. H., Anderson, K. N., Compston, A., Allen, C., Sharif, S., Enevoldson, P., Wilson, M., Hammans, S. R., Turnbull, D. M., McFarland, R., & Taylor, R. W. (2013). Pathogenic mitochondrial tRNA point mutations: nine novel mutations affirm their importance as a cause of mitochondrial disease. *Human Mutation*, *34*(9), 1260–1268. https://doi.org/10.1002/humu.22358

Böttcher, R., Hoogland, A. M., Dits, N., Verhoef, E. I., Kweldam, C., Waranecki, P., Bangma, C. H., van Leenders, G. J. L. H., & Jenster, G. (2015). Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. *Oncotarget*, *6*(6), 4036–4050. https://doi.org/10.18632/oncotarget.2879

Bowen, C., Bubendorf, L., Voeller, H. J., Slack, R., Willi, N., Sauter, G., Gasser, T. C., Koivisto, P., Lack, E. E., Kononen, J., Kallioniemi, O. P., & Gelmann, E. P. (2000). Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. *Cancer Research*, *60*(21), 6111–6115.

Braadland, P. R., & Urbanucci, A. (2019). Chromatin reprogramming as an adaptation mechanism in advanced prostate cancer. *Endocrine-Related Cancer*, *26*(4), R211–R235. https://doi.org/10.1530/ERC-18-0579

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436. https://doi.org/10.1038/sj.bjc.6601119

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Brierley, J. D., Gospodarowicz, M. K., Wittekind, C., & the Union for International Cancer Control. (2017). *TNM Classification of Malignant Tumours*. Wiley-Blackwell.

Bussemakers, M. J., Bokhoven, A. van, Verhaegh, G. W., Smit, F. P., Karthaus, H. F., Schalken, J. A., Debruyne, F. M., Ru, N., & Isaacs, W. B. (1999). DD3: A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer. *Cancer Research*, *59*(23), 5975–5979.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927. https://doi.org/10.1101/gad.17446611

Cai, C., Wang, H., He, H. H., Chen, S., He, L., Ma, F., Mucci, L., Wang, Q., Fiore, C., Sowalsky, A. G., Loda, M., Liu, X. S., Brown, M., Balk, S. P., & Yuan, X. (2013). ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. *The Journal of Clinical Investigation*, *123*(3), 1109–1122. https://doi.org/10.1172/JCI66666 [doi]

Cai, C., Wang, H., Xu, Y., Chen, S., & Balk, S. P. (2009). Reactivation of androgen receptor-regulated TMPRSS2:ERG gene expression in castration-resistant prostate cancer. *Cancer Research*, *69*(15), 6027–6032. https://doi.org/10.1158/0008-5472.CAN-09-0395

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A., & Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA (New York, N.Y.)*, *22*(6), 867–882. https://doi.org/10.1261/rna.053561.115

Carver, B. S., Chapinski, C., Wongvipat, J., Hieronymus, H., Chen, Y., Chandarlapaty, S., Arora, V. K., Le, C., Koutcher, J., Scher, H., Scardino, P. T., Rosen, N., & Sawyers, C. L. (2011). Reciprocal Feedback Regulation of PI3K and Androgen Receptor Signaling in PTEN-Deficient Prostate Cancer. *Cancer Cell*, *19*(5), 575–586. https://doi.org/10.1016/j.ccr.2011.04.008

Carver, B. S., Tran, J., Chen, Z., Carracedo-Perez, A., Alimonti, A., Nardella, C., Gopalan, A., Scardino, P. T., Cordon-Cardo, C., Gerald, W., & Pandolfi, P. P. (2009). ETS rearrangements and prostate cancer initiation. *Nature*, *457*(7231), E1; discussion E2-3. https://doi.org/10.1038/nature07738

Catalona, W. J., Partin, A. W., Slawin, K. M., Brawer, M. K., Flanigan, R. C., Patel, A., Richie, J. P., deKernion, J. B., Walsh, P. C., Scardino, P. T., Lange, P. H., Subong, E. N., Parson, R. E., Gasior, G. H., Loveland, K. G., & Southwick, P. C. (1998). Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. *JAMA*, *279*(19), 1542–1547. https://doi.org/10.1001/jama.279.19.1542

Catalona, W. J., Richie, J. P., Ahmann, F. R., Hudson, M. A., Scardino, P. T., Flanigan, R. C., DeKernion, J. B., Ratliff, T. L., Kavoussi, L. R., Dalkin, B. L., Waters, W. B., MacFarlane, M. T., & Southwick, P. C. (1994). Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *The Journal of Urology*, *151*(5), 1283–1290. https://doi.org/10.1016/S0022-5347(17)35233-3

Catena, V., & Fanciulli, M. (2017). Deptor: not only a mTOR inhibitor. *Journal of Experimental & Clinical Cancer Research : CR*, *36*, 12. https://doi.org/10.1186/s13046-016-0484-y

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, *58*(5), 235–263. https://doi.org/10.1002/EM.22087

Chaux, A., Peskoe, S. B., Gonzalez-Roibon, N., Schultz, L., Albadine, R., Hicks, J., De Marzo, A. M., Platz, E. A., & Netto, G. J. (2012). Loss of PTEN expression is associated with increased risk of recurrence after prostatectomy for clinically localized prostate cancer. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, *25*(11), 1543–1549. https://doi.org/10.1038/modpathol.2012.104

Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A., & Kuiper, M. (2013). TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics (Oxford, England)*, *29*(19), 2519–2520. https://doi.org/10.1093/bioinformatics/btt432

Chen, Y., Chi, P., Rockowitz, S., Iaquinta, P. J., Shamu, T., Shukla, S., Gao, D., Sirota, I., Carver, B. S., Wongvipat, J., Scher, H. I., Zheng, D., & Sawyers, C. L. (2013). ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nature Medicine 2013 19:8*, *19*(8), 1023–1029. https://doi.org/10.1038/nm.3216

Chng, K. R., Chang, C. W., Tan, S. K., Yang, C., Hong, S. Z., Sng, N. Y. W., & Cheung, E. (2012). A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. *The EMBO Journal*, *31*(12), 2810–2823. https://doi.org/10.1038/emboj.2012.112

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., Nevins, S. A., Chin, S.-F., Caldas, C., Liu, S. J., Horlbeck, M. A., Lim, D. A., Weissman, J. S., Curtis, C., & Chang, H. Y. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell*, *173*(6), 1398-1412.e22. https://doi.org/10.1016/j.cell.2018.03.068

Chung, S., Nakagawa, H., Uemura, M., Piao, L., Ashikawa, K., Hosono, N., Takata, R., Akamatsu, S., Kawaguchi, T., Morizono, T., Tsunoda, T., Daigo, Y., Matsuda, K., Kamatani, N., Nakamura, Y., & Kubo, M. (2011). Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Science*, *102*(1), 245–252. https://doi.org/10.1111/j.1349-7006.2010.01737.x

Claessens, F., Joniau, S., & Helsen, C. (2017). Comparing the rules of engagement of androgen and glucocorticoid receptors. *Cellular and Molecular Life Sciences : CMLS*, *74*(12), 2217–2228. https://doi.org/10.1007/s00018-017-2467-3

Clark, J., Merson, S., Jhavar, S., Flohr, P., Edwards, S., Foster, C. S., Eeles, R., Martin, F. L., Phillips, D. H., Crundwell, M., Christmas, T., Thompson, A., Fisher, C., Kovacs, G., & Cooper, C. S. (2007). Diversity of TMPRSS2-ERG fusion transcripts in the human prostate. *Oncogene*, *26*(18), 2667–2673. https://doi.org/10.1038/sj.onc.1210070

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232–238. https://doi.org/10.1038/sj.bjc.6601118

Coco, C., Sgarra, L., Potenza, M. A., Nacci, C., Pasculli, B., Barbano, R., Parrella, P., & Montagnani, M. (2019). Can Epigenetics of Endothelial Dysfunction Represent the Key to Precision Medicine in Type 2 Diabetes Mellitus? *International Journal of Molecular Sciences*, *20*(12), Article 12. https://doi.org/10.3390/ijms20122949

Connal, S., Cameron, J. M., Sala, A., Brennan, P. M., Palmer, D. S., Palmer, J. D., Perlow, H., & Baker, M. J. (2023). Liquid biopsies: the future of cancer early detection. *Journal of Translational Medicine*, *21*, 118. https://doi.org/10.1186/s12967-023-03960-8

Copeland, B. T., Pal, S. K., Bolton, E. C., & Jones, J. O. (2018). The androgen receptor malignancy shift in prostate cancer. *The Prostate*, *78*(7), 521–531. https://doi.org/10.1002/pros.23497

Crea, F., Watahiki, A., Quagliata, L., Xue, H., Pikor, L., Parolia, A., Wang, Y., Lin, D., Lam, W. L., Farrar, W. L., Isogai, T., Morant, R., Castori-Eppenberger, S., Chi, K. N., Wang, Y., & Helgason, C. D. (2014). Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. *Oncotarget*, *5*(3), 764–774. https://doi.org/10.18632/oncotarget.1769

Dadaev, T., Saunders, E. J., Newcombe, P. J., Anokian, E., Leongamornlert, D. A., Brook, M. N., Cieza-Borrella, C., Mijuskovic, M., Wakerell, S., Olama, A. A. A., Schumacher, F. R., Berndt, S. I., Benlloch, S., Ahmed, M., Goh, C., Sheng, X., Zhang, Z., Muir, K., Govindasami, K., … Kote-Jarai, Z. (2018). Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nature Communications*, *9*(1), 1–19. https://doi.org/10.1038/s41467-018-04109-8

Dai, C., Heemers, H., & Sharifi, N. (2017). Androgen Signaling in Prostate Cancer. *Cold Spring Harbor Perspectives in Medicine*, *7*(9). https://doi.org/10.1101/cshperspect.a030452

D'Amico, A. V., Whittington, R., Malkowicz, S. B., Schultz, D., Blank, K., Broderick, G. A., Tomaszewski, J. E., Renshaw, A. A., Kaplan, I., Beard, C. J., & Wein, A. (1998). Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*, *280*(11), 969–974. https://doi.org/10.1001/jama.280.11.969

Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., Narayanan, A. K., Onate, K. C., Graham, K., Miyasato, S. R., Dreszer, T. R., Strattan, J. S., Jolanki, O., Tanaka, F. Y., & Cherry, J. M. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research*, *46*(D1), D794–D801. https://doi.org/10.1093/NAR/GKX1081

de Kok, J. B., Verhaegh, G. W., Roelofs, R. W., Hessels, D., Kiemeney, L. A., Aalders, T. W., Swinkels, D. W., & Schalken, J. A. (2002). DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *CANCER RESEARCH*, *62*(9), 2695–2698.

Dehm, S. M., & Tindall, D. J. (2011). Alternatively spliced androgen receptor variants. *Endocrine-Related Cancer*, *18*(5), 183. https://doi.org/10.1530/ERC-11-0141

Deniz, E., & Erman, B. (2017). Long noncoding RNA (lincRNA), a new paradigm in gene expression control. *Functional & Integrative Genomics*, *17*(2), 135–143. https://doi.org/10.1007/s10142-016-0524-x

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/gr.132159.111

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature (London)*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

D'Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., Budnik, B. A., Lykke-Andersen, J., Saghatelian, A., & Slavoff, S. A. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nature Chemical Biology*, *13*(2), 174–180. https://doi.org/10.1038/nchembio.2249

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Donovan, J. L., Hamdy, F. C., Lane, J. A., Mason, M., Metcalfe, C., Walsh, E., Blazeby, J. M., Peters, T. J., Holding, P., Bonnington, S., Lennon, T., Bradshaw, L., Cooper, D., Herbert, P., Howson, J., Jones, A., Lyons, N., Salter, E., Thompson, P., … Neal, D. E. (2016). Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *The New England Journal of Medicine*, *375*(15), 1425–1437. https://doi.org/10.1056/NEJMoa1606221

Du, Z., Fei, T., Verhaak, R. G. W., Su, Z., Zhang, Y., Brown, M., Chen, Y., & Liu, X. S. (2013). Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nature Structural & Molecular Biology*, *20*(7), 908–913. https://doi.org/10.1038/nsmb.2591

Duffy, M. J. (2020). Biomarkers for prostate cancer: Prostate-specific antigen and beyond. *Clinical Chemistry and Laboratory Medicine*, *58*(3), 326–339. https://doi.org/10.1515/CCLM-2019-0693/XML

Dyba, T., Randi, G., Bray, F., Martos, C., Giusti, F., Nicholson, N., Gavin, A., Flego, M., Neamtiu, L., Dimitrova, N., Negrão Carvalho, R., Ferlay, J., & Bettio, M. (2021). The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. *European Journal of Cancer*, *157*, 308–347. https://doi.org/10.1016/j.ejca.2021.07.039

EAU. (2023). *EAU Guidelines. Edn. presented at the EAU Annual Congress Milan 2023*. EAU Guidelines Office, Arnhem, The Netherlands.

Ellinger, J., Kahl, P., von der Gathen, J., Rogenhofer, S., Heukamp, L. C., Gütgemann, I., Walter, B., Hofstädter, F., Büttner, R., Müller, S. C., Bastian, P. J., & von Ruecker, A. (2010). Global levels of histone modifications predict prostate cancer recurrence. *The Prostate*, *70*(1), 61–69. https://doi.org/10.1002/pros.21038

EMA. (2023a, February 23). *European Medicines Agency: Akeega - niraparib / abiraterone acetate*. https://www.ema.europa.eu/en/medicines/human/EPAR/akeega

EMA. (2023b, November 9). *European Medicines Agency: Talzenna - opinion on variation to marketing authorisation*. https://www.ema.europa.eu/en/medicines/human/variation/talzenna

Eminaga, O., Hinkelammert, R., Abbas, M., Titze, U., Eltze, E., Bettendorf, O., & Semjonow, A. (2013). High-Grade Prostatic Intraepithelial Neoplasia (HGPIN) and topographical distribution in 1,374 prostatectomy specimens: Existence of HGPIN near prostate cancer. *The Prostate*, *73*(10), 1115–1122. https://doi.org/10.1002/pros.22660

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., & Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature (London)*, *539*(7629), 452–455. https://doi.org/10.1038/nature20149

Epstein, J. I. (2011). Prognostic significance of tumor volume in radical prostatectomy and needle biopsy specimens. *The Journal of Urology*, *186*(3), 790–797. https://doi.org/10.1016/j.juro.2011.02.2695

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., Sun, L., Zhang, M. Q., Chen, R., & Zhao, Y. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Research*, *46*(D1), D308–D314. https://doi.org/10.1093/NAR/GKX1107

Felsenfeld, G. (2014). A Brief History of Epigenetics. *Cold Spring Harbor Perspectives in Biology*, *6*(1), a018200. https://doi.org/10.1101/cshperspect.a018200

Fenton, J. J., Weyrich, M. S., Durbin, S., Liu, Y., Bang, H., & Melnikow, J. (2018). Prostate-Specific Antigen–Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*, *319*(18), 1914–1931. https://doi.org/10.1001/JAMA.2018.3712

Foley, C., & Mitsiades, N. (2016). Moving Beyond the Androgen Receptor (AR): Targeting AR-Interacting Proteins to Treat Prostate Cancer. *Hormones & Cancer*, *7*(2), 84–103. https://doi.org/10.1007/s12672-015-0239-9

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Sala, S. C., Chrast, J., Cunningham, F., Domenico, T. D., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766–D773. https://doi.org/10.1093/nar/gky955

Fraser, M., & Rouette, A. (2019). Prostate Cancer Genomic Subtypes. *Advances in Experimental Medicine and Biology*, *1210*, 87–110. https://doi.org/10.1007/978-3-030-32656-2_5

Fraser, M., Sabelnykova, V. Y., Yamaguchi, T. N., Heisler, L. E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A. P., Fox, N. S., Xie, M., Prokopec, S. D., Berlin, A., Lalonde, E., Ahmed, M., Trudel, D., Luo, X., Beck, T. A., … Boutros, P. C. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. *Nature*, *541*(7637), 359–364. https://doi.org/10.1038/nature20788

Fruman, D. A., Chiu, H., Hopkins, B. D., Bagrodia, S., Cantley, L. C., & Abraham, R. T. (2017). The PI3K Pathway in Human Disease. *Cell*, *170*(4), 605–635. https://doi.org/10.1016/j.cell.2017.07.029

Gao, P., Xia, J.-H., Sipeky, C., Dong, X.-M., Zhang, Q., Yang, Y., Zhang, P., Cruz, S. P., Zhang, K., Zhu, J., Lee, H.-M., Suleman, S., Giannareas, N., Liu, S., Tammela, T. L. J., Auvinen, A., Wang, X., Huang, Q., Wang, L., … Wei, G.-H. (2018). Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell*, *174*(3), 576-589.e18. https://doi.org/10.1016/j.cell.2018.06.003

Gittelman, M. C., Hertzman, B., Bailen, J., Williams, T., Koziol, I., Henderson, R. J., Efros, M., Bidair, M., & Ward, J. F. (2013). PCA3 molecular urine test as a predictor of repeat

prostate biopsy outcome in men with previous negative biopsies: a prospective multicenter clinical study. *The Journal of Urology*, *190*(1), 64–69. https://doi.org/10.1016/J.JURO.2013.02.018

Glaviano, A., Foo, A. S. C., Lam, H. Y., Yap, K. C. H., Jacot, W., Jones, R. H., Eng, H., Nair, M. G., Makvandi, P., Geoerger, B., Kulke, M. H., Baird, R. D., Prabhu, J. S., Carbone, D., Pecoraro, C., Teh, D. B. L., Sethi, G., Cavalieri, V., Lin, K. H., … Kumar, A. P. (2023). PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. *Molecular Cancer*, *22*(1), 138. https://doi.org/10.1186/s12943-023-01827-6

Gleason, D. F. (1966). Classification of prostatic carcinomas. *Cancer Chemotherapy Reports*, *50*(3), 125–128.

Gong, C., & Maquat, L. E. (2011). lncRNAs transactivate Staufen1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. *Nature*, *470*(7333), 284–288. https://doi.org/10.1038/nature09701

Graff, J. R., Konicek, B. W., Lynch, R. L., Dumstorf, C. A., Dowless, M. S., McNulty, A. M., Parsons, S. H., Brail, L. H., Colligan, B. M., Koop, J. W., Hurst, B. M., Deddens, J. A., Neubauer, B. L., Stancato, L. F., Carter, H. W., Douglass, L. E., & Carter, J. H. (2009). eIF4E activation is commonly elevated in advanced human prostate cancers and significantly related to reduced patient survival. *Cancer Research*, *69*(9), 3866–3873. https://doi.org/10.1158/0008-5472.CAN-08-3472

Graham, L. S., & Schweizer, M. T. (2022). Mismatch repair deficiency and clinical implications in prostate cancer. *The Prostate*, *82*(S1), S37–S44. https://doi.org/10.1002/pros.24343

Greten, F. R., & Grivennikov, S. I. (2019). Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity (Cambridge, Mass.)*, *51*(1), 27–41. https://doi.org/10.1016/j.immuni.2019.06.025

Groelly, F. J., Fawkes, M., Dagg, R. A., Blackford, A. N., & Tarsounas, M. (2023). Targeting DNA damage response pathways in cancer. *Nature Reviews Cancer*, *23*(2), Article 2. https://doi.org/10.1038/s41568-022-00535-5

Groff, A. F., Sanchez-Gomez, D. B., Soruco, M. M. L., Gerhardinger, C., Barutcu, A. R., Li, E., Elcavage, L., Plana, O., Sanchez, L. V., Lee, J. C., Sauvageau, M., & Rinn, J. L. (2016). In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements. *Cell Reports*, *16*(8), 2178–2186. https://doi.org/10.1016/j.celrep.2016.07.050

Groskopf, J., Aubin, S. M., Deras, I. L., Blase, A., Bodrug, S., Clark, C., Brentano, S., Mathis, J., Pham, J., Meyer, T., Cass, M., Hodge, P., Macairan, M. L., Marks, L. S., & Rittenhouse, H. (2006). APTIMA PCA3 Molecular Urine Test: Development of a Method to Aid in the Diagnosis of Prostate Cancer. *Clinical Chemistry*, *52*(6), 1089–1095. https://doi.org/10.1373/CLINCHEM.2005.063289

Guo, H., Ahmed, M., Zhang, F., Yao, C. Q., Li, S., Liang, Y., Hua, J., Soares, F., Sun, Y., Langstein, J., Li, Y., Poon, C., Bailey, S. D., Desai, K., Fei, T., Li, Q., Sendorek, D. H., Fraser, M., Prensner, J. R., … He, H. H. (2016). Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nature Genetics*, *48*(10), 1142–1150. https://doi.org/10.1038/ng.3637

Guo, W., Li, L., He, J., Liu, Z., Han, M., Li, F., Xia, X., Zhang, X., Zhu, Y., Wei, Y., Li, Y., Aji, R., Dai, H., Wei, H., Li, C., Chen, Y., Chen, L., & Gao, D. (2020). Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nature Genetics*, *52*(9), 908–918. https://doi.org/10.1038/s41588-020-0642-1

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B. W., Cassady, J. P., Cabili, M. N., Jaenisch, R., Mikkelsen, T. S., Jacks, T., Hacohen, N., Bernstein, B. E., Kellis, M., Regev, A., Rinn, J. L., & Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227. https://doi.org/10.1038/nature07672

Haffner, M. C., Aryee, M. J., Toubaji, A., Esopi, D. M., Albadine, R., Gurel, B., Isaacs, W. B., Bova, G. S., Liu, W., Xu, J., Meeker, A. K., Netto, G., Marzo, A. M. D., Nelson, W. G., & Yegnasubramanian, S. (2010). Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nature Genetics*, *42*(8), 668–675. https://doi.org/10.1038/ng.613

Hägglöf, C., & Bergh, A. (2012). The Stroma—A Key Regulator in Prostate Function and Malignancy. *Cancers*, *4*(2), 531. https://doi.org/10.3390/CANCERS4020531

Hamdy, F. C., Donovan, J. L., Lane, J. A., Metcalfe, C., Davis, M., Turner, E. L., Martin, R. M., Young, G. J., Walsh, E. I., Bryant, R. J., Bollina, P., Doble, A., Doherty, A., Gillatt, D., Gnanapragasam, V., Hughes, O., Kockelbergh, R., Kynaston, H., Paul, A., … ProtecT Study Group. (2023). Fifteen-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *The New England Journal of Medicine*, *388*(17), 1547–1558. https://doi.org/10.1056/NEJMoa2214122

Han, B., Mehra, R., Lonigro, R. J., Wang, L., Suleman, K., Menon, A., Palanisamy, N., Tomlins, S. A., Chinnaiyan, A. M., & Shah, R. B. (2009). Fluorescence in situ hybridization study shows association of PTEN deletion with ERG rearrangement during prostate cancer progression. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, *22*(8), 1083–1093. https://doi.org/10.1038/modpathol.2009.69

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell (Cambridge)*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hankey, W., Chen, Z., & Wang, Q. (2020). Shaping Chromatin States in Prostate Cancer by Pioneer Transcription Factors. *Cancer Research*, *80*(12), 2427–2436. https://doi.org/10.1158/0008-5472.CAN-19-3447

Hawkes, E. J., Hennelly, S. P., Novikova, I. V., Irwin, J. A., Dean, C., & Sanbonmatsu, K. Y. (2016). COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. *Cell Reports*, *16*(12), 3087–3096. https://doi.org/10.1016/j.celrep.2016.08.045

He, F., Wei, R., Zhou, Z., Huang, L., Wang, Y., Tang, J., Zou, Y., Shi, L., Gu, X., Davis, M. J., & Su, Z. (2019). Integrative Analysis of Somatic Mutations in Non-coding Regions Altering RNA Secondary Structures in Cancer Genomes. *Scientific Reports*, *9*, 8205. https://doi.org/10.1038/s41598-019-44489-5

Henry, G. H., Malewska, A., Joseph, D. B., Malladi, V. S., Lee, J., Torrealba, J., Mauck, R. J., Gahan, J. C., Raj, G. V., Roehrborn, C. G., Hon, G. C., MacConmara, M. P., Reese, J. C., Hutchinson, R. C., Vezina, C. M., & Strand, D. W. (2018). A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Reports*, *25*(12), 3530-3542.e5. https://doi.org/10.1016/j.celrep.2018.11.086

Hessels, D., Klein Gunnewiek, J. M. T., Van Oort, I., Karthaus, H. F. M., Van Leenders, G. J. L., Van Balken, B., Kiemeney, L. A., Witjes, J. A., Schalken, J. A., & Culig, Z. (2003). DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *European Urology*, *44*(1), 8–16. https://doi.org/10.1016/S0302-2838(03)00201-X

Higgins, J., Brogley, M., Palanisamy, N., Mehra, R., Ittmann, M. M., Li, J. Z., Tomlins, S. A., & Robins, D. M. (2015). Interaction of the Androgen Receptor, ETV1, and PTEN Pathways in Mouse Prostate Varies with Pathological Stage and Predicts Cancer Progression. *Hormones & Cancer*, *6*(2–3), 67–86. https://doi.org/10.1007/s12672-014-0215-9

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T. S., Harte, R. A., Hsu, F., Hillman-Jackson, J., Kuhn, R. M., Pedersen, J. S., Pohl, A., Raney, B. J., Rosenbloom, K. R., Siepel, A., Smith, K. E., Sugnet, C. W., … Kent, W. J. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*, *34*(Database issue), D590-598. https://doi.org/10.1093/nar/gkj144

Hollenhorst, P. C., Chandler, K. J., Poulsen, R. L., Johnson, W. E., Speck, N. A., & Graves, B. J. (2009). DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genetics*, *5*(12), e1000778. https://doi.org/10.1371/journal.pgen.1000778 [doi]

Hollenhorst, P. C., Ferris, M. W., Hull, M. A., Chae, H., Kim, S., & Graves, B. J. (2011). Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes & Development*, *25*(20), 2147–2157. https://doi.org/10.1101/gad.17546311 [doi]

Hollenhorst, P. C., Jones, D. A., & Graves, B. J. (2004). Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Research*, *32*(18), 5693–5702. https://doi.org/10.1093/nar/gkh906

Hollenhorst, P. C., McIntosh, L. P., & Graves, B. J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual Review of Biochemistry*, *80*, 437–471. https://doi.org/10.1146/annurev.biochem.79.081507.103945

Hologic. (2019). Progensa PCA3 Assay [package insert]. In *502083-IFU-PI Rev. 003*. Hologic, Inc.

Hu, X., Wang, Q., Tang, M., Barthel, F., Amin, S., Yoshihara, K., Lang, F. M., Martinez-Ledesma, E., Lee, S. H., Zheng, S., & Verhaak, R. G. W. (2018). TumorFusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Research*, *46*(D1), D1144–D1149. https://doi.org/10.1093/nar/gkx1018

Hua, J. T., Ahmed, M., Guo, H., Zhang, Y., Chen, S., Soares, F., Lu, J., Zhou, S., Wang, M., Li, H., Larson, N. B., McDonnell, S. K., Patel, P. S., Liang, Y., Yao, C. Q., van der Kwast, T., Lupien, M., Feng, F. Y., Zoubeidi, A., … He, H. H. (2018). Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. *Cell*, *174*(3), 564-575.e18. https://doi.org/10.1016/j.cell.2018.06.014

Huang, J., Wu, C., di Sant'Agnese, P. A., Yao, J. L., Cheng, L., & Na, Y. (2007). Function and molecular mechanisms of neuroendocrine cells in prostate cancer. *Analytical and Quantitative Cytology and Histology*, *29*(3), 128–138.

Huggins, C., & Hodges, C. V. (1941). Studies on Prostatic Cancer. I. The Effect of Castration, of Estrogen and of Androgen Injection on Serum Phosphatases in Metastatic Carcinoma of the Prostate*. *Cancer Research*, *1*(4), 293–297.

Humphrey, P. A. (2004). Gleason grading and prognostic factors in carcinoma of the prostate. *Modern Pathology*, *17*(3), Article 3. https://doi.org/10.1038/modpathol.3800054

Hussain, M., Tangen, C. M., Higano, C., Schelhammer, P. F., Faulkner, J., Crawford, E. D., Wilding, G., Akdas, A., Small, E. J., Donnelly, B., MacVicar, G., & Raghavan, D. (2006). Absolute prostate-specific antigen value after androgen deprivation is a strong independent predictor of survival in new metastatic prostate cancer: data from Southwest Oncology Group Trial 9346 (INT-0162). *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *24*(24), 3984–3990. https://doi.org/10.1200/JCO.2006.06.4246

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. (2020). Pan-cancer analysis of whole genomes. *Nature*, *578*(7793), 82–93. https://doi.org/10.1038/s41586-020-1969-6

Iglesias-Gato, D., Thysell, E., Tyanova, S., Crnalic, S., Santos, A., Lima, T. S., Geiger, T., Cox, J., Widmark, A., Bergh, A., Mann, M., Flores-Morales, A., & Wikström, P. (2018). The Proteome of Prostate Cancer Bone Metastasis Reveals Heterogeneity with Prognostic Implications. *Clinical Cancer Research*, *24*(21), 5433–5444. https://doi.org/10.1158/1078-0432.CCR-18-1229

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. https://doi.org/10.1038/nature03001

Ittmann, M. (2018). Anatomy and Histology of the Human and Murine Prostate. *Cold Spring Harbor Perspectives in Medicine*, *8*(5), a030346. https://doi.org/10.1101/CSHPERSPECT.A030346

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T. R., Prensner, J. R., Evans, J. R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S. M., Wu, Y.-M., Robinson, D. R., Beer, D. G., Feng, F. Y., Iyer, H. K., & Chinnaiyan, A. M. (2015). The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nature Genetics*, *47*(3), 199–208. https://doi.org/10.1038/ng.3192

Jackson, A. L., & Linsley, P. S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews Drug Discovery*, *9*(1), Article 1. https://doi.org/10.1038/nrd3010

Jean, S., & Kiger, A. A. (2014). Classes of phosphoinositide 3-kinases at a glance. *Journal of Cell Science*, *127*(5), 923–928. https://doi.org/10.1242/jcs.093773

Jernberg, E., Bergh, A., & Wikström, P. (2017). Clinical relevance of androgen receptor alterations in prostate cancer. *Endocrine Connections*, *6*(8), R146–R161. https://doi.org/10.1530/EC-17-0118

Jin, G., Sun, J., Isaacs, S. D., Wiley, K. E., Kim, S.-T., Chu, L. W., Zhang, Z., Zhao, H., Zheng, S. L., Isaacs, W. B., & Xu, J. (2011). Human polymorphisms at long non-coding RNAs (lncRNAs) and association with prostate cancer risk. *Carcinogenesis*, *32*(11), 1655–1659. https://doi.org/10.1093/carcin/bgr187

Johnson, R., & Guigó, R. (2014). The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA (New York, N.Y.)*, *20*(7), 959–976. https://doi.org/10.1261/rna.044560.114

Kannan, K., Wang, L., Wang, J., Ittmann, M. M., Li, W., & Yen, L. (2011). Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9172–

9177. https://doi.org/10.1073/PNAS.1100489108/-/DCSUPPLEMENTAL/PNAS.201100489SI.PDF

Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., & Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genetics*, *9*(4), e1003470. https://doi.org/10.1371/journal.pgen.1003470

Karim, F. D., Urness, L. D., Thummel, C. S., Klemsz, M. J., McKercher, S. R., Celada, A., Beveren, C. V., Maki, R. A., Gunther, C. V., & Nye, J. A. (1990). The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes & Development*, *4*(9), 1451–1453. https://doi.org/10.1101/gad.4.9.1451

Karthaus, W. R., Hofree, M., Choi, D., Linton, E. L., Turkekul, M., Bejnood, A., Carver, B., Gopalan, A., Abida, W., Laudone, V., Biton, M., Chaudhary, O., Xu, T., Masilionis, I., Manova, K., Mazutis, L., Pe'er, D., Regev, A., & Sawyers, C. L. (2020). Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science (New York, N.Y.)*, *368*(6490), 497–505. https://doi.org/10.1126/science.aay0267

Kedage, V., Selvaraj, N., Nicholas, T. R., Budka, J. A., Plotnik, J. P., Jerde, T. J., & Hollenhorst, P. C. (2016). An Interaction with Ewing's Sarcoma Breakpoint Protein EWS Defines a Specific Oncogenic Mechanism of ETS Factors Rearranged in Prostate Cancer. *Cell Reports*, *17*(5), 1289–1301. https://doi.org/10.1016/j.celrep.2016.10.001

Kelley, D., & Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology*, *13*(11), R107. https://doi.org/10.1186/gb-2012-13-11-r107

King, J. C., Xu, J., Wongvipat, J., Hieronymus, H., Carver, B. S., Leung, D. H., Taylor, B. S., Sander, C., Cardiff, R. D., Couto, S. S., Gerald, W. L., & Sawyers, C. L. (2009). Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nature Genetics*, *41*(5), 524–526. https://doi.org/10.1038/ng.371

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., & Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Science Signaling*, *3*(107), ra8. https://doi.org/10.1126/scisignal.2000568

Koivisto, P., Kononen, J., Palmberg, C., Tammela, T., Hyytinen, E., Isola, J., Trapman, J., Cleutjens, K., Noordzij, A., Visakorpi, T., & Kallioniemi, O. P. (1997). Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer Research*, *57*(2), 314–319.

Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, *172*(3), 393–407. https://doi.org/10.1016/j.cell.2018.01.011

Kron, K. J., Murison, A., Zhou, S., Huang, V., Yamaguchi, T. N., Shiah, Y. J., Fraser, M., van der Kwast, T., Boutros, P. C., Bristow, R. G., & Lupien, M. (2017). TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nature Genetics*, *49*(9), 1336–1345. https://doi.org/10.1038/ng.3930 [doi]

Krumbholz, M., Agaimy, A., Stoehr, R., Burger, M., Wach, S., Taubert, H., Wullich, B., Hartmann, A., & Metzler, M. (2019). Molecular Composition of Genomic TMPRSS2-ERG Rearrangements in Prostate Cancer. *Disease Markers*, *2019*, 5085373. https://doi.org/10.1155/2019/5085373

Kumar, S., Gonzalez, E. A., Rameshwar, P., & Etchegaray, J. P. (2020). Non-Coding RNAs as Mediators of Epigenetic Changes in Malignancies. *Cancers*, *12*(12), 1–32. https://doi.org/10.3390/CANCERS12123657

Kumar-Sinha, C., Tomlins, S. A., & Chinnaiyan, A. M. (2008). Recurrent Gene Fusions in Prostate Cancer. *Nature Reviews. Cancer*, *8*(7), 497–511. https://doi.org/10.1038/nrc2402

Kuriyama, M., Wang, M. C., Papsidero, L. D., Killian, C. S., Shimano, T., Valenzuela;, L., Nishiura, T., Murphy, G. P., & Chu, T. M. (1980). Quantitation of Prostate-specific Antigen in Serum by a Sensitive Enzyme Immunoassay. *Cancer Research*, *40*(12), 4658–4662.

Lagarde, J., Uszczynska-Ratajczak, B., Carbonell, S., Pérez-Lluch, S., Abad, A., Davis, C., Gingeras, T. R., Frankish, A., Harrow, J., Guigo, R., & Johnson, R. (2017). High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nature Genetics*, *49*(12), 1731–1740. https://doi.org/10.1038/ng.3988

Lang, S. H., Swift, S. L., White, H., Misso, K., Kleijnen, J., & Quek, R. G. W. (2019). A systematic review of the prevalence of DNA damage response gene mutations in prostate cancer. *International Journal of Oncology*, *55*(3), 597–616. https://doi.org/10.3892/ijo.2019.4842

Latonen, L., Afyounian, E., Jylhä, A., Nättinen, J., Aapola, U., Annala, M., Kivinummi, K. K., Tammela, T. T. L., Beuerman, R. W., Uusitalo, H., Nykter, M., & Visakorpi, T. (2018). Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-018-03573-6

Lau, A. N., & Heiden, M. G. V. (2020). Metabolism in the Tumor Microenvironment. *Annual Review of Cancer Biology*, *4*(1), 17–40. https://doi.org/10.1146/annurev-cancerbio-030419-033333

Laudet, V., Hänni, C., Stéhelin, D., & Duterque-Coquillaud, M. (1999). Molecular phylogeny of the ETS gene family. *Oncogene*, *18*(6), 1351–1359. https://doi.org/10.1038/sj.onc.1202444

Laxman, B., Tomlins, S. A., Mehra, R., Morris, D. S., Wang, L., Helgeson, B. E., Shah, R. B., Rubin, M. A., Wei, J. T., & Chinnaiyan, A. M. (2006). Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia (New York, N.Y.)*, *8*(10), 885–888. https://doi.org/10.1593/neo.06625

Lee, C. H., Akin-Olugbade, O., & Kirschenbaum, A. (2011). Overview of prostate anatomy, histology, and pathology. *Endocrinology and Metabolism Clinics of North America*, *40*(3), 565–575. https://doi.org/10.1016/J.ECL.2011.05.012

Lee, H., Zhang, Z., & Krause, H. M. (2019). Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? *Trends in Genetics*, *35*(12), 892–902. https://doi.org/10.1016/j.tig.2019.09.006

Leinonen, K. A., Saramäki, O. R., Furusato, B., Kimura, T., Takahashi, H., Egawa, S., Suzuki, H., Keiger, K., Hahm, S. H., Isaacs, W. B., Tolonen, T. T., Stenman, U.-H., Tammela, T. L. J., Nykter, M., Bova, G. S., & Visakorpi, T. (2013). Loss of PTEN is associated with aggressive behavior in ERG-positive prostate cancer. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *22*(12), 2333–2344. https://doi.org/10.1158/1055-9965.EPI-13-0333-T

Leinonen, K. A., Tolonen, T. T., Bracken, H., Stenman, U.-H., Tammela, T. L. J., Saramäki, O. R., & Visakorpi, T. (2010). Association of SPINK1 expression and

TMPRSS2:ERG fusion with prognosis in endocrine-treated prostate cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *16*(10), 2845–2851. https://doi.org/10.1158/1078-0432.CCR-09-2505

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Xu, C., Lee, H. J., Ren, S., Zi, X., Zhang, Z., Wang, H., Yu, Y., Yang, C., Gao, X., Hou, J., Wang, L., Yang, B., Yang, Q., Ye, H., Zhou, T., Lu, X., Wang, Y., Qu, M., … Sun, Y. (2020). A genomic and epigenomic atlas of prostate cancer in Asian populations. *Nature*, *580*(7801), Article 7801. https://doi.org/10.1038/s41586-020-2135-x

Li, J. Z. H., Gao, W., Ho, W.-K., Lei, W. B., Wei, W. I., Chan, J. Y.-W., & Wong, T.-S. (2016). The clinical association of programmed cell death protein 4 (PDCD4) with solid tumors and its prognostic significance: a meta-analysis. *Chinese Journal of Cancer*, *35*(1), 95. https://doi.org/10.1186/s40880-016-0158-3

Liao, Y., & Xu, K. (2019). Epigenetic regulation of prostate cancer: the theories and the clinical implications. *Asian Journal of Andrology*, *21*(3), 279. https://doi.org/10.4103/AJA.AJA_53_18

Lilja, H. (1985). A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. *The Journal of Clinical Investigation*, *76*(5), 1899–1903. https://doi.org/10.1172/JCI112185

Lilja, H., Christensson, A., Dahlén, U., Matikainen, M. T., Nilsson, O., Pettersson, K., & Lövgren, T. (1991). Prostate-specific antigen in serum occurs predominantly in complex with alpha 1-antichymotrypsin. *Clinical Chemistry*, *37*(9), 1618–1625. https://doi.org/10.1093/clinchem/37.9.1618

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B., Ohgi, K., Zhang, J., Rose, D. W., Fu, X.-D., Glass, C. K., & Rosenfeld, M. G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell*, *139*(6), 1069–1083. https://doi.org/10.1016/j.cell.2009.11.030

Linja, M. J., Savinainen, K. J., Saramäki, O. R., Tammela, T. L. J., Vessella, R. L., & Visakorpi, T. (2001). Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. *Cancer Research*, *61*(9), 3550–3555.

Little, G. H., Noushmehr, H., Baniwal, S. K., Berman, B. P., Coetzee, G. A., & Frenkel, B. (2012). Genome-wide Runx2 occupancy in prostate cancer cells suggests a role in regulating secretion. *Nucleic Acids Research*, *40*(8), 3538–3547. https://doi.org/10.1093/nar/gkr1219

Liu, K., Gao, L., Ma, X., Huang, J.-J., Chen, J., Zeng, L., Ashby, C. R., Zou, C., & Chen, Z.-S. (2020). Long non-coding RNAs regulate drug resistance in cancer. *Molecular Cancer*, *19*(1), 54. https://doi.org/10.1186/s12943-020-01162-0

Liu, Q.-P., Lin, J.-Y., An, P., Chen, Y.-Y., Luan, X., & Zhang, H. (2021). LncRNAs in tumor microenvironment: The potential target for cancer treatment with natural compounds and chemical drugs. *Biochemical Pharmacology*, *193*, 114802. https://doi.org/10.1016/j.bcp.2021.114802

Liu, S. J., Horlbeck, M. A., Cho, S. W., Birk, H. S., Malatesta, M., He, D., Attenello, F. J., Villalta, J. E., Cho, M. Y., Chen, Y., Mandegar, M. A., Olvera, M. P., Gilbert, L. A., Conklin, B. R., Chang, H. Y., Weissman, J. S., & Lim, D. A. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science (New York, N.Y.)*, *355*(6320). https://doi.org/10.1126/science.aah7111

Loblaw, A., Souter, L. H., Canil, C., Breau, R. H., Haider, M., Jamnicky, L., Morash, R., Surchin, M., & Matthew, A. (2017). Follow-up Care for Survivors of Prostate Cancer - Clinical Management: a Program in Evidence-Based Care Systematic Review and Clinical Practice Guideline. *Clinical Oncology (Royal College of Radiologists (Great Britain))*, *29*(11), 711–717. https://doi.org/10.1016/j.clon.2017.08.004

Loda, A., & Heard, E. (2019). Xist RNA in action: Past, present, and future. *PLoS Genetics*, *15*(9), e1008333. https://doi.org/10.1371/journal.pgen.1008333

Lotan, T. L., Carvalho, F. L., Peskoe, S. B., Hicks, J. L., Good, J., Fedor, H., Humphreys, E., Han, M., Platz, E. A., Squire, J. A., De Marzo, A. M., & Berman, D. M. (2015). PTEN loss is associated with upgrading of prostate cancer from biopsy to radical prostatectomy. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, *28*(1), 128–137. https://doi.org/10.1038/modpathol.2014.85

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S., Mesirov, J. P., Cech, T. R., & Chang, H. Y. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell*, *165*(5), 1267–1279. https://doi.org/10.1016/j.cell.2016.04.028

Lubelsky, Y., & Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*, *555*(7694), 107–111. https://doi.org/10.1038/nature25757

Lyu, Q. R., Zhang, S., Zhang, Z., & Tang, Z. (2023). Functional knockout of long non-coding RNAs with genome editing. *Frontiers in Genetics*, *14*. https://www.frontiersin.org/articles/10.3389/fgene.2023.1242129

Mani, R. S., Amin, M. A., Li, X., Kalyana-Sundaram, S., Veeneman, B. A., Wang, L., Ghosh, A., Aslam, A., Ramanand, S. G., Rabquer, B. J., Kimura, W., Tran, M., Cao, X., Roychowdhury, S., Dhanasekaran, S. M., Palanisamy, N., Sadek, H. A., Kapur, P., Koch, A. E., & Chinnaiyan, A. M. (2016). Inflammation-Induced Oxidative Stress Mediates Gene Fusion Formation in Prostate Cancer. *Cell Reports*, *17*(10), 2620–2631. https://doi.org/10.1016/j.celrep.2016.11.019

Mani, R. S., & Chinnaiyan, A. M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nature Reviews. Genetics*, *11*(12), 819–829. https://doi.org/10.1038/nrg2883

Mani, R. S., Tomlins, S. A., Callahan, K., Ghosh, A., Nyati, M. K., Varambally, S., Palanisamy, N., & Chinnaiyan, A. M. (2009). Induced chromosomal proximity and gene fusions in prostate cancer. *Science (New York, N.Y.)*, *326*(5957), 1230. https://doi.org/10.1126/science.1178124

Manning, B. D., & Toker, A. (2017). AKT/PKB Signaling: Navigating the Network. *Cell*, *169*(3), 381–405. https://doi.org/10.1016/j.cell.2017.04.001

Markowski, M. C., Bowen, C., & Gelmann, E. P. (2008). Inflammatory cytokines induce phosphorylation and ubiquitination of prostate suppressor protein NKX3.1. *Cancer Research*, *68*(17), 6896–6901. https://doi.org/10.1158/0008-5472.CAN-08-0578

Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., & Campbell, P. J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, *171*(5), 1029-1041.e21. https://doi.org/10.1016/J.CELL.2017.09.042

Marzo, A. M. D., Platz, E. A., Sutcliffe, S., Xu, J., Grönberg, H., Drake, C. G., Nakai, Y., Isaacs, W. B., & Nelson, W. G. (2007). Inflammation in prostate carcinogenesis. *Nature Reviews. Cancer*, *7*(4), 256–269. https://doi.org/10.1038/nrc2090

Massie, C. E., Adryan, B., Barbosa-Morais, N. L., Lynch, A. G., Tran, M. G., Neal, D. E., & Mills, I. G. (2007). New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Reports*, *8*(9), 871–878. https://doi.org/7401046 [pii]

Massie, C. E., Lynch, A., Ramos-Montoya, A., Boren, J., Stark, R., Fazli, L., Warren, A., Scott, H., Madhu, B., Sharma, N., Bon, H., Zecchini, V., Smith, D.-M., Denicola, G. M., Mathews, N., Osborne, M., Hadfield, J., Macarthur, S., Adryan, B., … Mills, I. G. (2011). The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *The EMBO Journal*, *30*(13), 2719–2733. https://doi.org/10.1038/emboj.2011.158

Matsuhashi, S., Manirujjaman, M., Hamajima, H., & Ozaki, I. (2019). Control Mechanisms of the Tumor Suppressor PDCD4: Expression and Functions. *International Journal of Molecular Sciences*, *20*(9), Article 9. https://doi.org/10.3390/ijms20092304

Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L.-L., Chen, R., Dean, C., Dinger, M. E., Fitzgerald, K. A., Gingeras, T. R., Guttman, M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence, J. B., Lee, J. T., … Wu, M. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*, *24*(6), Article 6. https://doi.org/10.1038/s41580-022-00566-8

McGlincy, N. J., & Ingolia, N. T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods (San Diego, Calif.)*, *126*, 112–129. https://doi.org/10.1016/j.ymeth.2017.05.028

McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., & Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–236. https://doi.org/10.1038/nature14443

McMenamin, M. E., Soung, P., Perera, S., Kaplan, I., Loda, M., & Sellers, W. R. (1999). Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage. *Cancer Research*, *59*(17), 4291–4296.

McNeal, J. E. (1981). The zonal anatomy of the prostate. *The Prostate*, *2*(1), 35–49. https://doi.org/10.1002/PROS.2990020105

McNeal, J. E., Redwine, E. A., Freiha, F. S., & Stamey, T. A. (1988). Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. *The American Journal of Surgical Pathology*, *12*(12), 897–906. https://doi.org/10.1097/00000478-198812000-00001

Mehra, R., Shi, Y., Udager, A. M., Prensner, J. R., Sahu, A., Iyer, M. K., Siddiqui, J., Cao, X., Wei, J., Jiang, H., Feng, F. Y., & Chinnaiyan, A. M. (2014). A Novel RNA In Situ Hybridization Assay for the Long Noncoding RNA SChLAP1 Predicts Poor Clinical Outcome After Radical Prostatectomy in Clinically Localized Prostate Cancer. *Neoplasia*, *16*(12), 1121–1127. https://doi.org/10.1016/J.NEO.2014.11.006

Mehra, R., Udager, A. M., Ahearn, T. U., Cao, X., Feng, F. Y., Loda, M., Petimar, J. S., Kantoff, P., Mucci, L. A., & Chinnaiyan, A. M. (2016). Overexpression of the Long Non-coding RNA SChLAP1 Independently Predicts Lethal Prostate Cancer. *European Urology*, *70*(4), 549–552. https://doi.org/10.1016/J.EURURO.2015.12.003

Melé, M., Mattioli, K., Mallard, W., Shechner, D. M., Gerhardinger, C., & Rinn, J. L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Research*, *27*(1), 27–37. https://doi.org/10.1101/gr.214205.116

Mercer, T. R., Munro, T., & Mattick, J. S. (2022). The potential of long noncoding RNA therapies. *Trends in Pharmacological Sciences*, *43*(4), 269–280. https://doi.org/10.1016/j.tips.2022.01.008

Mesquita, D., Barros-Silva, J. D., Santos, J., Skotheim, R. I., Lothe, R. A., Paulo, P., & Teixeira, M. R. (2015). Specific and redundant activities of ETV1 and ETV4 in prostate cancer aggressiveness revealed by co-overexpression cellular contexts. *Oncotarget*, *6*(7), 5217–5236. https://doi.org/10.18632/oncotarget.2847

Metzger, E., Wissmann, M., Yin, N., Müller, J. M., Schneider, R., Peters, A. H. F. M., Günther, T., Buettner, R., & Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature 2005 437:7057*, *437*(7057), 436–439. https://doi.org/10.1038/nature04020

Mills, I. G. (2014). Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. *Nature Reviews. Cancer*, *14*(3), 187–198. https://doi.org/10.1038/nrc3678

Minotti, L., Agnoletto, C., Baldassari, F., Corrà, F., & Volinia, S. (2018). SNPs and Somatic Mutation on Long Non-Coding RNA: New Frontier in the Cancer Studies? *High-Throughput*, *7*(4), 34. https://doi.org/10.3390/ht7040034

Mirzaei, S., Paskeh, M. D. A., Okina, E., Gholami, M. H., Hushmandi, K., Hashemi, M., Kalu, A., Zarrabi, A., Nabavi, N., Rabiee, N., Sharifi, E., Karimi-Maleh, H., Ashrafizadeh, M., Kumar, A. P., & Wang, Y. (2022). Molecular Landscape of LncRNAs in Prostate Cancer: A focus on pathways and therapeutic targets for intervention. *Journal of Experimental & Clinical Cancer Research*, *41*(1), 214. https://doi.org/10.1186/s13046-022-02406-1

Mitobe, Y., Takayama, K., Horie-Inoue, K., & Inoue, S. (2018). Prostate cancer-associated lncRNAs. *Cancer Letters*, *418*, 159–166. https://doi.org/10.1016/j.canlet.2018.01.012

Mizushima, N., & Yoshimori, T. (2007). How to Interpret LC3 Immunoblotting. *Autophagy*, *3*(6), 542–545. https://doi.org/10.4161/auto.4600

Morris, W. J., Tyldesley, S., Rodda, S., Halperin, R., Pai, H., McKenzie, M., Duncan, G., Morton, G., Hamm, J., & Murray, N. (2017). Androgen Suppression Combined with Elective Nodal and Dose Escalated Radiation Therapy (the ASCENDE-RT Trial): An Analysis of Survival Endpoints for a Randomized Trial Comparing a Low-Dose-Rate Brachytherapy Boost to a Dose-Escalated External Beam Boost for High- and Intermediate-risk Prostate Cancer. *International Journal of Radiation Oncology, Biology, Physics*, *98*(2), 275–285. https://doi.org/10.1016/j.ijrobp.2016.11.026

Mounir, Z., Korn, J. M., Westerling, T., Lin, F., Kirby, C. A., Schirle, M., McAllister, G., Hoffman, G., Ramadan, N., Hartung, A., Feng, Y., Kipp, D. R., Quinn, C., Fodor, M., Baird, J., Schoumacher, M., Meyer, R., Deeds, J., Buchwalter, G., … Pagliarini, R. A. (2016). ERG signaling in prostate cancer is driven through PRMT5-dependent methylation of the Androgen Receptor. *eLife*, *5*, 10.7554/eLife.13964. https://doi.org/10.7554/eLife.13964 [doi]

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., Czene, K., Havelick, D. J., Scheike, T., Graff, R. E., Holst, K., Möller, S., Unger, R. H., McIntosh, C., Nuttall, E., Brandt, I., Penney, K. L., Hartman, M., Kraft, P., Parmigiani, G., Christensen, K., Koskenvuo, M., …

Kaprio, J. (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, *315*(1), 68–76. https://doi.org/10.1001/jama.2015.17703

Mulholland, D. J., Tran, L. M., Li, Y., Cai, H., Morim, A., Wang, S., Plaisier, S., Garraway, I. P., Huang, J., Graeber, T. G., & Wu, H. (2011). Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. *Cancer Cell*, *19*(6), 792–804. https://doi.org/10.1016/j.ccr.2011.05.006

Nadler, R. B., Humphrey, P. A., Smith, D. S., Catalona, W. J., & Ratliff, T. L. (1995). Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels. *The Journal of Urology*, *154*(2 Pt 1), 407–413. https://doi.org/10.1097/00005392-199508000-00023

Natesan, R., Aras, S., Effron, S. S., & Asangani, I. A. (2019). Epigenetic Regulation of Chromatin in Prostate Cancer. *Advances in Experimental Medicine and Biology*, *1210*, 379–407. https://doi.org/10.1007/978-3-030-32656-2_17

Neal, D. E., Metcalfe, C., Donovan, J. L., Lane, J. A., Davis, M., Young, G. J., Dutton, S. J., Walsh, E. I., Martin, R. M., Peters, T. J., Turner, E. L., Mason, M., Bryant, R., Bollina, P., Catto, J., Doherty, A., Gillatt, D., Gnanapragasam, V., Holding, P., … Hamdy, F. C. (2020). Ten-year Mortality, Disease Progression, and Treatment-related Side Effects in Men with Localised Prostate Cancer from the ProtecT Randomised Controlled Trial According to Treatment Received. *European Urology*, *77*(3), 320–330. https://doi.org/10.1016/j.eururo.2019.10.030

Nepple, K. G., Stephenson, A. J., Kallogjeri, D., Michalski, J., Grubb, R. L., Strope, S. A., Haslag-Minoff, J., Piccirillo, J. F., Ciezki, J. P., Klein, E. A., Reddy, C. A., Yu, C., Kattan, M. W., & Kibel, A. S. (2013). Mortality after prostate cancer treatment with radical prostatectomy, external-beam radiation therapy, or brachytherapy in men without comorbidity. *European Urology*, *64*(3), 372–378. https://doi.org/10.1016/j.eururo.2013.03.005

Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., & Graves, B. J. (1992). Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes & Development*, *6*(6), 975–990. https://doi.org/10.1101/gad.6.6.975

Oesterling, J. E., Jacobsen, S. J., Klee, G. G., Pettersson, K., Piironen, T., Abrahamsson, P. A., Stenman, U. H., Dowell, B., Lövgren, T., & Lilja, H. (1995). Free, complexed and total serum prostate specific antigen: the establishment of appropriate reference ranges for their concentrations and ratios. *The Journal of Urology*, *154*(3), 1090–1095. https://doi.org/10.1016/s0022-5347(01)66984-2

Pearson, H. B., Li, J., Meniel, V. S., Fennell, C. M., Waring, P., Montgomery, K. G., Rebello, R. J., Macpherson, A. A., Koushyar, S., Furic, L., Cullinane, C., Clarkson, R. W., Smalley, M. J., Simpson, K. J., Phesse, T. J., Shepherd, P. R., Humbert, P. O., Sansom, O. J., & Phillips, W. A. (2018). Identification of Pik3ca Mutation as a Genetic Driver of Prostate Cancer That Cooperates with Pten Loss to Accelerate Progression and Castration-Resistant Growth. *Cancer Discovery*, *8*(6), 764–779. https://doi.org/10.1158/2159-8290.CD-17-0867

Perner, S., Mosquera, J. M., Demichelis, F., Hofer, M. D., Paris, P. L., Simko, J., Collins, C., Bismar, T. A., Chinnaiyan, A. M., De Marzo, A. M., & Rubin, M. A. (2007). TMPRSS2-ERG fusion prostate cancer: An early molecular event associated with invasion. *American Journal of Surgical Pathology*, *31*(6), 882–888. https://doi.org/10.1097/01.PAS.0000213424.38503.AA

Pettersson, A., Graff, R. E., Bauer, S. R., Pitt, M. J., Lis, R. T., Stack, E. C., Martin, N. E., Kunz, L., Penney, K. L., Ligon, A. H., Suppan, C., Flavin, R., Sesso, H. D., Rider, J. R., Sweeney, C., Stampfer, M. J., Fiorentino, M., Kantoff, P. W., Sanda, M. G., … Mucci, L. A. (2012). The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *21*(9), 1497–1509. https://doi.org/10.1158/1055-9965.EPI-12-0042

Pierorazio, P. M., Walsh, P. C., Partin, A. W., & Epstein, J. I. (2013). Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU International*, *111*(5), 753–760. https://doi.org/10.1111/j.1464-410X.2012.11611.x

Pinsky, P. F., Parnes, H. L., & Andriole, G. (2014). Mortality and complications after prostate biopsy in the Prostate, Lung, Colorectal and Ovarian Cancer Screening (PLCO) trial. *BJU International*, *113*(2), 254–259. https://doi.org/10.1111/bju.12368

Pomerantz, M. M., Li, F., Takeda, D. Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R. A., Bowden, M., Lis, R., Hahn, W. C., Kantoff, P. W., Brown, M., Loda, M., Long, H. W., & Freedman, M. L. (2015). The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nature Genetics*, *47*(11), 1346–1351. https://doi.org/10.1038/ng.3419 [doi]

Ponjavic, J., Ponting, C. P., & Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Research*, *17*(5), 556–565. https://doi.org/10.1101/gr.6036807

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell*, *136*(4), 629–641. https://doi.org/10.1016/j.cell.2009.02.006

Pound, C. R., Partin, A. W., Eisenberger, M. A., Chan, D. W., Pearson, J. D., & Walsh, P. C. (1999). Natural History of Progression After PSA Elevation Following Radical Prostatectomy. *JAMA*, *281*(17), 1591–1597. https://doi.org/10.1001/jama.281.17.1591

Prensner, J. R., Iyer, M. K., Balbin, O. A., Dhanasekaran, S. M., Cao, Q., Brenner, J. C., Laxman, B., Asangani, I. A., Grasso, C. S., Kominsky, H. D., Cao, X., Jing, X., Wang, X., Siddiqui, J., Wei, J. T., Robinson, D., Iyer, H. K., Palanisamy, N., Maher, C. A., & Chinnaiyan, A. M. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature Biotechnology*, *29*(8), 742–749. https://doi.org/10.1038/nbt.1914

Prensner, J. R., Iyer, M. K., Sahu, A., Asangani, I. A., Cao, Q., Patel, L., Vergara, I. A., Davicioni, E., Erho, N., Ghadessi, M., Jenkins, R. B., Triche, T. J., Malik, R., Bedenis, R., McGregor, N., Ma, T., Chen, W., Han, S., Jing, X., … Chinnaiyan, A. M. (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature Genetics*, *45*(11), 1392–1398. https://doi.org/10.1038/ng.2771

Prensner, J. R., Zhao, S., Erho, N., Schipper, M., Iyer, M. K., Dhanasekaran, S. M., Magi-Galluzzi, C., Mehra, R., Sahu, A., Siddiqui, J., Davicioni, E., Den, R. B., Dicker, A. P., Karnes, R. J., Wei, J. T., Klein, E. A., Jenkins, R. B., Chinnaiyan, A. M., & Feng, F. Y. (2014). Nomination and validation of the long noncoding RNA SChLAP1 as a risk factor for metastatic prostate cancer progression: a multi-institutional high-throughput analysis. *The Lancet. Oncology*, *15*(13), 1469. https://doi.org/10.1016/S1470-2045(14)71113-1

Pritchard, C. C., Mateo, J., Walsh, M. F., Sarkar, N. D., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., Elemento, O., Rubin, M. A., Robinson, D., Lonigro, R., Hussain, M., Chinnaiyan, A., Vinson, J., Filipenko, J., Garraway, L., … Nelson, P. S. (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *New England Journal of Medicine*, *375*(5), 443–453. https://doi.org/10.1056/NEJMoa1603144

Pungsrinont, T., Kallenbach, J., & Baniahmad, A. (2021). Role of PI3K-AKT-mTOR Pathway as a Pro-Survival Signaling and Resistance-Mediating Mechanism to Therapy of Prostate Cancer. *International Journal of Molecular Sciences*, *22*(20), Article 20. https://doi.org/10.3390/ijms222011088

Qian, Y., Shi, L., & Luo, Z. (2020). Long Non-coding RNAs in Cancer: Implications for Diagnosis, Prognosis, and Therapy. *Frontiers in Medicine*, *7*, 612393. https://doi.org/10.3389/fmed.2020.612393

Quigley, D. A., Dang, H. X., Zhao, S. G., Lloyd, P., Aggarwal, R., Alumkal, J. J., Foye, A., Kothari, V., Perry, M. D., Bailey, A. M., Playdle, D., Barnard, T. J., Zhang, L., Zhang, J., Youngren, J. F., Cieslik, M. P., Parolia, A., Beer, T. M., Thomas, G., … Feng, F. Y. (2018). Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell*, *174*(3), 758-769.e9. https://doi.org/10.1016/j.cell.2018.06.039

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews. Genetics*, *17*(1), 47–62. https://doi.org/10.1038/nrg.2015.10

Ramanand, S. G., & Mani, R. S. (2019). Genetic, Environmental, and Nuclear Factors Governing Genomic Rearrangements. *Advances in Experimental Medicine and Biology*, *1210*, 57–66. https://doi.org/10.1007/978-3-030-32656-2_3

Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nature Reviews.Molecular Cell Biology*, *19*(3), 143–157. https://doi.org/10.1038/nrm.2017.104 [doi]

Rawla, P. (2019). Epidemiology of Prostate Cancer. *World Journal of Oncology*, *10*(2), 63–89. https://doi.org/10.14740/wjon1191

Rebuzzi, S. E., Rescigno, P., Catalano, F., Mollica, V., Vogl, U. M., Marandino, L., Massari, F., Pereira Mestre, R., Zanardi, E., Signori, A., Buti, S., Bauckneht, M., Gillessen, S., Banna, G. L., & Fornarini, G. (2022). Immune Checkpoint Inhibitors in Advanced Prostate Cancer: Current Data and Future Perspectives. *Cancers*, *14*(5), 1245. https://doi.org/10.3390/cancers14051245

Reid, A. H. M., Attard, G., Ambroisine, L., Fisher, G., Kovacs, G., Brewer, D., Clark, J., Flohr, P., Edwards, S., Berney, D. M., Foster, C. S., Fletcher, A., Gerald, W. L., Møller, H., Reuter, V. E., Scardino, P. T., Cuzick, J., de Bono, J. S., & Cooper, C. S. (2010). Molecular characterisation of ERG, ETV1 and PTEN gene loci identifies patients at low and high risk of death from prostate cancer. *British Journal of Cancer*, *102*(4), 678–684. https://doi.org/10.1038/sj.bjc.6605554

Ren, S., Peng, Z., Mao, J.-H., Yu, Y., Yin, C., Gao, X., Cui, Z., Zhang, J., Yi, K., Xu, W., Chen, C., Wang, F., Guo, X., Lu, J., Yang, J., Wei, M., Tian, Z., Guan, Y., Tang, L., … Sun, Y. (2012). RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and

aberrant alternative splicings. *Cell Research*, *22*(5), 806–821. https://doi.org/10.1038/cr.2012.30

Riggs, A. D., Martienssen, R. A., & Russo, V. E. A. (1996). Introduction. In *Epigenetic mechanisms of gene regulation* (pp. 1–4). Cold Spring Harbor Laboratory Press.

Risbridger, G. P., Taylor, R. A., Clouston, D., Sliwinski, A., Thorne, H., Hunter, S., Li, J., Mitchell, G., Murphy, D., Frydenberg, M., Pook, D., Pedersen, J., Toivanen, R., Wang, H., Papargiris, M., Lawrence, M. G., & Bolton, D. M. (2015). Patient-derived xenografts reveal that intraductal carcinoma of the prostate is a prominent pathology in BRCA2 mutation carriers with prostate cancer and correlates with poor prognosis. *European Urology*, *67*(3), 496–503. https://doi.org/10.1016/j.eururo.2014.08.007

Roach, M., Weinberg, V., Nash, M., Sandler, H. M., McLaughlin, P. W., & Kattan, M. W. (2006). Defining high risk prostate cancer with risk groups and nomograms: implications for designing clinical trials. *The Journal of Urology*, *176*(6 Pt 2), 16. https://doi.org/10.1016/j.juro.2006.06.081

Robinson, D., Allen, E. M. V., Wu, Y.-M., Schultz, N., Lonigro, R. J., Mosquera, J.-M., Montgomery, B., Taplin, M.-E., Pritchard, C. C., Attard, G., Beltran, H., Abida, W., Bradley, R. K., Vinson, J., Cao, X., Vats, P., Kunju, L. P., Hussain, M., Feng, F. Y., … Chinnaiyan, A. M. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell*, *161*(5), 1215–1228. https://doi.org/10.1016/j.cell.2015.05.001

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), Article 1. https://doi.org/10.1038/nbt.1754

Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F., & Crespi, M. (2018). Splicing regulation by long noncoding RNAs. *Nucleic Acids Research*, *46*(5), 2169–2184. https://doi.org/10.1093/nar/gky095

Roux, P. P., & Topisirovic, I. (2018). Signaling Pathways Involved in the Regulation of mRNA Translation. *Molecular and Cellular Biology*, *38*(12), e00070-18. https://doi.org/10.1128/MCB.00070-18

Sahu, B., Laakso, M., Ovaska, K., Mirtti, T., Lundin, J., Rannikko, A., Sankila, A., Turunen, J. P., Lundin, M., Konsti, J., Vesterinen, T., Nordling, S., Kallioniemi, O., Hautaniemi, S., & Jänne, O. A. (2011). Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *The EMBO Journal*, *30*(19), 3962–3976. https://doi.org/10.1038/emboj.2011.328 [doi]

Sahu, B., Pihlajamaa, P., Dubois, V., Kerkhofs, S., Claessens, F., & Jänne, O. A. (2014). Androgen receptor uses relaxed response element stringency for selective chromatin binding and transcriptional regulation in vivo. *Nucleic Acids Research*, *42*(7), 4230–4240. https://doi.org/10.1093/nar/gkt1401

Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell*, *146*(3), 353–358. https://doi.org/10.1016/j.cell.2011.07.014

Sartor, O., de Bono, J., Chi, K. N., Fizazi, K., Herrmann, K., Rahbar, K., Tagawa, S. T., Nordquist, L. T., Vaishampayan, N., El-Haddad, G., Park, C. H., Beer, T. M., Armour, A., Pérez-Contreras, W. J., DeSilvio, M., Kpamegan, E., Gericke, G., Messmann, R. A., Morris, M. J., & Krause, B. J. (2021). Lutetium-177–PSMA-617 for Metastatic Castration-Resistant Prostate Cancer. *New England Journal of Medicine*, *385*(12), 1091–1103. https://doi.org/10.1056/NEJMoa2107322

Sartori, D. A., & Chan, D. W. (2014). Biomarkers in prostate cancer: What's new? *Current Opinion in Oncology*, *26*(3), 259–264. https://doi.org/10.1097/CCO.0000000000000065

Saxton, R. A., & Sabatini, D. M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. *Cell*, *168*(6), 960–976. https://doi.org/10.1016/j.cell.2017.02.004

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, *65*(1), 25–38. https://doi.org/10.1016/j.molcel.2016.11.029

Schmitt, A. M., & Chang, H. Y. (2016). Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, *29*(4), 452–463. https://doi.org/10.1016/j.ccell.2016.03.010

Seppä, K., Tanskanen, T., Heikkinen, S., Malila, N., & Pitkäniemi, J. (2023). Cancer in Finland 2021. *Cancer Society of Finland*, *Helsinki*. https://cancerregistry.fi/

Sfanos, K. S., Yegnasubramanian, S., Nelson, W. G., & Marzo, A. M. D. (2018). The inflammatory microenvironment and microbiome in prostate cancer development. *Nature Reviews. Urology*, *15*(1), 11–24. https://doi.org/10.1038/nrurol.2017.167

Shao, L., Tekedereli, I., Wang, J., Yuca, E., Tsang, S., Sood, A., Lopez-Berestein, G., Ozpolat, B., & Ittmann, M. (2012). Highly specific targeting of the TMPRSS2/ERG fusion gene using liposomal nanovectors. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *18*(24), 6648–6657. https://doi.org/10.1158/1078-0432.CCR-12-2715

Sharma, N. L., Massie, C. E., Ramos-Montoya, A., Zecchini, V., Scott, H. E., Lamb, A. D., MacArthur, S., Stark, R., Warren, A. Y., Mills, I. G., & Neal, D. E. (2013). The Androgen Receptor Induces a Distinct Transcriptional Program in Castration-Resistant Prostate Cancer in Man. *Cancer Cell*, *23*(1), 35–47. https://doi.org/10.1016/j.ccr.2012.11.010

Shi, Y., Parag, S., Patel, R., Lui, A., Murr, M., Cai, J., & Patel, N. A. (2019). Stabilization of lncRNA GAS5 by a Small Molecule and Its Implications in Diabetic Adipocytes. *Cell Chemical Biology*, *26*(3), 319-330.e6. https://doi.org/10.1016/j.chembiol.2018.11.012

Shorning, B. Y., Dass, M. S., Smalley, M. J., & Pearson, H. B. (2020). The PI3K-AKT-mTOR Pathway and Prostate Cancer: At the Crossroads of AR, MAPK, and WNT Signaling. *International Journal of Molecular Sciences*, *21*(12), 4507. https://doi.org/10.3390/ijms21124507

Shukla, S., Zhang, X., Niknafs, Y. S., Xiao, L., Mehra, R., Cieślik, M., Ross, A., Schaeffer, E., Malik, B., Guo, S., Freier, S. M., Bui, H.-H., Siddiqui, J., Jing, X., Cao, X., Dhanasekaran, S. M., Feng, F. Y., Chinnaiyan, A. M., & Malik, R. (2016). Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer. *Neoplasia (New York, N.Y.)*, *18*(8), 489–499. https://doi.org/10.1016/j.neo.2016.07.001

Smith, A. F. A., Hubley, R., & Green, P. (2013). *RepeatMasker Open-4.0*. http://www.repeatmasker.org

Smith, D. F., & Toft, D. O. (2008). Minireview: The Intersection of Steroid Receptors with Molecular Chaperones: Observations and Questions. *Molecular Endocrinology*, *22*(10), 2229–2240. https://doi.org/10.1210/me.2008-0089

Smith, S. F., Collins, S. E., & Charest, P. G. (2020). Ras, PI3K and mTORC2 – three's a crowd? *Journal of Cell Science*, *133*(19), jcs234930. https://doi.org/10.1242/jcs.234930

Smola, M. J., Calabrese, J. M., & Weeks, K. M. (2015). Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry*, *54*(46), 6867–6875. https://doi.org/10.1021/acs.biochem.5b00977

Soghli, N., Yousefi, T., Abolghasemi, M., & Qujeq, D. (2021). NORAD, a critical long non-coding RNA in human cancers. *Life Sciences*, *264*, 118665. https://doi.org/10.1016/j.lfs.2020.118665

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., & Pyle, A. M. (2015). HOTAIR forms an intricate and modular secondary structure. *Molecular Cell*, *58*(2), 353–361. https://doi.org/10.1016/j.molcel.2015.03.006

Song, C., & Chen, H. (2018). Predictive significance of TMRPSS2-ERG fusion in prostate cancer: a meta-analysis. *Cancer Cell International*, *18*, 177. https://doi.org/10.1186/s12935-018-0672-2

Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., & Chang, H. Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, *519*(7544), 486–490. https://doi.org/10.1038/nature14263

Srikantan, V., Zou, Z., Petrovics, G., Xu, L., Augustus, M., Davis, L., Livezey, J. R., Connell, T., Sesterhenn, I. A., Yoshino, K., Buzard, G. S., Mostofi, F. K., McLeod, D. G., Moul, J. W., & Srivastava, S. (2000). PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 12216. https://doi.org/10.1073/PNAS.97.22.12216

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, *22*(2), Article 2. https://doi.org/10.1038/s41580-020-00315-9

Stelloo, S., Nevedomskaya, E., van der Poel, H. G., de Jong, J., Leenders, G., Jenster, G., Wessels, L., Bergman, A., & Zwart, W. (2015). Androgen receptor profiling predicts prostate cancer outcome. *EMBO Molecular Medicine*, *7*(11), 1450–1464. https://doi.org/10.15252/emmm.201505424

Stenman, U. H., Leinonen, J., Alfthan, H., Rannikko, S., Tuhkanen, K., & Alfthan, O. (1991). A complex between prostate-specific antigen and alpha 1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer. *Cancer Research*, *51*(1), 222–226.

Strand, D. W., & Goldstein, A. S. (2015). The many ways to make a luminal cell and a prostate cancer cell. *Endocrine-Related Cancer*, *22*(6), 187. https://doi.org/10.1530/ERC-15-0195

Takayama, K., Tsutsumi, S., Katayama, S., Okayama, T., Horie-Inoue, K., Ikeda, K., Urano, T., Kawazu, C., Hasegawa, A., Ikeo, K., Gojyobori, T., Ouchi, Y., Hayashizaki, Y., Aburatani, H., & Inoue, S. (2011). Integration of cap analysis of gene expression and chromatin immunoprecipitation analysis on array reveals genome-wide androgen receptor signaling in prostate cancer cells. *Oncogene*, *30*(5), Article 5. https://doi.org/10.1038/onc.2010.436

Takayama, K.-I., Horie-Inoue, K., Katayama, S., Suzuki, T., Tsutsumi, S., Ikeda, K., Urano, T., Fujimura, T., Takagi, K., Takahashi, S., Homma, Y., Ouchi, Y., Aburatani, H., Hayashizaki, Y., & Inoue, S. (2013). Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. *The EMBO Journal*, *32*(12), 1665–1680. https://doi.org/10.1038/emboj.2013.99

Takeda, D. Y., Spisák, S., Seo, J.-H., Bell, C., O'Connor, E., Korthauer, K., Ribli, D., Csabai, I., Solymosi, N., Szállási, Z., Stillman, D. R., Cejas, P., Qiu, X., Long, H. W., Tisza, V., Nuzzo, P. V., Rohanizadegan, M., Pomerantz, M. M., Hahn, W. C., & Freedman, M. L. (2018). A Somatically Acquired Enhancer of the Androgen Receptor Is a

Noncoding Driver in Advanced Prostate Cancer. *Cell*, *174*(2), 422-432.e13. https://doi.org/10.1016/j.cell.2018.05.037

Tan, M. E., Li, J., Xu, H. E., Melcher, K., & Yong, E. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*, *36*(1), Article 1. https://doi.org/10.1038/aps.2014.18

Tandefelt, D. G., Boormans, J., Hermans, K., & Trapman, J. (2014). ETS fusion genes in prostate cancer. *Endocrine-Related Cancer*, *21*(3), R143–R152. https://doi.org/10.1530/ERC-13-0390

Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., Arora, V. K., Kaushik, P., Cerami, E., Reva, B., Antipin, Y., Mitsiades, N., Landers, T., Dolgalev, I., Major, J. E., Wilson, M., Socci, N. D., Lash, A. E., Heguy, A., … Gerald, W. L. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell*, *18*(1), 11–22. https://doi.org/10.1016/j.ccr.2010.05.026

Taylor, R. A., Cowin, P. A., Cunha, G. R., Pera, M., Trounson, A. O., Pedersen, J., & Risbridger, G. P. (2006). Formation of human prostate tissue from embryonic stem cells. *Nature Methods*, *3*(3), 179–181. https://doi.org/10.1038/nmeth855

Tellez-Gabriel, M., & Heymann, D. (2019). Exosomal lncRNAs: the newest promising liquid biopsy. *Cancer Drug Resistance*, *2*(4), 1002–1017. https://doi.org/10.20517/cdr.2019.69

Tettey, T. T., Rinaldi, L., & Hager, G. L. (2023). Long-range gene regulation in hormone-dependent cancer. *Nature Reviews Cancer*, *23*(10), Article 10. https://doi.org/10.1038/s41568-023-00603-4

The Breast Cancer Linkage Consortium. (1999). Cancer Risks in BRCA2 Mutation Carriers. *JNCI: Journal of the National Cancer Institute*, *91*(15), 1310–1316. https://doi.org/10.1093/jnci/91.15.1310

The Cancer Genome Atlas Research Network. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, *163*(4), 1011–1025. https://doi.org/10.1016/J.CELL.2015.10.025

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature 2012 489:7414*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Tomlins, S. A., Bjartell, A., Chinnaiyan, A. M., Jenster, G., Nam, R. K., Rubin, M. A., & Schalken, J. A. (2009). ETS gene fusions in prostate cancer: from discovery to daily clinical practice. *European Urology*, *56*(2), 275–286. https://doi.org/10.1016/j.eururo.2009.04.036

Tomlins, S. A., Day, J. R., Lonigro, R. J., Hovelson, D. H., Siddiqui, J., Kunju, L. P., Dunn, R. L., Meyer, S., Hodge, P., Groskopf, J., Wei, J. T., & Chinnaiyan, A. M. (2016). Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. *European Urology*, *70*(1), 45–53. https://doi.org/10.1016/j.eururo.2015.04.039

Tomlins, S. A., Laxman, B., Dhanasekaran, S. M., Helgeson, B. E., Cao, X., Morris, D. S., Menon, A., Jing, X., Cao, Q., Han, B., Yu, J., Wang, L., Montie, J. E., Rubin, M. A., Pienta, K. J., Roulston, D., Shah, R. B., Varambally, S., Mehra, R., & Chinnaiyan, A. M. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*, *448*(7153), 595–599. https://doi.org/10.1038/nature06024

Tomlins, S. A., Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B. E., Cao, Q., Prensner, J. R., Rubin, M. A., Shah, R. B., Mehra, R., & Chinnaiyan, A. M. (2008).

Role of the TMPRSS2-ERG Gene Fusion in Prostate Cancer. *Neoplasia*, *10*(2), 177-IN9. https://doi.org/10.1593/neo.07822

Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J. E., Shah, R. B., Pienta, K. J., Rubin, M. A., & Chinnaiyan, A. M. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, N.Y.)*, *310*(5748), 644–648. https://doi.org/10.1126/science.1117679

Tortorella, E., Giantulli, S., Sciarra, A., & Silvestri, I. (2023). AR and PI3K/AKT in Prostate Cancer: A Tale of Two Interconnected Pathways. *International Journal of Molecular Sciences*, *24*(3), Article 3. https://doi.org/10.3390/ijms24032046

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A., Blencowe, B. J., Prasanth, S. G., & Prasanth, K. V. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Molecular Cell*, *39*(6), 925–938. https://doi.org/10.1016/j.molcel.2010.08.011

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., … Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03828-1

Urbanucci, A., Barfeld, S. J., Kytölä, V., Itkonen, H. M., Coleman, I. M., Vodák, D., Sjöblom, L., Sheng, X., Tolonen, T., Minner, S., Burdelski, C., Kivinummi, K. K., Kohvakka, A., Kregel, S., Takhar, M., Alshalalfa, M., Davicioni, E., Erho, N., Lloyd, P., … Mills, I. G. (2017). Androgen Receptor Deregulation Drives Bromodomain-Mediated Chromatin Alterations in Prostate Cancer. *Cell Reports*, *19*(10), 2045–2059. https://doi.org/10.1016/j.celrep.2017.05.049

Uusi-Mäkelä, J., Afyounian, E., Tabaro, F., Häkkinen, T., Lussana, A., Shcherban, A., Annala, M., Nurminen, R., Kivinummi, K., Tammela, T. L. J., Urbanucci, A., Latonen, L., Kesseli, J., Granberg, K. J., Visakorpi, T., & Nykter, M. (2020). *Chromatin accessibility analysis uncovers regulatory element landscape in prostate cancer progression* (p. 2020.09.08.287268). bioRxiv. https://doi.org/10.1101/2020.09.08.287268

Van Leenders, G. J. L. H., Van Der Kwast, T. H., Grignon, D. J., Evans, A. J., Kristiansen, G., Kweldam, C. F., Litjens, G., McKenney, J. K., Melamed, J., Mottet, N., Paner, G. P., Samaratunga, H., Schoots, I. G., Simko, J. P., Tsuzuki, T., Varma, M., Warren, A. Y., Wheeler, T. M., Williamson, S. R., & Iczkowski, K. A. (2020). The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *American Journal of Surgical Pathology*, *44*(8), E87–E99. https://doi.org/10.1097/PAS.0000000000001497

Verze, P., Cai, T., & Lorenzetti, S. (2016). The role of the prostate in male fertility, health and disease. *Nature Reviews. Urology*, *13*(7), 379–386. https://doi.org/10.1038/NRUROL.2016.89

Visakorpi, T., Hyytinen, E., Koivisto, P., Tanner, M., Keinänen, R., Palmberg, C., Palotie, A., Tammela, T., Isola, J., & Kallioniemi, O. P. (1995). In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nature Genetics*, *9*(4), 401–406. https://doi.org/10.1038/ng0495-401

Viswanathan, S. R., Ha, G., Hoff, A. M., Wala, J. A., Carrot-Zhang, J., Whelan, C. W., Haradhvala, N. J., Freeman, S. S., Reed, S. C., Rhoades, J., Polak, P., Cipicchio, M., Wankowicz, S. A., Wong, A., Kamath, T., Zhang, Z., Gydush, G. J., Rotem, D., Love,

J. C., … Meyerson, M. (2018). Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. *Cell*, *174*(2), 433-447.e19. https://doi.org/10.1016/j.cell.2018.05.036

Vitorino, R., Guedes, S., Amado, F., Santos, M., & Akimitsu, N. (2021). The role of micropeptides in biology. *Cellular and Molecular Life Sciences: CMLS*, *78*(7), 3285–3298. https://doi.org/10.1007/s00018-020-03740-3

Waltering, K. K., Helenius, M. A., Sahu, B., Manni, V., Linja, M. J., Jänne, O. A., & Visakorpi, T. (2009). Increased expression of androgen receptor sensitizes prostate cancer cells to low levels of androgens. *Cancer Research*, *69*(20), 8141–8149. https://doi.org/10.1158/0008-5472.CAN-09-0919

Wang, H.-L. V., & Chekanova, J. A. (2019). An Overview of Methodologies in Studying lncRNAs in the High-Throughput Era: When Acronyms ATTACK! *Methods in Molecular Biology (Clifton, N.J.)*, *1933*, 1. https://doi.org/10.1007/978-1-4939-9045-0_1

Wang, J., Ye, Q., & She, Q.-B. (2014). New insights into 4E-BP1-regulated translation in cancer progression and metastasis. *Cancer Cell & Microenvironment*, *1*(5), e331. https://doi.org/10.14800/ccm.331

Wang, M. C., Valenzuela, L. A., Murphy, G. P., & Chu, T. M. (1979). Purification of a human prostate specific antigen. *Investigative Urology*, *17*(2), 159–163.

Wang, Q., & Yang, H.-S. (2018). The role of Pdcd4 in tumour suppression and protein translation. *Biology of the Cell*, *110*(8), 169–177. https://doi.org/10.1111/boc.201800014

Wang, S., Gao, J., Lei, Q., Rozengurt, N., Pritchard, C., Jiao, J., Thomas, G. V., Li, G., Roy-Burman, P., Nelson, P. S., Liu, X., & Wu, H. (2003). Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell*, *4*(3), 209–221. https://doi.org/10.1016/s1535-6108(03)00215-0

Wasmuth, E. V., Hoover, E. A., Antar, A., Klinge, S., Chen, Y., & Sawyers, C. L. (2020). Modulation of androgen receptor DNA binding activity through direct interaction with the ETS transcription factor ERG. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8584–8592. https://doi.org/10.1073/pnas.1922159117 [doi]

Wei, G.-H., Badis, G., Berger, M. F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A. R., Yan, J., Talukder, S., Turunen, M., Taipale, M., Stunnenberg, H. G., Ukkonen, E., Hughes, T. R., Bulyk, M. L., & Taipale, J. (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *The EMBO Journal*, *29*(13), 2147–2160. https://doi.org/10.1038/emboj.2010.106

Wei, J. T., Feng, Z., Partin, A. W., Brown, E., Thompson, I., Sokoll, L., Chan, D. W., Lotan, Y., Kibel, A. S., Busby, J. E., Bidair, M., Lin, D. W., Taneja, S. S., Viterbo, R., Joon, A. Y., Dahlgren, J., Kagan, J., Srivastava, S., & Sanda, M. G. (2014). Can urinary PCA3 Supplement PSA in the early detection of prostate cancer? *Journal of Clinical Oncology*, *32*(36), 4066–4072. https://doi.org/10.1200/JCO.2013.52.8505

Weier, C., Haffner, M. C., Mosbruger, T., Esopi, D. M., Hicks, J., Zheng, Q., Fedor, H., Isaacs, W. B., Marzo, A. M. D., Nelson, W. G., & Yegnasubramanian, S. (2013). Nucleotide resolution analysis of TMPRSS2 and ERG rearrangements in prostate cancer. *The Journal of Pathology*, *230*(2), 174–183. https://doi.org/10.1002/path.4186

Wu, S. C., Kallin, E. M., & Zhang, Y. (2010). Role of H3K27 methylation in the regulation of lncRNA expression. *Cell Research*, *20*(10), Article 10. https://doi.org/10.1038/cr.2010.114

Xin, L. (2019). Cells of Origin for Prostate Cancer. *Advances in Experimental Medicine and Biology*, *1210*, 67–86. https://doi.org/10.1007/978-3-030-32656-2_4

Xu, T., Lin, C., Cheng, S., Min, J., Li, L., Meng, X., Huang, C., Zhang, L., Deng, Z., & Li, J. (2018). Pathological bases and clinical impact of long noncoding RNAs in prostate cancer: a new budding star. *Molecular Cancer*, *17*(1), 103. https://doi.org/10.1186/s12943-018-0852-7

Yamane, K., Toumazou, C., Tsukada, Y. ichi, Erdjument-Bromage, H., Tempst, P., Wong, J., & Zhang, Y. (2006). JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell*, *125*(3), 483–495. https://doi.org/10.1016/J.CELL.2006.03.027

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., Zhang, Y., Yang, L., Shan, W., He, Q., Fan, L., Kandalaft, L. E., Tanyi, J. L., Li, C., Yuan, C.-X., Zhang, D., Yuan, H., Hua, K., Lu, Y., … Zhang, L. (2015). Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell*, *28*(4), 529–540. https://doi.org/10.1016/j.ccell.2015.09.006

Yang, G., Lu, X., & Yuan, L. (2014). LncRNA: A link between RNA and cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1839*(11), 1097–1109. https://doi.org/10.1016/J.BBAGRM.2014.08.012

Yang, X., Liu, M., Li, M., Zhang, S., Hiju, H., Sun, J., Mao, Z., Zheng, M., & Feng, B. (2020). Epigenetic modulations of noncoding RNA: a novel dimension of Cancer biology. *Molecular Cancer 2020 19:1*, *19*(1), 1–12. https://doi.org/10.1186/S12943-020-01159-9

Yang, Y., Blee, A. M., Wang, D., An, J., Pan, Y., Yan, Y., Ma, T., He, Y., Dugdale, J., Hou, X., Zhang, J., Weroha, S. J., Zhu, W. G., Wang, Y. A., DePinho, R. A., Xu, W., & Huang, H. (2017). Loss of FOXO1 Cooperates with TMPRSS2-ERG Overexpression to Promote Prostate Tumorigenesis and Cell Invasion. *Cancer Research*, *77*(23), 6524–6537. https://doi.org/10.1158/0008-5472.CAN-17-0686 [doi]

Yao, R.-W., Wang, Y., & Chen, L.-L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology*, *21*(5), 542–551. https://doi.org/10.1038/s41556-019-0311-8

Yao, Y., Wang, H., Li, B., & Tang, Y. (2014). Evaluation of the TMPRSS2:ERG fusion for the detection of prostate cancer: a systematic review and meta-analysis. *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, *35*(3), 2157–2166. https://doi.org/10.1007/s13277-013-1286-x

Yeo, S. J., Ying, C., Fullwood, M. J., & Tergaonkar, V. (2023). Emerging regulatory mechanisms of noncoding RNAs in topologically associating domains. *Trends in Genetics*, *39*(3), 217–232. https://doi.org/10.1016/j.tig.2022.12.003

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J. L., De, S., Huarte, M., Zhan, M., Becker, K. G., & Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. *Molecular Cell*, *47*(4), 648–655. https://doi.org/10.1016/j.molcel.2012.06.027

Yoshida, T., Naito, Y., Yasuhara, H., Sasaki, K., Kawaji, H., Kawai, J., Naito, M., Okuda, H., Obika, S., & Inoue, T. (2019). Evaluation of off-target effects of gapmer antisense oligonucleotides using human cells. *Genes to Cells*, *24*(12), 827–835. https://doi.org/10.1111/gtc.12730

Yu, J., Yu, J., Mani, R.-S., Cao, Q., Brenner, C. J., Cao, X., Wang, X., Wu, L., Li, J., Hu, M., Gong, Y., Cheng, H., Laxman, B., Vellaichamy, A., Shankar, S., Li, Y., Dhanasekaran, S. M., Morey, R., Barrette, T., … Chinnaiyan, A. M. (2010). An integrated network of

androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*, *17*(5), 443–454. https://doi.org/10.1016/j.ccr.2010.03.018

Zelefsky, M. J., Eastham, J. A., Cronin, A. M., Fuks, Z., Zhang, Z., Yamada, Y., Vickers, A., & Scardino, P. T. (2010). Metastasis after radical prostatectomy or external beam radiotherapy for patients with clinically localized prostate cancer: a comparison of clinical cohorts adjusted for case mix. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *28*(9), 1508–1513. https://doi.org/10.1200/JCO.2009.22.2265

Zennami, K., Choi, S. M., Liao, R., Li, Y., Dinalankara, W., Marchionni, L., Rafiqi, F. H., Kurozumi, A., Hatano, K., & Lupold, S. E. (2019). PDCD4 Is an Androgen-Repressed Tumor Suppressor that Regulates Prostate Cancer Growth and Castration Resistance. *Molecular Cancer Research*, *17*(2), 618–627. https://doi.org/10.1158/1541-7786.MCR-18-0837

Zhang, A., Zhao, J. C., Kim, J., Fong, K.-W., Yang, Y. A., Chakravarti, D., Mo, Y.-Y., & Yu, J. (2015). LncRNA HOTAIR Enhances the Androgen-Receptor-Mediated Transcriptional Program and Drives Castration-Resistant Prostate Cancer. *Cell Reports*, *13*(1), 209–221. https://doi.org/10.1016/j.celrep.2015.08.069

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *International Journal of Molecular Sciences 2019, Vol. 20, Page 5573*, *20*(22), 5573. https://doi.org/10.3390/IJMS20225573

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Y., Pitchiaya, S., Cieślik, M., Niknafs, Y. S., Tien, J. C.-Y., Hosono, Y., Iyer, M. K., Yazdani, S., Subramaniam, S., Shukla, S. K., Jiang, X., Wang, L., Liu, T.-Y., Uhl, M., Gawronski, A. R., Qiao, Y., Xiao, L., Dhanasekaran, S. M., Juckette, K. M., … Chinnaiyan, A. M. (2018). Analysis of the androgen receptor–regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nature Genetics*, *50*(6), Article 6. https://doi.org/10.1038/s41588-018-0120-1

Zhao, S. G., Chen, W. S., Li, H., Foye, A., Zhang, M., Sjöström, M., Aggarwal, R., Playdle, D., Liao, A., Alumkal, J. J., Das, R., Chou, J., Hua, J. T., Barnard, T. J., Bailey, A. M., Chow, E. D., Perry, M. D., Dang, H. X., Yang, R., … Feng, F. Y. (2020). The DNA methylation landscape of advanced prostate cancer. *Nature Genetics*, *52*(8), 778–789. https://doi.org/10.1038/s41588-020-0648-8

Zuckerman, B., Ron, M., Mikl, M., Segal, E., & Ulitsky, I. (2020). Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex. *Molecular Cell*, *79*(2), 251-267.e6. https://doi.org/10.1016/j.molcel.2020.05.013

PUBLICATIONS

PUBLICATION I

Transcriptome Sequencing Reveals *PCAT5* as a Novel ERG-Regulated Long Noncoding RNA in Prostate Cancer

Antti Ylipää*, Kati Kivinummi*, Annika Kohvakka, Matti Annala, Leena Latonen, Mauro Scaravilli, Kimmo Kartasalo, Simo-Pekka Leppänen, Serdar Karakurt, Janne Seppälä, Olli Yli-Harja, Teuvo L. J. Tammela, Wei Zhang, Tapio Visakorpi, and Matti Nykter

Cancer Research, 2015, 75:4026–4031
doi: 10.1158/0008-5472.CAN-15-0217

*These authors contributed equally to the work.

Priority Report

Cancer Research

Transcriptome Sequencing Reveals *PCAT5* as a Novel ERG-Regulated Long Noncoding RNA in Prostate Cancer

Antti Ylipää[1,2], Kati Kivinummi[1,2], Annika Kohvakka[2,3], Matti Annala[1,2], Leena Latonen[2,3], Mauro Scaravilli[2,3], Kimmo Kartasalo[1,2], Simo-Pekka Leppänen[1,2], Serdar Karakurt[2,3], Janne Seppälä[1,2], Olli Yli-Harja[1], Teuvo L.J. Tammela[4], Wei Zhang[5], Tapio Visakorpi[2,3], and Matti Nykter[1,2]

Abstract

Castration-resistant prostate cancers (CRPC) that arise after the failure of androgen-blocking therapies cause most of the deaths from prostate cancer, intensifying the need to fully understand CRPC pathophysiology. In this study, we characterized the transcriptomic differences between untreated prostate cancer and locally recurrent CRPC. Here, we report the identification of 145 previously unannotated intergenic long noncoding RNA transcripts (lncRNA) or isoforms that are associated with prostate cancer or CRPC. Of the one third of these transcripts that were specific for CRPC, we defined a novel lncRNA termed *PCAT5* as a regulatory target for the transcription factor ERG, which is activated in approximately 50% of human prostate cancer. Genome-wide expression analysis of a *PCAT5*-positive prostate cancer after *PCAT5* silencing highlighted alterations in cell proliferation pathways. Strikingly, an *in vitro* validation of these alterations revealed a complex integrated phenotype affecting cell growth, migration, invasion, colony-forming potential, and apoptosis. Our findings reveal a key molecular determinant of differences between prostate cancer and CRPC at the level of the transcriptome. Furthermore, they establish *PCAT5* as a novel oncogenic lncRNA in ERG-positive prostate cancers, with implications for defining CRPC biomarkers and new therapeutic interventions. *Cancer Res; 75(19); 4026–31. ©2015 AACR.*

Introduction

The most frequent genomic lesion in prostate cancers is deletion of 21q22 in 50% of cases resulting in overexpression of ERG, an ETS family transcription factor. A translocation following the deletion fuses the regulatory sequence of an androgen-regulated gene, most often TMPRSS2, with the protein coding sequence of ERG bringing it under androgen regulation (1). *ERG* is a critical proto-oncogene that disrupts the ability of the cells to differentiate when activated. *ERG* fusions also contribute to development of androgen independence in prostate cancer through inducing repressive epigenetic programs via activation of a Polycomb methylatransferase EZH2, inhibiting androgen receptor (AR) expression, and disruption of AR signaling (2). Overexpression of ERG, or other ETS transcription factors, such as ETV1, and ETV4 activates cell invasion programs (3). ETS-negative prostate cancers have rare alternate driving events, such as SKIL-activating rearrangements (4). Generally, the molecular mechanisms of action for ERG are yet to be fully understood.

Recently, long noncoding RNA (lncRNA) molecules that are mainly transcribed from the intergenic regions of the genome (lncRNAs) have become a focus in transcriptome studies of cancers (5). These molecules form an integral part of many biologic processes, often through interactions with the Polycomb complex, which lead to silencing tumor-suppressive functions (6), but many other mechanisms have also been described previously (7). Few prostate cancer–specific lncRNAs have been well characterized to date, particularly *PCGEM1* (8), *PRNCR1* (9), *PCAT1* (10) and *SChLAP1* (11). *PCAT1* is a regulator of cell proliferation and a target of the Polycomb-Repressive Complex 2 (PRC2) that represses *BRCA2* tumor-suppressor and controls homologous recombination (12). *SChLAP1* antagonizes the regulatory functions of the SWI/SNF chromatin-modifying complex leading to increased invasiveness and metastasis *in vitro*, and its expression predicts poor outcome in clinical setting (11).

We hypothesized that there is still a significant amount of previously unexplored transcriptomic differences between hormone-naive and castration-resistant prostate cancer (CRPC), especially in the expression patterns of lncRNAs. To conduct the first comprehensive characterization of protein-coding genes, small RNAs and lncRNAs in these prostate cancers, we

[1]Department of Signal Processing, Tampere University of Technology, Tampere, Finland. [2]Institute of Biosciences and Medical Technology—BioMediTech, University of Tampere, Tampere, Finland. [3]Fimlab Laboratories, Tampere University Hospital, Tampere, Finland. [4]Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland. [5]Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, Texas.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

A. Ylipää and K. Kivinummi contributed equally to this work.

Corresponding Author: Matti Nykter, University of Tampere, Biokatu 6, Tampere 33520, Finland. Phone: 358-50-318-6869; Fax: 358-3-364-1291; E-mail: matti.nykter@uta.fi and Tapio Visakorpi, University of Tampere, Biokatu 6, Tampere 33520 Finland. Phone: 358-40-717-4402; Fax: 358-3-364-1247; E-mail: tapio.visakorpi@uta.fi

doi: 10.1158/0008-5472.CAN-15-0217

deep-sequenced transcriptomes of 12 benign prostatic hyperplasias (BPH), 28 untreated prostate cancers, and 13 CRPCs. In addition to identifying several CRPC-specific lncRNAs, we discovered PCAT5, an ERG-regulated tumor growth–associated lncRNA, that is exclusively expressed in ERG-positive prostate cancers and CRPCs. Its functional association in prostate cancer progression may partly explain how ERG exerts its widespread effect in gene regulation.

Materials and Methods

Patient samples and sequencing

Fresh-frozen tissue specimens from 12 BPHs, 28 prostate cancers, and 13 CRPCs were acquired from Tampere University Hospital (Tampere, Finland). The BPHs included both transition zone ($n = 4$) and peripheral zone ($n = 8$) samples received either by transurethral resection of the prostate (TURP) or cystoprostatectomy, respectively, from patients without prostate cancer diagnosis. All cancer samples contained a minimum of 70% cancerous or hyperplastic epithelial cells. Prostate cancer samples were obtained by radical prostatectomy and locally recurrent CRPCs by TURP sequenced. Libraries were prepared for paired-end analysis on the Illumina HiSeq 2000. On average, we obtained 110 million 90 bp-long paired-end reads from the whole transcriptome sequencing (RNA-seq), and 8.2 million 50 bp-long single-end reads from the small RNA sequencing (sRNA-seq). The sequencing reads were subsequently aligned to the human genome, and expression estimates for all expressed transcripts were computed. On average were able to align 92% of the reads (and minimum of 84%), indicating sufficient quality reads from all samples (Supplementary Table S1). More detailed description of the experimental setup can be found in Supplementary Methods.

Transcriptome assembly

To fully characterize the wealth of expressed transcripts in different stages of prostate cancer, we assembled a consensus transcriptome from all the samples using Cufflinks (13) RABT (reference annotation–based transcript) assembly approach with NCBI 37.2/hg19 genome build. Comparing the assembled prostate cancer transcriptome with all the exonic and intronic sequences in human reference transcriptomes (UCSC hg19, NCBI build 37.2, Ensembl GRCh37, Gencode version 12e) resulted in identification of 99,120 novel loci of expression. Transcripts overlapping exonic sequences were labeled as known sequences (and not included in the 99,120 novel loci), transcripts fully contained in an intron were labeled as intragenic (32,744; 33%), and transcripts not overlapping with exonic or intronic sequences were labeled as intergenic (66,376; 67%). To reduce the effect of noise, we filtered the lowly expressed transcripts (maximum normalized read count under 500), and included only transcripts that were differentially expressed across the tumor types using a negative binomial test and Mann–Whitney U test (adjusted $P < 0.001$ for both tests). Filtering reduced the number of loci to 152 intergenic and 25 intragenic prostate cancer–associated novel loci of transcription that were expressed at a significant level and were differentially expressed between BPH and prostate cancer or prostate cancer and CRPC samples (Supplementary Table S2). More detailed description of the data analysis can be found in Supplementary Methods.

While taking into account the computationally predicted sequences of the transcripts, we manually inferred putative exon structures, different isoforms, and strandedness for 145 transcripts or isoforms merging some of the adjacent loci of transcription. The curation from 152 loci into 145 isoforms was made based on the recurrent splice junctions in the paired-end read data coinciding with canonical intron splice site motifs. We were able to infer these structural details only for about half of the loci. The rest of the loci may either encode functional single-exon transcripts or be parts of ambiguously expressed large genomic regions such as *SChLAP1* (11). Following the previously adopted nomenclature, we named the transcripts tentatively as TPCATs (Tampere prostate cancer–associated transcripts) followed by chromosome and locus identifications (Supplementary Table S2). The annotation process is described in Supplementary Methods.

Results

Comprehensive transcriptome analysis reveals alterations in key regulatory pathways

We integrated the sequencing data into a comprehensive view of the prostate cancer and CRPC transcriptomes. Hierarchical clustering (Fig. 1A) and principal component analysis (PCA; Fig. 1B) of gene-expression profiles separated BPH, prostate cancer, and CRPC samples into distinct clusters. From PCA analysis, we observed two additional clusters that represented cancers with special features: One cluster contained two AR-negative tumors, whereas another contained tumors with strong AR amplification. We looked for differentially expressed genes using the Mann–Whitney U test with threshold for significant difference $P < 0.0001$, and absolute difference between medians of length-normalized read counts above 200 and $\log_2$-ratio above 1. In total, we identified 798 genes and 20 small RNAs differentially expressed between BPH and prostate cancer, and 330 genes and 43 small RNAs between prostate cancer and CRPC (Supplementary Table S3). When pathway analysis was run using the genes that were differentially expressed between prostate cancer and BPH, cytochrome p450 metabolism, cell adhesion, and TGFβ signaling pathways were identified as aberrated. Altered processes between CRPC and prostate cancer were dominated by regulatory pathways in which NR4A1, EGR family, FOS, DUSP1, and ATF3 play a key role (Supplementary Table S3). These genes were overexpressed in prostate cancers but not in CRPCs, and their mutual correlation (Pearson correlation > 0.9) indicated potentially shared regulation.

To highlight the pathway level changes in cell-cycle and androgen regulation, we constructed pathway models of these processes and projected the observed expression changes onto these models. In cell cycle, we noted a strong combined overexpression of the proliferation markers *MKI67*, *TOP2A*, *AURKA*, and *EZH2* in half of CRPCs, suggesting a high proliferation rate in these tumors. This high proliferation rate was also reflected in the expression of cyclins *CCNB1*, *CCNB2* and *CCNE2*, and the cyclin-dependent kinase *CDK1* (Supplementary Fig. S1). In the androgen regulation pathway, we observed overexpression of AR in 7 of 13 CRPCs. The AR coactivator *FOXA1* was overexpressed in untreated prostate cancer relative to BPH. Isozymes *SRD5A1* and *SRD5A2*, responsible for testosterone-to-DHT conversion, showed respective up- and downregulation in CRPC. Enzymes *AKR1C3* and *AKR1C2*, responsible for canonical androstenedione-to-testosterone

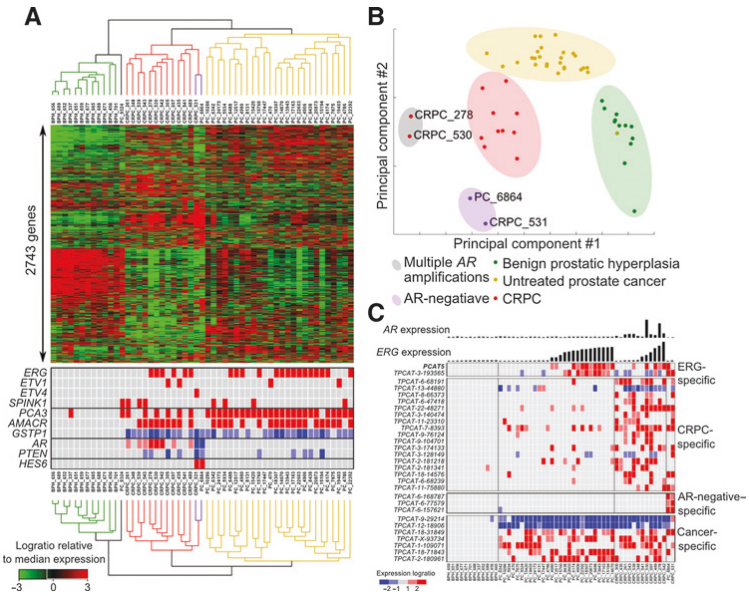


Figure 1.
Expression level characterization of prostate cancers. A, hierarchical clustering of annotated genes reveals distinct gene-expression signatures for BPH (green), prostate cancer (PC; yellow), and locally recurrent CRPC (red). High expressions of key marker genes, such as *ERG* and *AR*, have been highlighted for all the tumors in red and low expression in blue, different levels of shade indicating the level of expression difference from the median. Two tumors (PC_6864 and CRPC_531, purple) were negative for *AR* expression and positive for neuronal differentiation marker *HES6*. B, principal component analysis. BPH, prostate cancer, and CRPC samples are well separated into distinct clusters based on their gene-expression profiles. AR negative tumors (PC_6864 and CRPC_531) as well as tumors with strong *AR* amplification and overexpression (CRPC_530 and CRPC_278) formed separate clusters. C, association of found transcripts to disease phenotypes. A number of transcripts were associated with *ERG*-positive prostate cancers and CRPCs, including *PCAT5*. In addition, many transcripts were CRPC-specific and few showed specificity to untreated prostate cancers. Most transcripts were found in all cancer tissue types.

reduction, were overexpressed in 30% to 50% of CRPCs, with associated overexpression of *UGT2B15* and *UGT2B17*, enzymes responsible for glucuronidation of testosterone and DHT. Transcription factors *ERG* and *ETV1* were overexpressed in 25 of 41 prostate cancers corresponding to previously established frequency of fusions with the androgen regulated *TMPRSS2* (Supplementary Fig. S2; ref. 1).

The expression patterns of novel lncRNAs differentiate between prostate cancer and CRPC

Majority of the novel expressed loci were detected in CRPC only or in both prostate cancer and CRPC, but a few loci were expressed in all three sample groups, albeit at different rates, or were specific to the AR-negative samples (Fig. 1C). More than 30% of the transcripts were expressed on average at more than 10 times higher level in CRPCs than in prostate cancers or BPHs, which we considered highly CRPC-specific expression pattern. Some of the transcripts were expressed in only few samples corresponding to outlier expression pattern. The specificities of the TPCAT expression patterns were further validated using available RNA-sequencing data from 21 prostate cancer cell lines (10), 24 normal tissues (14), 2 human embryonic stem cells (PolyA-selected and non-selected; ref. 15), and two independent cohorts of prostate cancer tumors ($n = 30$ and $n = 34$, respectively; Fig. 2A and Supplementary Table S2; refs. 10, 16). Generally, TPCATs were minimally expressed in normal tissues, with testes most commonly being the normal tissue with the highest expression level. The expression patterns in cancer tissue were generally concordant in all three prostate cancer cohorts.

To investigate whether changes in DNA methylation or copy number bring about the expression of the novel transcripts in the samples that express them, we integrated DNA-sequencing and MeDIP-sequencing data from the same samples with the RNA-seq data (See Supplementary Methods). We computed Spearman correlations between transcript expression values and the copy number of the locus, and expression values and methylation values of nearby differentially methylated regions. In addition, we tested for differential expression between samples that had copy-number aberrations at the locus versus samples that had normal copy number for each TPCAT using the *t* test. We required a significant correlation between the expression and copy number, and significantly differential expression between samples with normal copy number and samples with copy-number aberration. Similar requirements were applied to methylation. None of the TPCATs were found to be significant taking account both criteria, indicating that the expression differences of TPCATs were

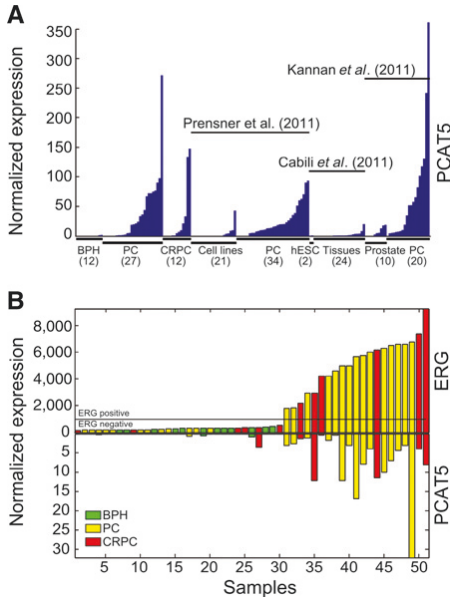


Figure 2.
Expression patterns of *PCAT5*. A, association of *PCAT5* expression to prostate cancer was confirmed by analyzing additional prostate cancer and healthy tissue RNA-seq datasets. B, *ERG* and *PCAT5* expressions derived from our RNA-seq data were plotted for each sample. A cutoff of 1,000 RPKs was chosen to differentiate ERG-negative and ERG-positive samples. The *ERG*-negative pool contained all the BPH samples, 12 prostate cancers, and 6 CRPC whereas the ERG-positive pool contained 15 prostate cancers and 6 CRPC samples.

not explained by these factors. This suggesting that, at least in general, TPCATs are transcriptionally regulated and that their expression does not arise due to genetic alterations or changes in DNA methylation (Supplementary Table S4).

We wanted to find prostate cancer–associated transcription factors that could act partly by regulating some of the lncRNAs we discovered. Spearman correlations were computed between the expression of the lncRNAs and eight transcription factors (*ERG*, *AR*, *FOXA1*, *EZH2*, *HDAC1*, *HDAC2*, *HDAC3*, and *RUNX2*), for which public ChIP-sequencing data in prostate cancer cell lines were available for validating the regulatory association. The correlation analysis associated the expression of several TPCATs with the expression of these transcriptional regulators (Supplementary Table S4). The strongest positive correlation ($r = 0.69$) was observed between *ERG* and transcript *TPCAT-10-36067* (officially termed *PCAT5*; Fig. 1C). Concordantly with *ERG* expression, *PCAT5* was expressed in a subset of prostate cancers and CRPCs (Fig. 2A and B). The expression was detected at a comparable frequency in both independent cohorts of prostate cancer, but not significantly in healthy tissues, including BPHs. The expression of *PCAT5* was quantified and validated in independent cohort of 76 primary prostate cancer samples as well as *ETV4*-positive PC-3 and *ERG*-positive VCaP cells using qRT-PCR (Supplementary Fig. S3). In addition, and the expression correlation between *PCAT5* and *ERG* was validated in this 76 sample set with ERG immunohistochemistry, and in LuCaP xenografts with qRT-PCR ($r = 0.78$; Supplementary Fig. S3). Expressions of four additional CRPC-expressed multiexon TPCATs were also validated with RT-PCR (Supplementary Figs. S4 and S5). Because ERG is a dominant feature in prostate cancers, we decided to concentrate our validation efforts to deciphering the exact structure of *PCAT5*, elucidating the regulatory connection between *PCAT5* and ERG, and investigating the functional relevance of *PCAT5*.

Inhibition of *PCAT5* expression reduces growth, migration, and invasion of ERG-positive prostate cancer cells

On the basis of the spliced read alignments, we inferred a three-exon structure for *PCAT5* with no components of viral open reading frames or other repetitive elements located on the exons (Fig. 3A). Both exon–exon junctions were validated with RT-PCR and Sanger sequencing in three clinical samples (Supplementary Fig. S3). To accurately identify both termini of the transcript, we performed 5′ and 3′ rapid amplification of cDNA ends (RACE). Open reading frame analysis indicated that the

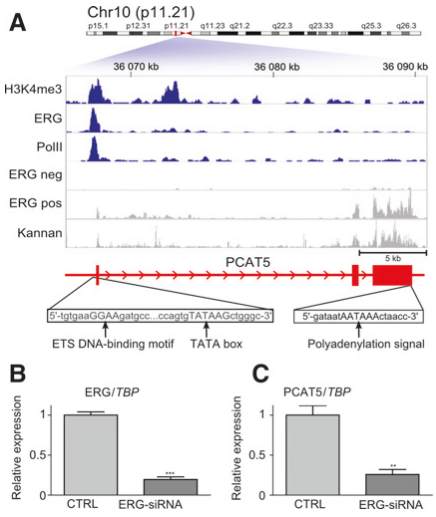


Figure 3.
Sequence analysis of *PCAT5*, located in 10p11.21. A, regulatory properties of *PCAT5* were investigated using ChIP-sequencing data of H3K4me3, ERG, and POL2 from *PCAT5*-positive VCaP cells (blue), and detailed expression profiles in *ERG*-negative and *ERG*-positive samples, and in an independent prostate cancer cohort (gray). The data were overlaid onto the inferred exon structure of *PCAT5* (in red), indicating open chromatin coinciding with ERG- and POL2-binding events in the promoter of VCaP cells, and active expression in *ERG*-positive samples compared with *ERG*-negative samples. An ETS DNA–binding domain and TATA box were identified at the suspected promoter region and a Poly-A signal sequence at the end of third exon. *ERG* (B) and *PCAT5* (C) expressions after 50 nmol/L ERG-siRNA or scrambled CTRL-siRNA treatment in VCaP cells; error bars, SEM; **, $P < 0.0083$; ***, $P < 0.001$, unpaired two-tailed *t* test ($n = 3$).

transcript lacks protein-coding potential. From available ChIP-sequencing data measured from ERG-positive VCaP cells (Supplementary Table S4), we identified open chromatin histone markers, such as H3K4 trimethylation, and binding events of ERG and RNA polymerase II at proximal promoter of *PCAT5* (Fig. 3A). Conversely, no ERG binding or H3K4 trimethylation was found at the *PCAT5* promoter in LNCaP cells, which do not express *PCAT5* (Supplementary Fig. S6). Sequence analysis revealed a canonical ETS family DNA-binding motif and a TATA-box coinciding with the locus that ERG was bound to, and a polyadenylation signal at the 3′-end of the transcript (Fig. 3A). We further validated the regulatory association by knocking down *ERG* in VCaP cells using an siRNA (Fig. 3B), which led to 75% inhibition of *PCAT5* expression (Fig. 3C). Similarly, we validated the association between *PCAT5* and another ETS-family transcription factor, ETV4, by knocking it down in ERG-negative PC-3 cells, leading to comparable inhibition of PCAT5 expression (Supplementary Fig. S7). These data indicate that *PCAT5* is under direct regulation by ERG, and likely other ETS family transcription factors as well.

To characterize and validate the function of *PCAT5*, we suppressed its expression with two different siRNAs in two cell lines: PC-3 cells, which expressed the transcript (Fig. 4A), and 22Rv1 cells, which did not express it. The genome-wide expression changes that the suppression induced to PC-3 cells were studied using expression arrays. Gene ontology enrichment analysis indicated that cell-cycle, mitosis, and Aurora signaling were the most extensively affected processes (Supplementary Table S5). Several functional assays validated the computationally identified biologic processes: The knockdown dramatically decreased cell growth (Fig. 4B) and invasiveness (Fig. 4C), and increased the rate of apoptosis (Fig. 4D). In addition, colony formation (Fig. 4E) and migration potential (Fig. 4F) of the transfected PC-3 cells decreased substantially compared with non-transfected PC-3 cells. Conversely, the growth rate of 22Rv1 cells that do not express *PCAT5* was unaffected by the siRNA suppression as expected (Supplementary Fig. S8) whereas the growth of ERG-positive DuCaP cells decreased after siRNA suppression of *PCAT5* (Supplementary Fig. S9). These results suggest that *PCAT5* has a key role in regulating tumor growth and malignancy in ETS-positive prostate cancers.

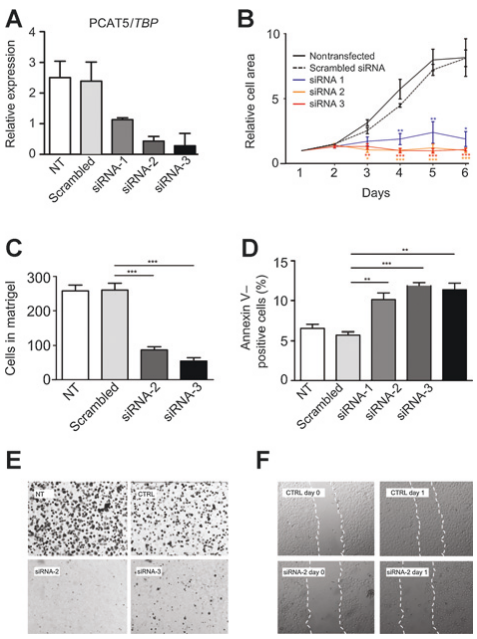


Figure 4.
Functional validation of *PCAT5*. A, successful silencing of *PCAT5* using multiple siRNA knockdowns was validated using qRT-PCR. B, growth of the *PCAT5*-positive PC-3 cell line was completely inhibited by siRNAs targeting the transcript. C, invasiveness of PC-3 cells was reduced by the siRNA knockdown. D, Annexin V assay indicated an increased rate of apoptosis after the knockdown; error bars, SEM; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$, unpaired two-tailed *t* test. E, colony formation was significantly reduced in the *PCAT5*-deficient PC-3 cells. F, wound-healing assay (triplicates, representative experiment shown) indicated a significantly impaired migration ability.

Discussion

Hundreds of lncRNAs, for which little more than an expression pattern is known, have been discovered by RNA-sequencing and stored in databases such as NONCODE (17). A growing interest toward lncRNAs in cancer research is sparked by the dozens of molecules that have been implicated as key players in cancer cells (5). In prostate tumorigenesis, differential expression of hundreds of lncRNAs is already a recognized phenomenon (8, 9, 11). However, the functional role of many cancer-associated lncRNAs remains undetermined. The expression of lncRNA may confer clinical information about disease outcomes, and thus have utility as diagnostic tests. One prostate cancer–specific biomarker lncRNA, PCA3, is currently in use (18). Evidence for effectively targeting tumor-specific lncRNAs as a therapeutic regimen (19), such as the telomerase lncRNA *TERC*, are accumulating. Therefore, the characterization of the noncoding RNA species and their functions are clinically important.

To identify novel transcripts, a comparable experimental and computational approach was taken by Prensner and colleagues (10), which resulted in discovery of 121 lncRNAs in untreated prostate cancer. Here, we extended the list of prostate tumor–specific transcripts with 145 distinct molecular entities by including CRPC samples to the cohort and performing much deeper sequencing. Outlier-type expression patterns of many lncRNAs discovered in this article may explain why many of them had not been discovered to this date despite the use of RNA-sequencing technologies. Furthermore, many lncRNAs were expressed in moderate-to-low levels, which may have caused previous studies to overlook them. Also, PCAT5 that was identified as a key molecule in ERG-positive prostate cancers is quite lowly but consistently transcribed in other cohorts.

Integration with DNA-seq and MeDIP-seq data indicated that the expression of none of the novel transcripts correlated with copy-number or DNA methylation status, and thus it seems that the expression of these transcripts is not driven by copy-number changes or differential DNA methylation. For example, in cell lines that have open chromatin at *PCAT5* promoter likely express

it due to a binding of an ETS family transcription factor. The mechanism is intriguing because *ERG* overexpression is one of the hallmark events of prostate cancer, whereas the mechanisms downstream of ERG remain poorly understood. Our siRNA experiments revealed that *PCAT5* affects both cell growth and invasiveness, which suggests that the transcript may be an integral mediator in the regulatory cascade downstream of ERG. Although low-expression level is probably not ideal for a biomarker, the strong phenotype combined with prostate cancer specificity, makes *PCAT5* a prospective target for therapy.

In conclusion, we performed the first transcriptomic analysis on CRPC and identified more than hundred novel lncRNAs that seem to be specific for either prostate cancer or CRPC. One of the lncRNAs, *PCAT5*, was shown to be regulated by ERG and has a dramatic effect on prostate cancer cells. Inclusion of more specimens, especially CRPCs, would likely result in identification of even more novel lncRNAs with outlier type of expression pattern. Biopsies of metastases might also reveal novel lncRNAs that are specific to these tumors and ones that are not expressed in primary tumors. However, the identified transcripts form an interesting pool of putative biomarker and mechanisms for prostate cancer progression.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A. Ylipää, K. Kivinummi, W. Zhang, T. Visakorpi, M. Nykter

Development of methodology: A. Ylipää, K. Kivinummi, S. Karakurt, M. Nykter

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. Kivinummi, A. Kohvakka, L. Latonen, M. Scaravilli, K. Kartasalo, S.-P. Leppänen, T.L.J. Tammela, T. Visakorpi

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Ylipää, K. Kivinummi, M. Annala, L. Latonen, M. Scaravilli, K. Kartasalo, J. Seppälä, W. Zhang, T. Visakorpi, M. Nykter

Writing, review, and/or revision of the manuscript: A. Ylipää, K. Kivinummi, A. Kohvakka, T.L.J. Tammela, W. Zhang, T. Visakorpi

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. Kartasalo, S.-P. Leppänen, O. Yli-Harja, T. Visakorpi

Study supervision: T.L.J. Tammela, T. Visakorpi, M. Nykter

Acknowledgments

The authors express gratitude to Dr. Robert L. Vessella for the LuCaP xenograft data. They also thank Paula Kosonen, Riina Kylätie, Päivi Martikainen, and Marika Vähä-Jaakkola for their technical assistance and CSC—IT Center for Science for providing computational resources.

Grant Support

The work was supported by grants from the Finnish Funding Agency for Technology and Innovation Finland Distinguished Professor program (O. Yli-Harja and M. Nykter), Academy of Finland (project nos. 269474 and 132877 to M. Nykter and project no. 251790 to O. Yli-Harja, and project no. 127187 to T. Visakorpi), Sigrid Juselius Foundation (M. Nykter and T. Visakorpi), Emil Aaltonen Foundation (M. Annala), Cancer Society of Finland (M Nykter and T. Visakorpi). Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (grant 9P053 to T.L.J. Tammela and grant 9N087 to T. Visakorpi), and EU-FP7 Marie Curie Integrated Training Network, PRO-NEST (T. Visakorpi), the NIH (U24CA143835; to W. Zhang).

Received January 29, 2015; revised May 19, 2015; accepted July 2, 2015; published OnlineFirst August 17, 2015.

References

1. 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.
2. 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.
3. 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.
4. Annala M, Kivinummi K, Tuominen J, Karakurt S, Granberg K, Latonen L, et al. Recurrent SKIL-activating rearrangements in ETS negative prostate cancer. Oncotarget 2015;6:6235–50.
5. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. Cancer Discov 2011;1:391–407.
6. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007;129:1311–23.
7. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904–14.
8. Srikantan V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L, et al. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. Proc Natl Acad Sci U S A 2000;97:12216–21.
9. Chung S, Nakagawa H, Uemura M, Piao L, Ashikawa K, Hosono N, et al. Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. Cancer Sci 2011;102:245–52.
10. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol 2011;29:742–9.
11. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, et al. The long noncoding RNA SchLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet 2013;45:1392–8.
12. Prensner JR, Chen W, Iyer MK, Cao Q, Ma T, Han S, et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. Cancer Res 2014;74:1651–60.
13. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511–5.
14. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;24:1915–27.
15. ENCODE Project Consortium Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
16. Kannan K, Wang L, Wang J, Ittmann MM, Li W, Yen L. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. Proc Natl Acad Sci U S A 2011;108:9172–7.
17. Bu D, Yu K, Sun S, Xie C, Skogerbø G, Miao R, et al. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res 2012;40:D210–5.
18. Hessels D, Schalken JA. The use of PCA3 in the diagnosis of prostate cancer. Nat Rev Urol 2009;6:255–61.
19. Li CH, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. Int J Biochem Cell Biol 2013;45:1895–910.

Supplementary Methods

Sample description

Fresh-frozen tissue specimens from 12 benign prostate hyperplasias, 28 untreated prostate cancers, and 13 castration resistant prostate cancers were acquired from Tampere University Hospital (Tampere, Finland). BPH samples were obtained by radical prostatectomy, cystoprostatectomy and by transurethral resection of the prostate (TURP). Untreated PC samples were obtained by radical prostatectomy and locally recurrent CRPCs by transurethral resection of the prostate (TURP). Samples were snap-frozen and stored in liquid nitrogen. Histological evaluation and Gleason grading were performed by a pathologist based on hematoxylin/eosin-stained slides. All samples contained a minimum of 70% cancerous or hyperplastic cells. CRPC patients had been treated with orchiectomy (6 cases), luteinizing hormone releasing hormone analog (LHRH, 3 cases), bicalutamide and orchiectomy (1 case), LHRH and bicalutamide (1 case) or with estrogen (1 case). The mean age at diagnosis was 60.8 years (range: 47.4-71.8), mean PSA at diagnosis was 10.8 ng/ml (range: 3.5-48.1), and the median time from the onset of androgen ablation to
castration resistant progression was 43.5 months (range: 15-81) (**Supplementary Table 1**).

Sample preparation

Fresh-frozen tissue blocks were cut into 10x20-micrometer sections using a cryotome. RNA and DNA were isolated simultaneously using an AllPrep RNA/DNA minikit (Qiagen, Valencia, CA, USA) according to manufacturer's protocol. For some samples, more total RNA was isolated using Trizol (Invitrogen, Carlsbad, CA, USA) extraction according to manufacturer's protocol. Three CRPC samples had RNA extracted using both Trizol and Qiagen AllPrep. The isolated RNA was quantified by 260nm absorbance and its purity assessed by the 260nm/280nm ratio. Integrity was checked using Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Reverse transcription of RNA to cDNA from clinical samples was carried out using SuperScript(TM)III (Invitrogen, Carlsbad, CA, USA) reverse transcriptase and AMV (Finnzymes, Espoo, Finland) for cell line samples with random hexamer-primers according to manufacturer's instructions.

Whole transcriptome library construction and sequencing

Total RNA for each sample was converted into a library of template molecules for sequencing on the Illumina HiSeq2000 at BGI, Hong Kong. Beads with Oligo(dT) were used to isolate poly(A) mRNA after collection of total RNA. Fragmentation buffer were added for interrupting mRNA to short fragments and used to synthesize the first-strand cDNA with random hexamer-primers. The second-strand cDNA was synthesized using buffer, dNTPs, RNaseH, and DNA polymerase I, respectively. Short fragments were purified with QiaQuick PCR extraction kit and resolved with EB buffer for end reparation and adding poly(A). After that, the short fragments were ligated to sequencing adapters and suitable fragments were selected for the PCR amplification as templates and separated with agarose gel electrophoresis before sequencing. Raw image files were processed by Illumina pipeline for base-calling with default

parameters resulting in 90-basepair-long paired-end reads. Reads with too many N bases (>10%) or low base quality (>50% bases with base quality <5) were discarded. 110 million good quality reads per sample were obtained, on average (**Supplementary Table 1**).

Small RNA library construction and sequencing

Small RNA (sRNA) library preparation was performed according to the manufacturer's protocol in "Preparing Samples for Analysis of Small RNA". Briefly, total RNA was size fractionated 18 to 30 nt. Resulting sRNAs were gel-purified and ligated to the 3' adaptor, leading to 36-50nt fragments. Next 5' adaptor was ligated giving 62-75nt fragments. Ligation products were gel-purified, reverse transcribed, and PCR amplified using Illumina's sRNA primer set. Libraries were sequenced using the Illumina HiSeq 2000 platform at BGI, Hong Kong. Raw image files were processed by Illumina pipeline for base-calling with default parameters resulting in 50-bp-long single-end reads. Reads with too many N bases (>10%) or low base quality (>50% bases with base quality <5) were discarded. 8.2 million good quality reads per sample were obtained, on average (**Supplementary Table 1**).

Validation cohort

76 additional hormonally untreated PC prostatectomy samples were acquired from the Tampere University Hospital (Tampere, Finland) (**Supplementary Table 1**). Samples were snap-frozen and stored in liquid nitrogen. Histological evaluation and Gleason grading were performed by a pathologist based on hematoxylin/eosin-stained slides. Samples were confirmed to contain a minimum of 70% cancerous or hyperplastic cells by hematoxylin-eosin staining. Mean age at diagnosis was 62.1 years (range: 47.4-71.8), mean PSA at diagnosis was 11.8 (range: 3.15-51.5). The use of clinical material was approved by the ethical committee of the Tampere University Hospital and the National Authority for Medicolegal Affairs. Written informed consent was obtained from the subjects donating freshly frozen samples.

Cell lines and xenografts

Prostate cancer cell lines PC-3, LNCaP, DU145, 22Rv1 were obtained from American Type Cell Collection (Manassas, VA, USA). LAPC-4 cell line was kindly provided by Dr. Charles Sawyers (University of California at Los Angeles, Los Angeles, CA, USA), VCaP and DuCaP by Dr. Jack Schalken (Radboud University Nijmegen Medical Center, Nijmegen, the Netherlands), and EP156T by Dr. Varda Rotter (Weizmann Institute of Science, Rehovot, Israel). All cell lines were cultured under recommended conditions. 22 previously established LuCaP-series xenografts were provided by R.L.V. These xenografts have been derived from primary and metastatic human prostate cancer and are maintained in vivo [1].

Deriving gene expression estimates for protein-coding genes

RNA-seq reads were aligned against RefSeq 38 human transcript sequences using Bowtie version 2.0.0-beta6 [2]. Expression values were normalized across all 53 samples using median-of-ratios normalization. Read counts for a given gene were divided by the total length of the gene's exons (in kilobases) to correct for gene size bias. For some genes, we observed a strong and systematic expression bias associated with the two different RNA isolation methods that we used (Trizol and Qiagen AllPrep

kits). To correct for this bias, we calculated for each gene an expression ratio from BPH samples from which we had extracted the RNA using both Trizol and Qiagen. The ratio was calculated by dividing the median expression in the Trizol group by the median expression in the Qiagen group. We then used an unpaired t-test to look for genes differentially expressed between Trizol and Qiagen treated samples. Genes with a p-value less than 0.0001 were considered sensitive to the RNA isolation method, and their expression levels were corrected by dividing the expression of Trizol samples with the Trizol/Qiagen expression ratio.

Calculation of small RNA expression levels

After quality filtering of smallRNA-seq reads, 3' adapters were trimmed out. Trimmed reads were then aligned against sequences annotated in miRBase 18 [3] and 3' isoforms of the sequences. Expression values between samples were normalized using median-of-ratios normalization. Only RNAs with a median expression greater or equal to 15 reads were used to calculate ratios for normalization.

Clustering of samples based on gene expression profiles

For each gene, log2 expression ratios were calculated relative to the gene's median expression. To reduce the effect of noise, lowly expressed genes (<1000 reads) were omitted similarly as stably expressed genes (standard deviation of logratios < 1.0). Hierarchical clustering across columns and rows was performed using the L1 distance metric ("city block distance").

Identification of differentially expressed genes

A gene was considered differentially expressed between two sample groups (e.g. PC vs BPH) if all of the following conditions were met:

- Absolute difference between medians of length-normalized read counts in the two groups was above 200
- Absolute log2-ratio between medians of length-normalized read counts in the two groups was above 1.0
- p-value from ranksum test comparing the equality of expression between the two groups was below 0.0001

A gene was considered to have outlier expression between two sample groups (e.g. PC vs BPH) if all of the following conditions were met:

- Absolute difference between the highest length-normalized read count in one group and the lowest length-normalized read count in the other group was above 2000
- Absolute log2-ratio between the highest length-normalized read count in one group and the lowest length-normalized read count in the other group was above 2.0
- Gene was not considered differentially expressed

The reason for comparing absolute read count differences between sample groups (rather than just logratios and p-values) was to exclude differentially expressed genes where the overall expression in both sample groups was so low as to render any biological significance unlikely.

Transcriptome assembly

RNA-seq reads were aligned to the UCSC human genome hg19 using TopHat version 1.4.0 for 64bit Linux x86 [4] with default parameters. TopHat uses Bowtie [5] to perform a gapped alignment of reads, which enables the discovery of novel splice junctions. Aligned RNA-seq reads were combined into four pools of 1-1.5 billion reads according to sample histology. Transcriptomes were then built individually for each pool using the Cufflinks [6] version 1.3.0 for 64-bit Linux x86 with NCBI 37.2/hg19 genome build and default parameters. Illumina iGenome transcriptome annotations were used as a reference to guide transcript assembly. Cufflinks assembles mapped reads into transfrags based on their genomic coordinates. Manual inspection of the transfrags suggested that despite pooling samples to reduce transcriptional noise, most of the transfrags actually resembled parts of poorly assembled transcripts, or sequencing artifacts more than complete transcripts. This was concluded based on the distribution of transfrags' proximity with each other, short median length (less than 1kbp), high frequency of transfrags without splice junctions, and low occurrence across the samples. To address this problem, a merged transcriptome was formed by combining all overlapping and proximal (within 250bp) transfrags. Classifying the transfrags to either previously annotated or unannotated transcripts based on reference annotations (UCSC hg19 (http://hgdownload.cse.ucsc.edu/downloads.html), NCBI build 37.2 (http://www.ncbi.nlm.nih.gov/genome/guide/human/), Ensembl GRCh37 (http://www.ensembl.org/info/website/archives/index.html) and Gencode version 12e (http://www.gencodegenes.org/)) resulted in 32,744 novel intragenic and 66,376 novel intergenic transfrags. Intragenic transcripts were defined as those that overlapped with previously annotated introns, but did not match with previously annotated exons.

Deriving a focused set of novel transcripts

Even after merging the proximal transfrags and partial assemblies, the set of intergenic transcripts still contained obvious sequencing artifacts and/or transcriptional noise, a number of filtering criteria for absolute expression, differential expression, and previous annotations were applied to identify transcripts that were novel and likely the most relevant to prostate cancers. First, we counted the reads that were mapped to the genomic loci of the unannotated transcripts for each sample. The statistically significant differences in read counts between the sample groups (BPH, PC, CRPC), excluding the two AR-negative samples, were computed using version 1.2.1 of DESeq algorithm [7]. Additionally, we computed p-values for differential expression of transcripts based on Mann-Whitney U-test to account for the bias from outlier expression profiles in DEseq. That is, we only wanted to include those transcripts that were also systematically differentially expressed and not only through outlier expression. Only transcripts with adjusted p-value of less than 0.001 from both DESeq and U-test, and normalized read count exceeding 500 reads, were included in the subsequent analyses. We also filtered

out previously identified Prostate Cancer Associated Transcripts (PCATs) [8], but not the transcripts that have been annotated in Noncode v3.0 database [9] since not all of these have been well characterized, and thus may be important in prostate cancer. Despite merging and filtering, many of the transcripts were still short in length, arose from close adjacent genomic loci, and shared a strikingly similar expression profiles. Therefore, we concluded that they were likely ill-assembled transfrags, i.e. parts of the same transcript, or were at least produced by the same transcriptional mechanism. To address this issue, we manually curated the transcripts in order to merge the remaining transfrags into transcripts, and to identify likely false positives (e.g. unannotated 3' UTRs, or simple DNA repeats) that had not yet been filtered out. Exon-structures, putative isoforms, and strandedness were manually inferred based on the recurrence of junctions in the paired-end read data coinciding with canonical intron splice site motifs. We were able to make these predictions for about half of the remaining transcripts, the rest may either be functional single-exon transcripts or parts of ambiguously expressed genomic loci, a phenomenon described earlier [8]. We chose not to filter the one-exon structures out. Taken together, we ended up with 128 novel prostate cancer specific transcripts. For some transcripts multiple isoforms was observed which were also categorized whenever possible, leading to a list of 145 distinct molecular entities (**Supplementary Table 2**).

General properties of the transcripts were in line with previous studies of novel lncRNAs [10]. Using RepeatMasker (http://repeatmasker.org) annotations of low complexity DNA sequences and interspersed repeats, we found that the transcripts often coincided with repetitive DNA elements. We did not use this as a filtering criterion since repeat-associated lncRNAs have been shown to be functionally relevant in the past [8]. Also, as reported earlier, the transcript expression often correlated with the expression of an adjacent gene [10]. Some of the expression correlated transcripts may represent parts of unannotated UTRs, but others may be regulators of these genes, as suggested by previous studies [10], and therefore we did not filter our transcripts based on this criterion. Most of the transcripts with exon structures had two or three exons, a bias that has been observed before [10], and did not show coding potential based on lengths of open reading frames.

Taking account the inferred exon structures, read counts were re-computed for the curated transcripts. The read counts were median-of-ratios (MOR) normalized with the transcripts from RefSeq 37 human transcriptome. Reads-per-kilobase (RPK) expression levels were computed for the transcripts (**Supplementary Table 2**). We inspected the expression profile of each transcript, and found that majority of the transcripts were exclusively expressed in either CRPC or in both CRPC and PCA. BPH specific transcripts were almost nonexistent. Many transcripts were expressed at a high level only in one or few samples, which may explain the reason why they have remained unannotated thus far.

Nomenclature of novel transcripts

To prevent confusion with earlier studies describing novel prostate cancer specific transcripts, we labeled our transcripts Prostate Cancer Associated Transcripts from Tampere cohort (TPCAT). Rather than using arbitrary numbering, we generated a unique identifier for each transcript based on the chromosomal coordinates of the locus (in kilobases). For example TPCAT-10-36067 is located in

chromosome 10 around genomic coordinate 36067kb. However, after submission to HGNC they will be named PCATs followed by a number, such as PCAT5 for TPCAT-10-36067.

Correlation of novel transcript expression with TF expression

Spearman correlations between transcripts expression and eight transcription factors (ERG, AR, FOXA1, EZH2, HDAC1, HDAC2, HDAC3, RUNX2) [11] were computed to look for transcripts whose expression could be induced by these transcription factors.

ChIP-sequencing data from untreated or vehicle treated cell lines 22RV1, LNCaP, and VCAP, as well as from doxycycline treated C4-2B cell line, and metastatic prostate cancer tissue samples were used to confirm transcription factor binding events in the promoters of TPCATs. Raw data from two published studies [12, 13] was downloaded from Gene Expression Omnibus (GEO) and aligned to GRCh37 using Bowtie (version 0.12.8.) and the non-default parameters –v2 and –m20. Transcription factor binding sites were inferred with MACS peak detection algorithm [14] with default parameters unless otherwise stated (**Supplementary Table 4**). Processed data from two additional studies [11, 15] were downloaded from GEO and converted from hg18 to hg19 using LiftOver [16] with default parameters.

In addition to the transcription factors, antibodies used in these experiments were specific for histone modifications H3K4me1-3, H3K9me3, H3K36me3, H3K37me3, and RNA polymerase II in VCaP cells under three different conditions. Any MACS-inferred peak region (center coordinate) located at maximum 1kbp upstream or 500bp downstream from the inferred TSSs was considered a binding event to the TPCATs proximal promoter. For TPCATs that we could not infer strandedness, we treated both ends of the transcript equally. All the TF binding data taken together, we identified 102 binding events to the proximal promoter regions of the TPCATs (**Supplementary Table 4**).

Correlation of novel transcript expression with DNA methylation levels and copy number changes

To investigate whether changes in DNA methylation or copy number bring about the expression of the novel transcripts in the samples that express them, we integrated DNA-sequencing and MeDIP-sequencing data from the same samples with the RNA-seq data.

DNA-seq reads were aligned against GRCh37 using Bowtie version 2.0.0-beta6 (ref. 36). Aligned read counts were calculated within overlapping 500 bp windows along the whole genome. Coverage logratios were calculated within each window by comparing against read count averages from four BPH controls. To normalize logratios within copy neutral regions to zero, we applied a median filter of length 50 to the logratios in each sample, rendered logratio histograms for each chromosome, and took the median of the histogram modes. This value was then subtracted from all logratios for that sample. Frequently aberrant chromosomes 8, 22, X, and Y were not included in calculating the median of modes. Normalized logratios were segmented using circular binary segmentation. Coverage logratios for individual genes were calculated by taking the median logratio over all intragenic windows. If a gene's length was shorter than 20 kb, the median window was extended on both sides so as to reach a length of 20 kb. Logratios were converted into copy number changes using the formula (ploidy * 2^logratio - ploidy), where ploidy was based on the chromosome in which the gene or genomic region was located.

Copy number changes were further multiplied by (1 / 0.7) to correct for the estimated 70% tumor sample purity in our samples. A gene was considered to be amplified or deleted if the corrected copy number change had an absolute value above 0.5.

MeDIP-seq reads were aligned against GRCh37 using Bowtie version 2.0.0-beta7 (ref.36). All read pairs that aligned within 5kb of one another were extended into long fragments, and fragment counts were tallied in 50 bp windows across the entire genome. Fragment density within each bin was normalized by the total amount of aligned fragments in each sample. To correct for copy number bias, the genome was divided into 500 kb windows, and fragment counts were calculated within each window. For each window, the median read count in BPH samples was calculated and used as a reference. For each sample, the read count within each window was multiplied so as to match the reference read count.

For the purposes of identifying differentially methylated regions (DMR), the genome was divided into a grid of overlapping 550 bp windows. A 550 bp genomic window was considered differentially methylated between BPH and PC or PC and CRPC if: i) The median MeDIP-seq fragment count between sample groups was changed 2-fold or more; ii) The difference in average MeDIP-seq fragment counts between the two sample groups was at least 15 fragments; iii) Ranksum test of MeDIP-seq fragment counts between sample groups produced a p-value less than 0.01; and iv) No other DMR with a stronger absolute logratio was located within 1000 bp. DMRs were classified TPCAT promoter associated if they were located within or at most 250 kb from either end of a TPCAT.

We computed Spearman correlations between transcript expression values and the copy number of the locus, and expression values and methylation values of nearby differentially methylated regions. In addition, we tested for differential expression between samples that had copy number aberrations at the locus versus samples that had normal copy number for each TPCAT using t-test. We required a significant correlation between the expression and copy number, and significantly differential expression between samples with normal copy number and samples with copy number aberration. Similar requirements were applied to methylation. None of the TPCATs were found to be significant taking account both criteria indicating that the expression differences of TPCATs were not explained by these factors. This suggesting that, at least in general, TPCATs are transcriptionally regulated and that their expression does not arise due to genetic alterations or changes in DNA methylation (**Supplementary Table 4**).

Processing of additional RNA-seq datasets

We studied the expression of the TPCATs using various other RNA-sequencing datasets. RNA-seq data from 24 normal human tissues and 21 normal prostate and prostate cancer cell lines was acquired from two published studies [8, 17]. In addition, RNA-sequencing data of Cold Spring Harbor Lab's PolyA-selected and non-PolyA-selected H1 human embryonic stem cells was downloaded from ENCODE [18]. Additionally, we acquired two publicly available prostate cancer RNA-sequencing datasets [8, 19] with a total of 64 samples. These data were re-aligned to hg19 with TopHat as described above, and the read counts for all the 145 novel transcripts or isoforms were computed (**Supplementary Table 2**).

Validating TSSs of transcripts expressed in LNCaP cells using GRO-seq

Three Global Run-On Sequencing (GRO-seq) datasets from DHT- or vehicle-treated, mock, or FOXA1 siRNA-transfected LNCaP prostate cancer cell line [20] were acquired from Gene Expression Omnibus (GEO). The data were realigned to hg19 with Bowtie2 [2] and default parameters. Forward and reverse reads were separately combined to gain greater coverage. Putative promoters for each TPCAT that was expressed in LNCaP cells were analyzed. Many TPCATs (such as TPCAT-15-21970 in **Supplementary Figure 5**) showed the distinctive pattern of divergent peaks of forward and reverse reads for transcription start site close to the one that we had inferred from our RNA-seq data.

Microarray expression analyses of siRNA knockdown cell lines

To study the effect of PCAT5 on the protein coding gene expression the siRNA 1 and siRNA 2 (50 nM) and scrambled control siRNA (50nM) were transiently transfected to PC-3 cell line expressing high level of PCAT5 transcript and the RNA was collected and purified as described above. 500 ng of total-RNAs of siRNA 1 and 2 as well as scrambled control treated cells was labeled with either Cy5 or Cy3 with correspond dye-swap controls using two-color gene expression system and hybridized on the Whole Genome Human 4x44K microarray chips (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. The microarray images were scanned with the Agilent microarray scanner and analyzed using Agilent Feature Extraction Software with default setting.

After siRNA-mediated PCAT5 knockdown experiments on PC-3 cell line with two different siRNAs, we measured global gene expression patterns with Agilent gene expression arrays. The experiments were done in duplicates with dye swapping. Log2 ratios were computed for all the genes over normal reference. Both siRNAs induced a comparable effect in the expression rates, and therefore mean log ratios were taken of all four experiments to reduce the effect of microarray measurement noise. Genes with mean log ratios of expression outside two standard deviations were considered differentially expressed in these experiments (**Supplementary Table 5**).

Pathway analysis

We acquired gene set data from five databases: Biocarta (http://www.biocarta.com/genes/index.asp, 4/19/2011), WikiPathways [21] (http://wikipathways.org, 1/3/2011), Pathway Commons [22] (http://www.pathwaycommons.org, 11/30/2010), Gene Ontology Biological Processes [23] v3.0 (http://www.geneontology.org, 1/3/2011), Kyoto Encyclopedia of Genes and Genomes [24] (http://www.genome.jp/kegg/pathway.html, 6/14/2010). Together, these databases contained over 2500 unique signaling or metabolic pathways or genes annotated to the same biological process. The gene sets were combined into a mutually compatible form by re-annotating the gene identifiers to a common namespace (HUGO nomenclature). Gene sets with less than 10 or more than 1000 constituent genes were filtered out resulting in 1994 pathways or processes.

Enriched pathways in differentially expressed genes were computed using the gene lists in **Supplementary Table 3** omitting genes that have been implicated in ischemic prostate tissues [25]. We used standard hypergeometric test (with p-value threshold of 0.05) with minimum of 10 genes in each

enriched pathway to concentrate on the more general level pathways. Processes were computed for genes that were differentially expressed between PC and BPH, and CRPC and PC (**Supplementary Table 3**). Further, using the lists of differentially expressed genes from siRNA microarray expression experiments, we computed the enriched gene sets in cells with downregulated PCAT5 expression (**Supplementary Table 5**).

List of primers and siRNAs used in validation

All primers used for sequence, expression and DNA-methylation validations by traditional PCR, qPCR, RT-PCR, qRT-PCR, and Sanger sequencing as well as siRNA sequences are listed in **Supplementary Table 5**.

PCR analysis

For all mRNA and TPCAT expression analysis, cDNA synthesis was performed from total RNA (1-4ug) using SuperScript(TM) III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) or AMV (Finnzymes, Espoo, Finland) and random hexamer primers (Thermo Scientific, Waltham, MA, USA) according to manufacturers' instructions and diluted 1-10-1:20 to nuclease free water. The expression of all targets genes was calculated relative to TATA-box binding protein (TBP) reference gene. Quantitative mRNA expression was studied by quantitative Real Time PCR (qRT-PCR) using Bio Rad CFX96 Real Time System. The final reaction mixture (22 µl) contained 2 µl cDNA, 0.125 µl reverse and forward primers and 11 µl 2X SYBR Green qPCR Master Mix and RNAase free ddH2O. In order to detect DNA contamination NTC (No template control) was used during the reaction. 95 °C for 20 seconds followed by 55 cycles of 95°C for 10 seconds, 56-60°C for 10 seconds, 72°C for 8-15 seconds, and finally the melting curve analysis with 1.5% agarose gel electrophoresis of the products were used to ensure that right size product was amplified in the reaction.

Sanger sequencing

Target amplicons were purified using QIAquick PCR purification columns (Qiagen Inc, Valencia, CA, USA) after which the sequencing was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and the ABI PRISM® 3100 sequencer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions.

Small interfering RNA knockdown of ERG and PCAT5

Small interfering RNA knockdown of ERG, ETV4, and PCAT5 ERG was knocked down with mixture of two commercial siRNA-molecules (ERG-1 3504332/3-F and ERG-2 3504334/5-F, Proligo® Sigma-Aldrich Corp. St. Louis, MO, USA) by reverse transfection with 25nM of ERG siRNA-mixture or 50 nM of scrambled control siRNA and Lipofectamine®RNAiMAX Transfection Reagent (Invitrogen/ Thermo Fisher Scientific, Waltham, Massachusetts, USA) on 12-well plate according to manufacturer's instructions. ETV4 was knocked down with predesigned siRNA-molecules (HSC.RNAI.N001986.12.1_2nm, Integrated DNA Technologies, Inc., Coralville, Iowa, USA) by reverse transfection with 25nM of ETV4 siRNA or scrambled control siRNA and Lipofectamine®RNAiMAX Transfection Reagent (Invitrogen/ Thermo Fisher Scientific,

Waltham, Massachusetts, USA) on 12-well plate according to manufacturer's instructions. PCAT5 was knocked down using three different siRNAs self-designed to target exon 3 of the transcript (Integrated DNA Technologies, Inc., Coralville, Iowa, USA) (Supplementary Table 5). siRNAs were designed with the IDTRNAi design tool (https://eu.idtdna.com/Scitools/Applications/RNAi/RNAi.aspx). Cells were seeded into 12well plates as triplicates and were transfected the following day with 50 nM PCAT5targeting siRNAs and the scrambled control siRNA. INTERFERin (PolyPlus Transfection, Strasbourg, France) and OptiMEM were used for cell transfection.

Growth and cell viability assays

For growth-curve analysis of transfected (Supplementary Table 5) PC-3 and control 22Rv1 cell lines were used. The wells were scanned each day using the Surveyor Software (Objective Imaging Ltd.) with camera (Imaging Inc., Canada) attached to the Olympus IX71 (Olympus, Tokyo, Japan) microscope and the area of the attached cells in each well were counted by analysis with ImageJ Software (Wayne Rasband, National Institutes of Health, Bethesda, MD) and divided by the mean area of day-1 for each following day. For cell viability assay 150 000 DuCaP cells were reverse transfected with 25nM of PCAT5-siRNA 2 and 3 each (Supplementary Table 5) or 50 nM of scrambled control siRNA using Lipofectamine®RNAiMAX Transfection Reagent (Invitrogen/ Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the manufacturer's instructions. Cells were equally seeded on 24-well plate as four replicates on day1 and alamarBlue® (Invitrogen/ Thermo Fisher Scientific) reagent was used in Day 5 to quantitatively measure the relative cell viability of the all transfected cells as well as non-transfected control cells by adding 60uL of alamarBlue® into the cell medium after which the plate was incubated 3h in cell incubator and the endpoint absorbance of all wells were measured with the EnVision, 2104 Multilabel Reader (Wallac, PerkinElmer Inc., Waltham, Massachusetts, USA) at 595 nm.

Cell invasion assay

Effects of PCAT5 knockdown on PC-3 cell invasion were evaluated in BioCoat Matrigel Invasion chambers (BD Biosciences, Bedford, MA, USA) coated with basement membrane matrix. Matrigel was rehydrated in growth medium for 2 h at 37 °C and 5% CO2. After transfection with gene specific siRNAs or control siRNAs (**Supplementary Table 5**) for two days, 10,000 PC-3 cells were harvested and resuspended in 1% FBS, and placed in the upper chamber of the transwell. The tumor cells in the upper chamber were allowed to migrate into the lower chamber containing 10% FBS and 5 µg/ml fibronectin containing medium (750 µl) by incubating for 22 h at 37 °C in a 5% CO2 atmosphere. Cells in the top well of the upper chamber were removed by wiping them off the top membrane with cotton swabs. The membranes were then fixed with 3.7% formaldehyde for 2 min, and permeabilized with 100% methanol for 10 min, and stained with 1% toluidine for 15 min at room temperature for visualization of cells. The cells that invaded to the lower surface were photographed under a microscope and counted.

Colony formation assay

PC-3 cells transfected with the gene specific-siRNAs targeting PCAT5 and scrambled, non-targeting control siRNAs (**Supplementary Table 5**) were grown on 6-well plates in duplicate (5000 cells/well).

Briefly, base agar containing 1% agar dissolved in Ham's F-12 mixed with 20% FBS, 2% L-glutamine and 2% penicillin-streptomycin and 1 ml mixture was transferred into the well. Then the top layer containing 0.7% agar including 10% FBS, 1% L-glutamine and 1% penicillin-streptomycin was prepared. Top agar was mixed with suspended cell line and transferred over base agar. Base and top agar were covered with normal growth medium and incubated at 37°C and 5% CO2 for 15 days. After incubation, colonies were fixed with 3.7% formaldehyde and stained with 0.1% toluidine blue. Excess dye was removed by washing with 10 mM phosphate buffer (pH 7.4).

Wound healing assay

The effects of PCAT5 knockdown on PC-3 cell mobility were assessed using a wound healing assay. After transfection with 50nM gene specific siRNAs or control siRNAs (**Supplementary Table 5**) for 2 days, PC-3 cells were cultured in a 12-well plate until confluent. The cell layer was carefully scratched using a sterile tip. After incubation with serum free medium for 18 h, the cells were photographed.

Flow cytometric analysis of apoptosis

PC-3 cells were transfected with INTERFERin™ (Polyplus Transfection, Strasbourg, France) and 50 mM siRNA (**Supplementary Table 5**) for 28 hours, harvested by trypsinization, washed with phosphate-buffered saline (PBS) and resuspended to annexin-binding buffer (10 mM HEPES, 140 mM NaCl, and 2.5 mM CaCl2, pH 7.4). Cells were stained with R-phycoerythrin-labelled Annexin V (Molecular Probes, Invitrogen, Carlsbad, CA, USA) and analyzed with Accuri C6 Flow cytometer.

Data access

The European Genome-phenome archive database accession number for the high throughput sequencing data reported in this paper is EGAS00001000526. Microarray data has been deposited in Gene Expression Omnibus under accession number GSE48211. The complete *PCAT5* sequence has been deposited in Genbank under accession number KF154780.1.

Supplementary figure legends

Supplementary Figure 1. Cell cycle regulation. (a) Pathway diagram. Each box represents a gene, the left half shows changes in untreated prostate cancer (PC), and the right half in castration resistant prostate cancer (CRPC). Percentages indicate the fraction of samples with RNA-level overexpression (red) or underexpression (blue) relative to BPH controls. (b) Matrix showing which samples exhibit overexpression (red) or underexpression (blue) in the sequencing cohort.

Supplementary Figure 2. Androgen biosynthesis and androgen receptor (AR) signaling. (a) Pathway diagram. Each box represents a gene, the left half shows changes in untreated prostate cancer (PC), and the right half in castration resistant prostate cancer (CRPC). Percentages indicate the fraction of samples

with RNA-level overexpression (red) or underexpression (blue) relative to BPH controls. (b) Matrix showing which samples exhibit overexpression (red) or underexpression (blue) in the sequencing cohort.

Supplementary figure 3. Validation of PCAT5 expression pattern by RT-PCR. (**a**) Expression pattern of PCAT5 was validated with qRT-PCR in 76 tumor samples from validation cohort. ERG status for each sample was investigated by IHC staining. Asterisk on top of PCAT5-expression level indicating bar denotes ERG positive samples, minus ERG negative samples, and m stands for missing ERG status data. (**b-c**) qRT-PCR shows similar consistency of expression patterns of PCAT5 and ERG in LuCaP xenograft series (Spearman correlation 0,78). The cell line with the highest expression was PC-3, in agreement with RNA-seq data. (**d-e**) Exon-exon junctions were identified with Sanger sequencing from samples PC_15760, PC_14670, and PC_17163. Sequence validation is shown for PC_15760.

Supplementary figure 4. Normalized read counts of four TPCATs are plotted in our BPH, PC, and CRPC samples, and complemented with prostate cancer cell lines, normal tissues, human embryonic stem cells, and prostate cancer samples from three additional data sets. Read counts were normalized using median-of-ratios normalization. (**a**) TPCAT-10-84917 is expressed in a subset of PC and CRPC tumors in our dataset. The transcript is not present in any cell lines, normal tissues, stem cells, or the other two PC datasets. (**b**) TPCAT-10-84941v3 is expressed in a subset of PC and CRPC samples as well as in the LNCaP-CDS-parent cell line. (**c**) TPCAT-11-23310 is highly expressed in two CRPC samples and the LNCaP-CDS-parent cell line. (**d**) TPCAT-15-21970v1 is expressed in a subset of PC and CRPC samples as well as brain and adrenal tissues. (**e**) RT-PCR validation for TPCAT-10-36067 in three clinical samples, TPCAT-10-84917 in three clinical samples, TPCAT-10-84941 in three clinical samples, TPCAT-15-21970 in one clinical sample and two PC cell lines (VCaP and LNCaP), TPCAT-11-23310 in four clinical samples.

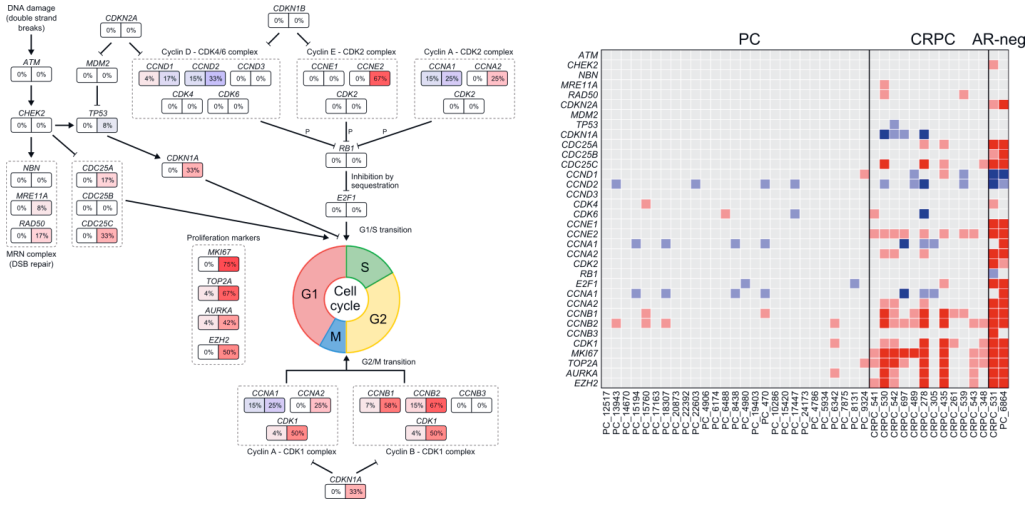
Supplementary Figure 5. Inferred exon structures for four novel transcripts that were validated with RT-PCR. (**a**) Expression profile of TPCAT-10-84919 is plotted for our pooled BPH, PC, and CRPC samples, and supplemented with profiles of pooled samples from two published prostate cancer cohorts (9, 30). Read junction indicated a three-exon structure. (**b**) Read data for TPCAT-10-84941 suggested a three-exon structure and three alternative splice sites in the last exon boundary. Different splice variants are denoted by v1-3. (**c**) Data for TPCAT-15-21979 indicated two variants, v1 with three exons, and v2 with two exons due to a skipped middle exon. (**d**) The expression profile of TPCAT-11-23310 is plotted for the two CRPC samples that high express the transcript. Read junctions indicated a two-exon structure.

Supplementary Figure 6. Open chromatin markers and ERG binding at the locus of PCAT5. Topmost track shows the RNA-seq data coverage plot for sample PC_14670. The transcript structure is shown on the bottom, with red boxes indicating exons. Binding of ERG was observed at the promoter of PCAT5 only in VCaP cells which expressed the transcript, and not in LNCaP or healthy prostate tissue which did not express PCAT5. Open chromatin markers (methylation of H3K4) were enriched in VCaP cells in comparison to LNCaP cells. No significant differences were observed in the promoter in other histone modifications between the cell lines.

Supplementary Figure 7. (**a**) ETV4 and (**b**) PCAT5 expression after 25nM ETV4-siRNA or scrambled CTRL-siRNA treatment in PC-3 cells. Error bars, s.e.m.; ***p<0.001, unpaired two-tailed t-test (n=2 or 3).

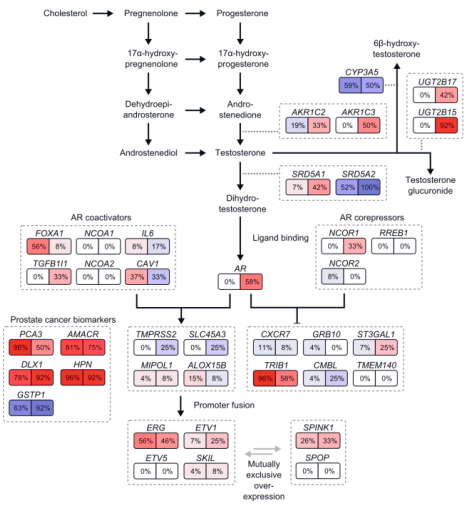
Supplementary Figure 8. Transfection with PCAT5-specific siRNAs did not affect the growth of PCAT5 negative prostate cancer cell line 22Rv1.

Supplementary Figure 9. Relative cell viability of PCAT5-siRNA transfected ERG-positive DuCaP cells. Pool of 50nM PCAT5-siRNA2/3 were used or negative controls. NT = Non-transfected control cells. CTRL = scrambled siRNA transfected cells. Error bars, s.e.m.; **P<0.0037, unpaired two-tailed t-test (n=4).


DNA damage (double strand breaks)
ATM
CHEK2
NBN
MRE11A
RAD50
MRN complex (DSB repair)
CDKN2A
MDM2
TP53
CDC25A
CDC25B
CDC25C
CDKN1A
CDKN1B
Cyclin D - CDK4/6 complex
CCND1
CCND2
CCND3
CDK4
CDK6
Cyclin E - CDK2 complex
CCNE1
CCNE2
CDK2
Cyclin A - CDK2 complex
CCNA1
CCNA2
CDK2
RB1
Inhibition by sequestration
E2F1
G1/S transition
Proliferation markers
MKI67
TOP2A
AURKA
EZH2
Cell cycle
G1
S
G2
M
G2/M transition
Cyclin A - CDK1 complex
Cyclin B - CDK1 complex
CCNA1
CCNA2
CDK1
CCNB1
CCNB2
CCNB3
CDK1
CDKN1A

PC
CRPC
AR-neg
ATM
CHEK2
NBN
MRE11A
RAD50
CDKN2A
MDM2
TP53
CDKN1A
CDC25A
CDC25B
CDC25C
CCND1
CCND2
CCND3
CDK4
CDK6
CCNE1
CCNE2
CCNA1
CCNA2
CDK2
RB1
E2F1
CCNA1
CCNA2
CCNB1
CCNB2
CCNB3
CDK1
MKI67
TOP2A
AURKA
EZH2

Figure S1


Cholesterol
Pregnenolone
Progesterone
17α-hydroxy-pregnenolone
17α-hydroxy-progesterone
6β-hydroxy-testosterone
CYP3A5
59%
50%
UGT2B17
0%
42%
Dehydroepi-androsterone
Andro-stenedione
AKR1C2
19%
33%
AKR1C3
0%
50%
UGT2B15
0%
92%
Androstenediol
Testosterone
Testosterone glucuronide
SRD5A1
7%
42%
SRD5A2
52%
100%
Dihydro-testosterone
AR coactivators
FOXA1
56%
8%
NCOA1
0%
0%
IL6
8%
17%
TGFB1I1
0%
33%
NCOA2
0%
0%
CAV1
37%
33%
Ligand binding
AR corepressors
NCOR1
0%
33%
RREB1
0%
0%
NCOR2
8%
0%
AR
0%
58%
Prostate cancer biomarkers
PCA3
90%
50%
AMACR
81%
75%
DLX1
78%
92%
HPN
96%
92%
GSTP1
63%
92%
TMPRSS2
0%
25%
SLC45A3
0%
25%
MIPOL1
4%
8%
ALOX15B
15%
8%
CXCR7
11%
8%
GRB10
4%
0%
ST3GAL1
7%
25%
TRIB1
96%
58%
CMBL
4%
25%
TMEM140
0%
0%
Promoter fusion
ERG
56%
46%
ETV1
7%
25%
ETV5
0%
0%
SKIL
4%
8%
Mutually exclusive over-expression
SPINK1
26%
33%
SPOP
0%
0%

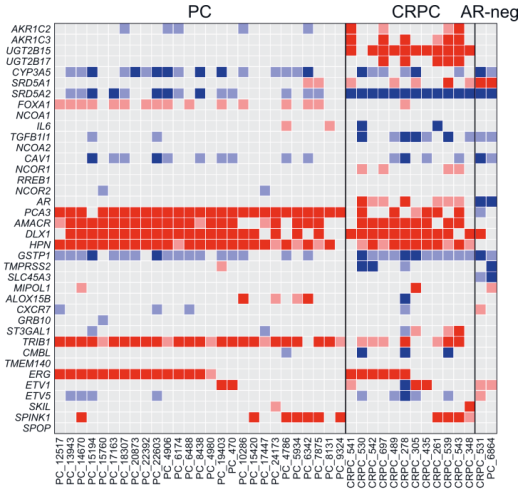

PC
CRPC
AR-neg
AKR1C2
AKR1C3
UGT2B15
UGT2B17
CYP3A5
SRD5A1
SRD5A2
FOXA1
NCOA1
IL6
TGFB1I1
NCOA2
CAV1
NCOR1
RREB1
NCOR2
AR
PCA3
AMACR
DLX1
HPN
GSTP1
TMPRSS2
SLC45A3
MIPOL1
ALOX15B
CXCR7
GRB10
ST3GAL1
TRIB1
CMBL
TMEM140
ERG
ETV1
ETV5
SKIL
SPINK1
SPOP
PC_12517
PC_13943
PC_14670
PC_15194
PC_15760
PC_17765
PC_18307
PC_20873
PC_22392
PC_22603
PC_4066
PC_6174
PC_6488
PC_8438
PC_8490
PC_19403
PC_470
PC_10288
PC_11447
PC_24173
PC_4786
PC_5934
PC_6542
PC_7875
PC_8131
PC_9324
CRPC_541
CRPC_530
CRPC_542
CRPC_697
CRPC_298
CRPC_278
CRPC_305
CRPC_435
CRPC_261
CRPC_258
CRPC_543
CRPC_348
CRPC_531
PC_6864

Figure S2

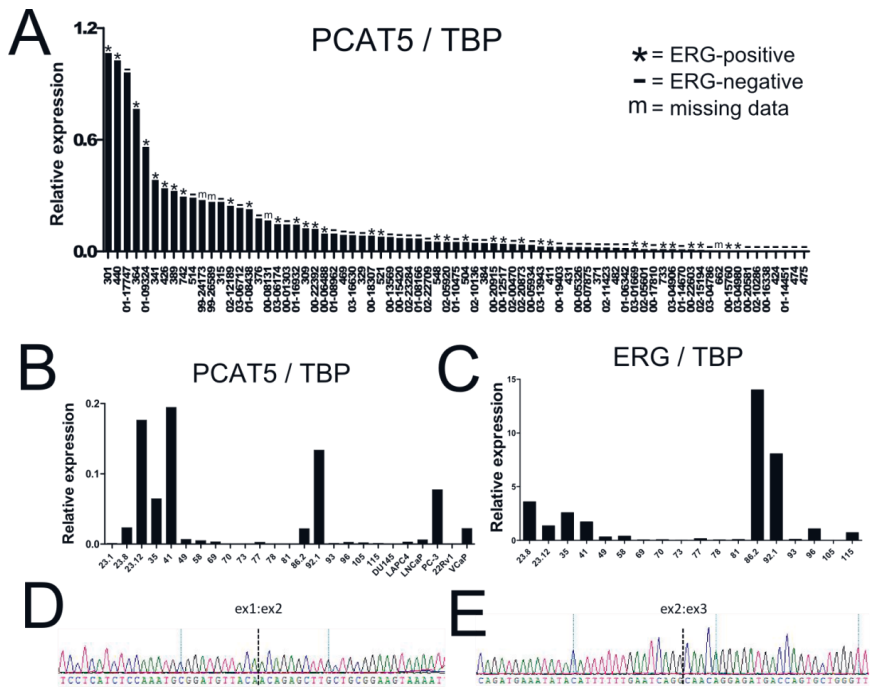

A
PCAT5 / TBP
* = ERG-positive
– = ERG-negative
m = missing data
Relative expression
1.2
0.6
0.0
B
PCAT5 / TBP
Relative expression
C
ERG / TBP
Relative expression
D
ex1:ex2
E
ex2:ex3

Figure S3

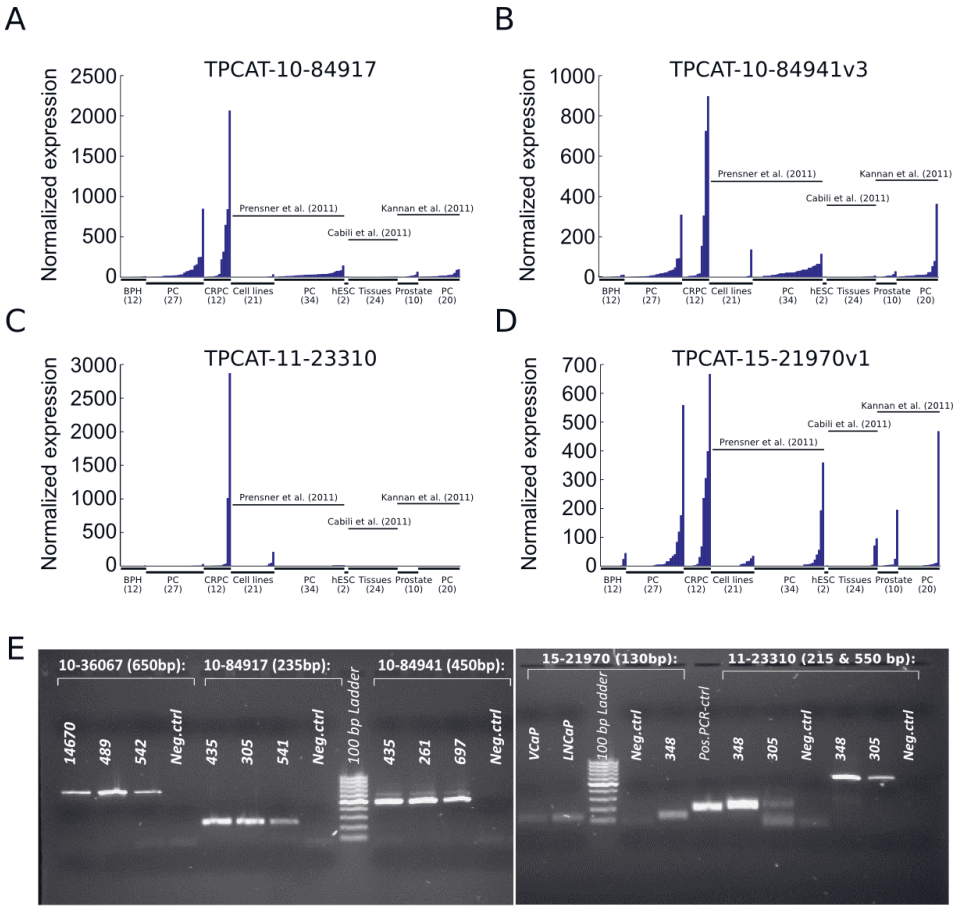

A
TPCAT-10-84917
Normalized expression
2500
2000
1500
1000
500
0
Prensner et al. (2011)
Kannan et al. (2011)
Cabili et al. (2011)
BPH (12)
PC (27)
CRPC (12)
Cell lines (21)
PC (34)
hESC (2)
Tissues (24)
Prostate (10)
PC (20)
B
TPCAT-10-84941v3
1000
800
600
400
200
C
TPCAT-11-23310
3000
D
TPCAT-15-21970v1
700
E
10-36067 (650bp):
10-84917 (235bp):
10-84941 (450bp):
14670
489
542
Neg.ctrl
435
305
541
100 bp Ladder
261
697
15-21970 (130bp):
11-23310 (215 & 550 bp):
VCaP
LNCaP
348
Pos.PCR-ctrl

Figure S4

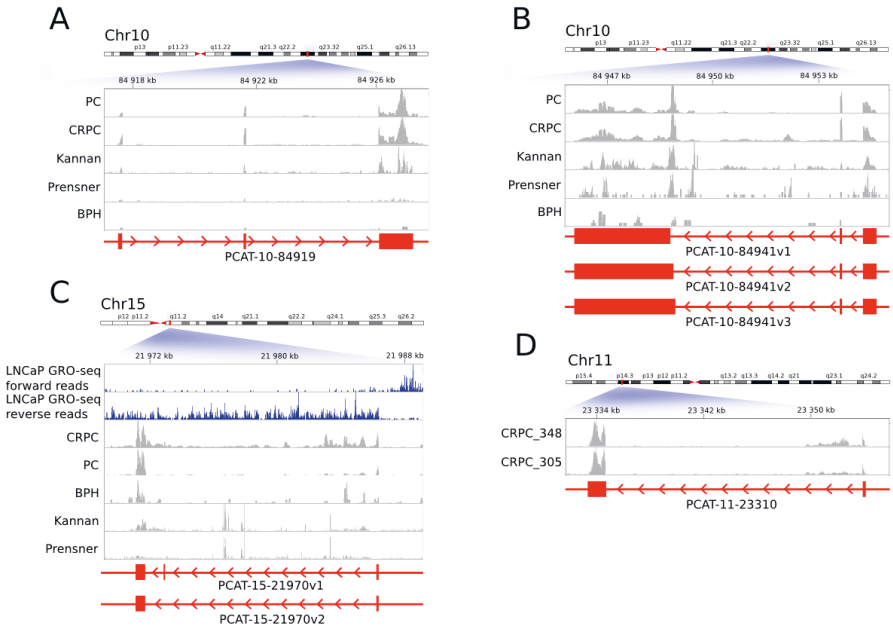

A
Chr10
84 918 kb
84 922 kb
84 926 kb
PC
CRPC
Kannan
Prensner
BPH
PCAT-10-84919
B
Chr10
84 947 kb
84 950 kb
84 953 kb
PC
CRPC
Kannan
Prensner
BPH
PCAT-10-84941v1
PCAT-10-84941v2
PCAT-10-84941v3
C
Chr15
21 972 kb
21 980 kb
21 988 kb
LNCaP GRO-seq forward reads
LNCaP GRO-seq reverse reads
CRPC
PC
BPH
Kannan
Prensner
PCAT-15-21970v1
PCAT-15-21970v2
D
Chr11
23 334 kb
23 342 kb
23 350 kb
CRPC_348
CRPC_305
PCAT-11-23310

Figure S5

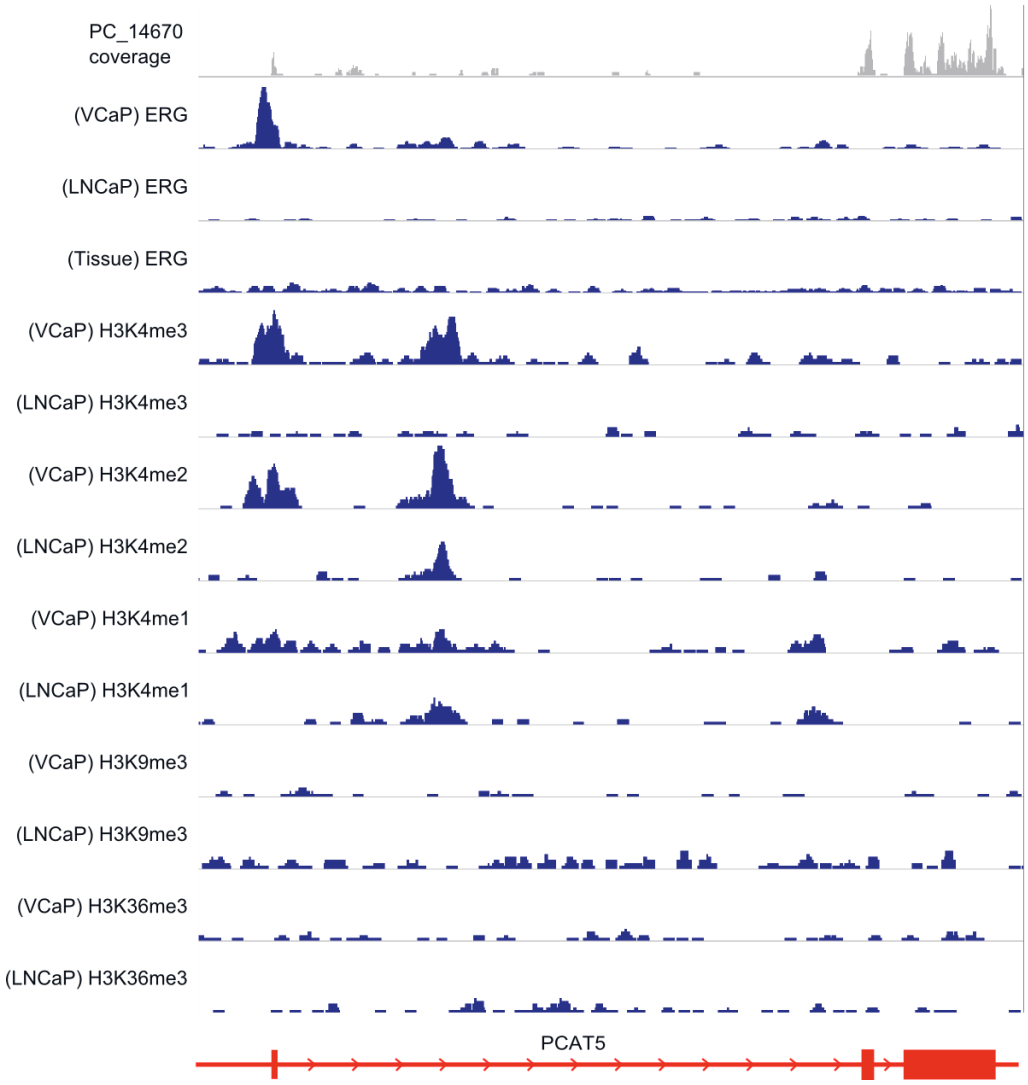

PC_14670
coverage
(VCaP) ERG
(LNCaP) ERG
(Tissue) ERG
(VCaP) H3K4me3
(LNCaP) H3K4me3
(VCaP) H3K4me2
(LNCaP) H3K4me2
(VCaP) H3K4me1
(LNCaP) H3K4me1
(VCaP) H3K9me3
(LNCaP) H3K9me3
(VCaP) H3K36me3
(LNCaP) H3K36me3
PCAT5

Figure S6

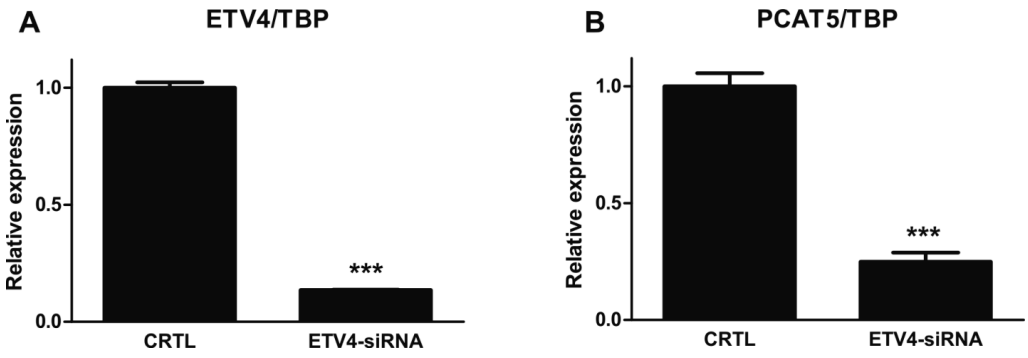

A
ETV4/TBP
Relative expression
1.0
0.5
0.0
***
CRTL
ETV4-siRNA
B
PCAT5/TBP
Relative expression
1.0
0.5
0.0
***
CRTL
ETV4-siRNA

Figure S7

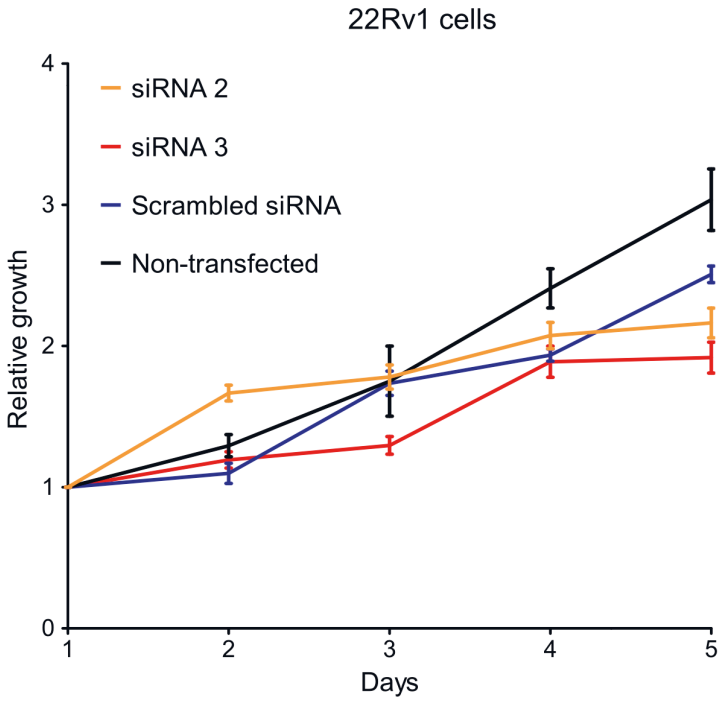

22Rv1 cells
siRNA 2
siRNA 3
Scrambled siRNA
Non-transfected
Relative growth
Days
0
1
2
3
4
1
2
3
4
5

Figure S8

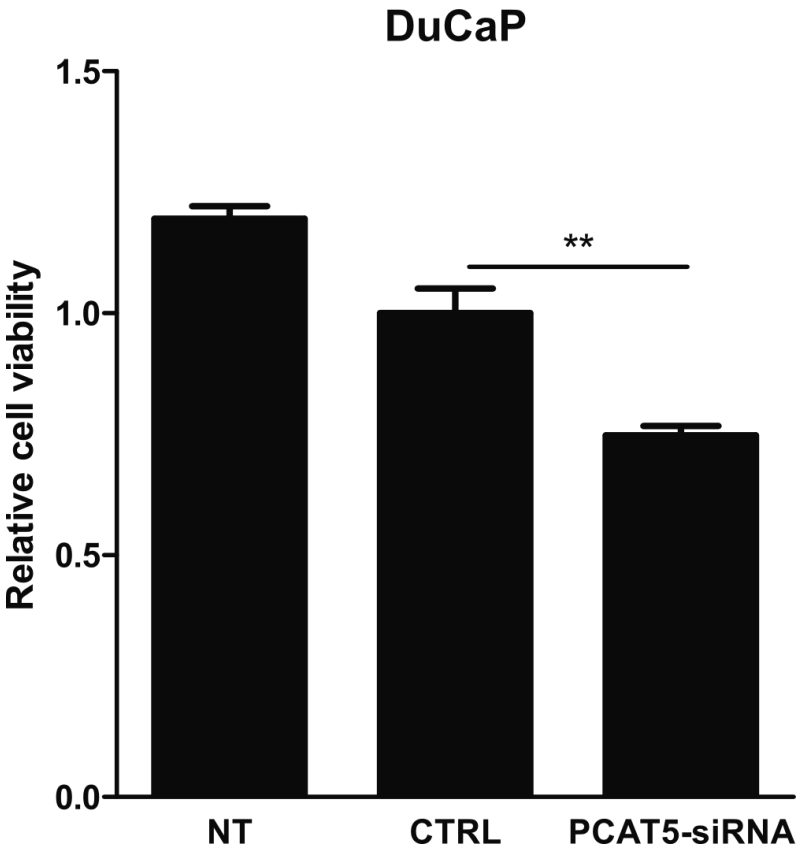

DuCaP
Relative cell viability
1.5
1.0
0.5
0.0
**
NT
CTRL
PCAT5-siRNA

Figure S9

Supplementary table legends

Supplementary Table 1. Clinical information for 28 freshly frozen untreated clinical specimens showing the age at diagnosis, PSA level, TNM classification, and Gleason scoring. Individual treatment modalities for the 13 castration resistant prostate tissues are also shown. Corresponding information is provided for samples in the validation cohort. Sequencing information. Read counts for each sample are given.

Supplementary Table 2. Manually inferred exon structures, read counts and median of ratios (MOR) normalized reads per kilobases (RPKs) for the novel transcripts in our samples, and read counts in cell lines, normal tissues, hESCs, Kannan cohort, and Prensner cohort.

Supplementary Table 3. Lists of genes that were differentially expressed between PC and BPH or CRPC and PC samples. Gene expression values in the table were normalized across samples using median-of-ratios normalization, and read counts were divided by gene length (in kilobases). Lists of microRNAs that were differentially expressed between PC and BPH or CRPC and PC samples. MicroRNA expression values in the table were normalized across samples using median-of-ratios normalization. Biological pathways enriched for genes that were differentially expressed between PC and BPH or CRPC and PC samples.

Supplementary Table 4. Correlations between expressions of TPCATs and TFs with available ChIP-sequencing data, copy number, and DNA methylation data. List of ChIP-sequencing datasets used to analyze DNA methylation changes in the proximity of transcription factor binding sites, and to discover putative transcriptional regulators of TPCATs. Observed TF binding events near TPCAT promoters in cell lines.

Supplementary Table 5. Overexpressed and underexpressed genes after siRNA knockdown of PCAT5. Pathway analysis is also shown. List of all primers used in experiments involving PCR-reactions and Sanger sequencing. The table also includes siRNA sequences that were used to knockdown PCAT5 *in vitro*.

Supplementary References

1. Corey E, Vessella RL. Xenograft models of human prostate cancer. In: Chung L, Isaacs W, Simons J, editors. Prostate Cancer: Biology, Genetics, and the New Therapeutics, 2nd Edition. Totowa: Humana Press; 2007. pp.3-31.
2. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9:357-359.
3. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research. 2014;42:D68-73.
4. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105-11.
5. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009;10:R25.
6. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology. 2010;28:511-5.
7. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology. 2010;11:R106.
8. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nature Biotechnology. 2011;29:742-749.
9. Bu D, Yu K, Sun S, Xie C, Skogerbø G, Miao R, Xiao H, Liao Q, Luo H, Zhao G, Zhao H, Liu Z, Liu C, Chen R, Zhao Y. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Research. 2012;40:D210-5.
10. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Research. 2012;22:1775-1789.
11. Chng KR, Chang CW, Tan SK, Yang C, Hong SZ, Sng NY, Cheung E. A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. The EMBO Journal. 2012;31:2810-2823.
12. Massie CE, Lynch A, Ramos-Montoya A, Boren J, Stark R, Fazli L, Warren A, Scott H, Madhu B, Sharma N, Bon H, Zecchini V, Smith DM, Denicola GM, Mathews N, Osborne M, Hadfield J, Macarthur S, Adryan B, Lyons SK, Brindle KM, Griffiths J, Gleave ME, Rennie PS, Neal DE, Mills IG. The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. The EMBO Journal. 2011;30:2719-33.
13. Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y, Cheng H, Laxman B, Vellaichamy A, Shankar S, Li Y, Dhanasekaran SM, Morey R, Barrette T, Lonigro RJ, Tomlins SA, Varambally S, Qin ZS, Chinnaiyan AM. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer Cell. 2010;17:443-454.
14. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biology. 2008;9:R137.
15. Little GH, Noushmehr H, Baniwal SK, Berman BP, Coetzee GA, Frenkel B. Genome-wide Runx2 occupancy in prostate cancer cells suggests a role in regulating secretion. Nucleic Acids Research. 2012;40:3538-47.
16. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR,

Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. Nucleic Acids Research. 2006;34:D590-8.
17. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & Development. 2011;24:1915-27.
18. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57-74.
19. Kannan K, Wang L, Wang J, Ittmann MM, Li W, Yen L. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. PNAS. 2011;108:9172-9177.
20. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011;474:390-4.
21. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR. WikiPathways: building research communities on biological pathways. Nucleic Acids Research. 2012;40:D1301-7.
22. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Research. 2011;39:D685-690.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics. 2000;25:25-9.
24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000;28:27-30.
25. Dash A, Maine IP, Varambally S, Shen R, Chinnaiyan AM, Rubin MA. Changes in differential gene expression because of warm ischemia time of radical prostatectomy specimens. American Journal of Pathology. 2002;161:1743-1748.

PUBLICATION II

AR and ERG drive the expression of prostate cancer specific long noncoding RNAs

Annika Kohvakka, Mina Sattari, Anastasia Shcherban, Matti Annala, Alfonso Urbanucci, Juha Kesseli, Teuvo L. J. Tammela, Kati Kivinummi, Leena Latonen, Matti Nykter, and Tapio Visakorpi

Oncogene, 2020, 39(30):5241–5251
doi: 10.1038/s41388-020-1365-6

Oncogene
https://doi.org/10.1038/s41388-020-1365-6

ARTICLE

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] · Tapio Visakorpi[1,5]

Received: 6 March 2020 / Revised: 25 May 2020 / Accepted: 8 June 2020

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

Supplementary information The online version of this article (https://doi.org/10.1038/s41388-020-1365-6) contains supplementary material, which is available to authorized users.

✉ Tapio Visakorpi
tapio.visakorpi@tuni.fi

[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

Published online: 17 June 2020

overexpression of ERG. High ERG expression has been suggested to promote invasion and progression of PC cells [10, 11].

Long noncoding RNAs (lncRNA) are over 200 nucleotide long nonprotein-coding transcripts that are involved in various biological and pathological processes, including cancer [12]. In PC, 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.

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 (Fig. 1).

To further assess the observed ERG association, we divided the PC samples into ERG-positive and ERG-negative groups based on their ERG gene fusion status and expression [21] (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 Fig. 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 Fig. 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 ten 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 with 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) (Fig. 2; Supplementary Table S5). In addition, the vast majority of all the TPCAT-associated ERG

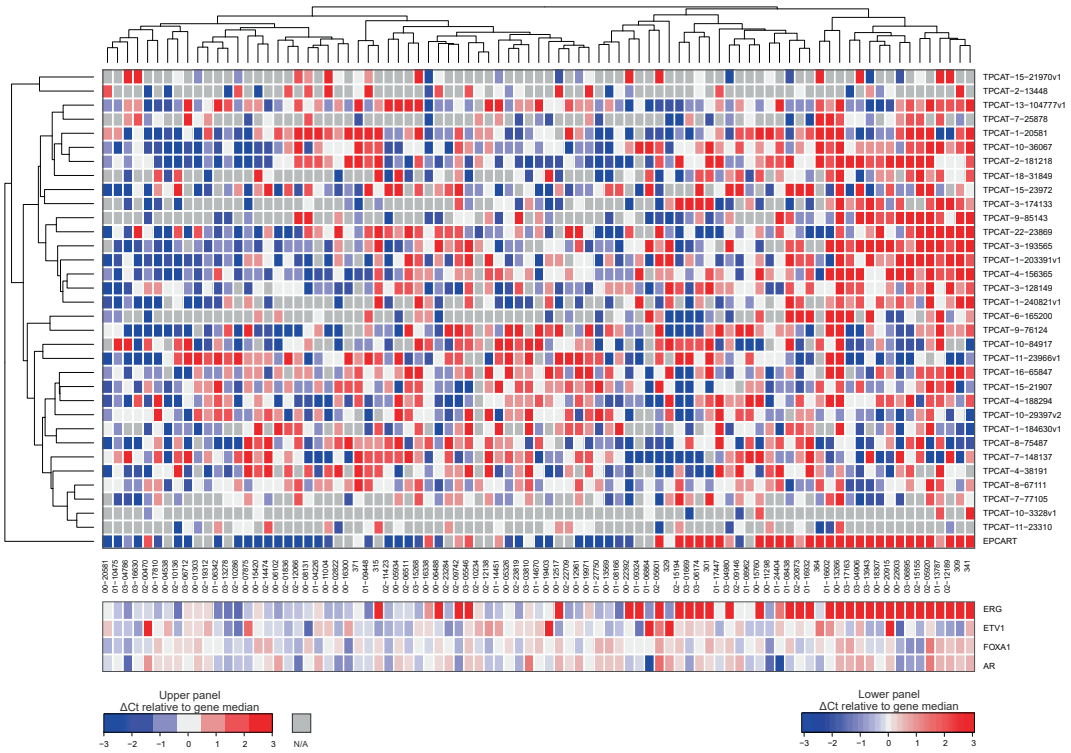


Fig. 1 ERG overexpression correlates with the expression of TPCATs. The expression of 34 TPCATs was analyzed in 87 prostatectomy specimens by qRT-PCR using Fluidigm Biomark HD. Hierarchical clustering revealed multiple TPCATs that were abundantly expressed in samples overexpressing ERG.

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 Fig. 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 (Fig. 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 [23–26], 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 (Fig. 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 six times more ARBS 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 Fig. 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, seven TPCATs were affected in opposite ways by both treatments in the same cell line; however, a similar but

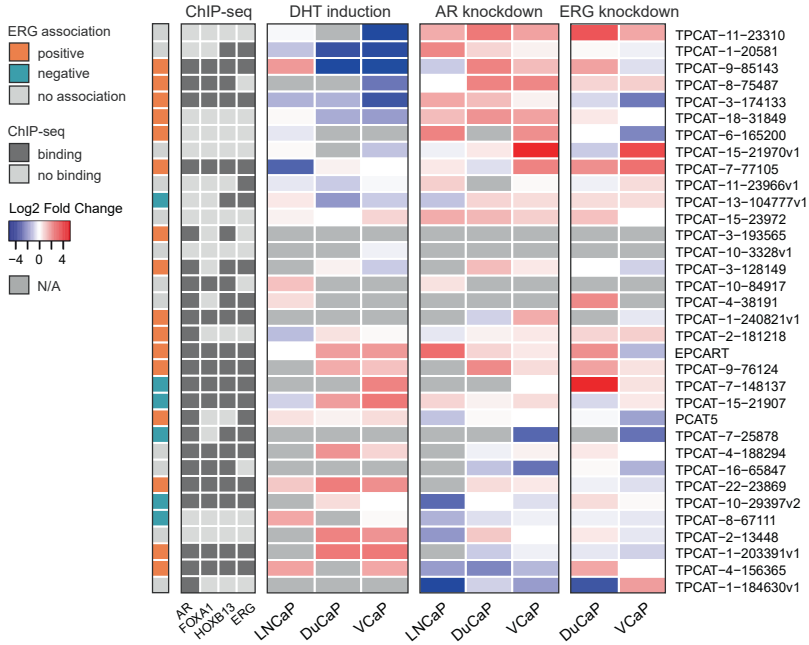


Fig. 2 Several TPCATs are regulated by AR and ERG. The ERG association of TPCATs based on the expression of TPCATs in clinical samples (Supplementary Fig. S1a, b) is marked in the column on the left. ChIP-seq peaks for different TFs (AR, ERG, FOXA1, and HOXB13) found in the regulatory region (−15 kb/+2 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.

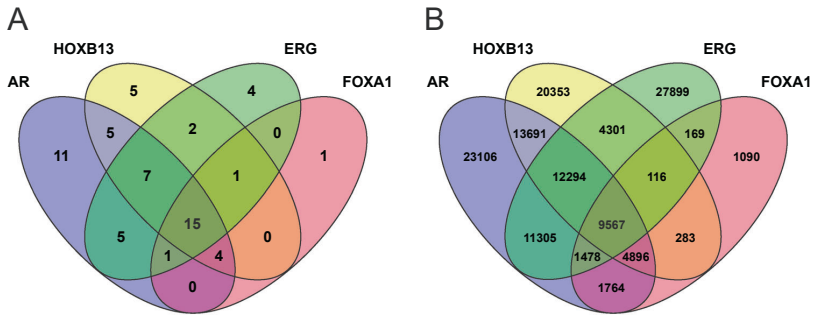


Fig. 3 TFs that drive PC colocalize in the regulatory regions of TPCATs. a Number of peaks detected in ChIP-seq data for AR, ERG, FOXA1, and HOXB13 in the regulatory region (−15 kb/+2 kb from TSS) of TPCATs. **b** Total number of AR, ERG, FOXA1, and HOXB13 ChIP-seq peaks detected in the genome.

weaker effect was also noticeable with several additional TPCATs (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 2; Supplementary Table S5). HOXB13 binding (39 peaks) was observed more frequently than FOXA1 binding (22 peaks) (Fig. 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 (Fig. 3a) was slightly, but not significantly, higher than what was globally detected in PC (62%) (Fig. 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) (Fig. 3a, b). One-third of the TPCATs (13 out of 34) had at least one binding site from one of the four TFs (Fig. 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

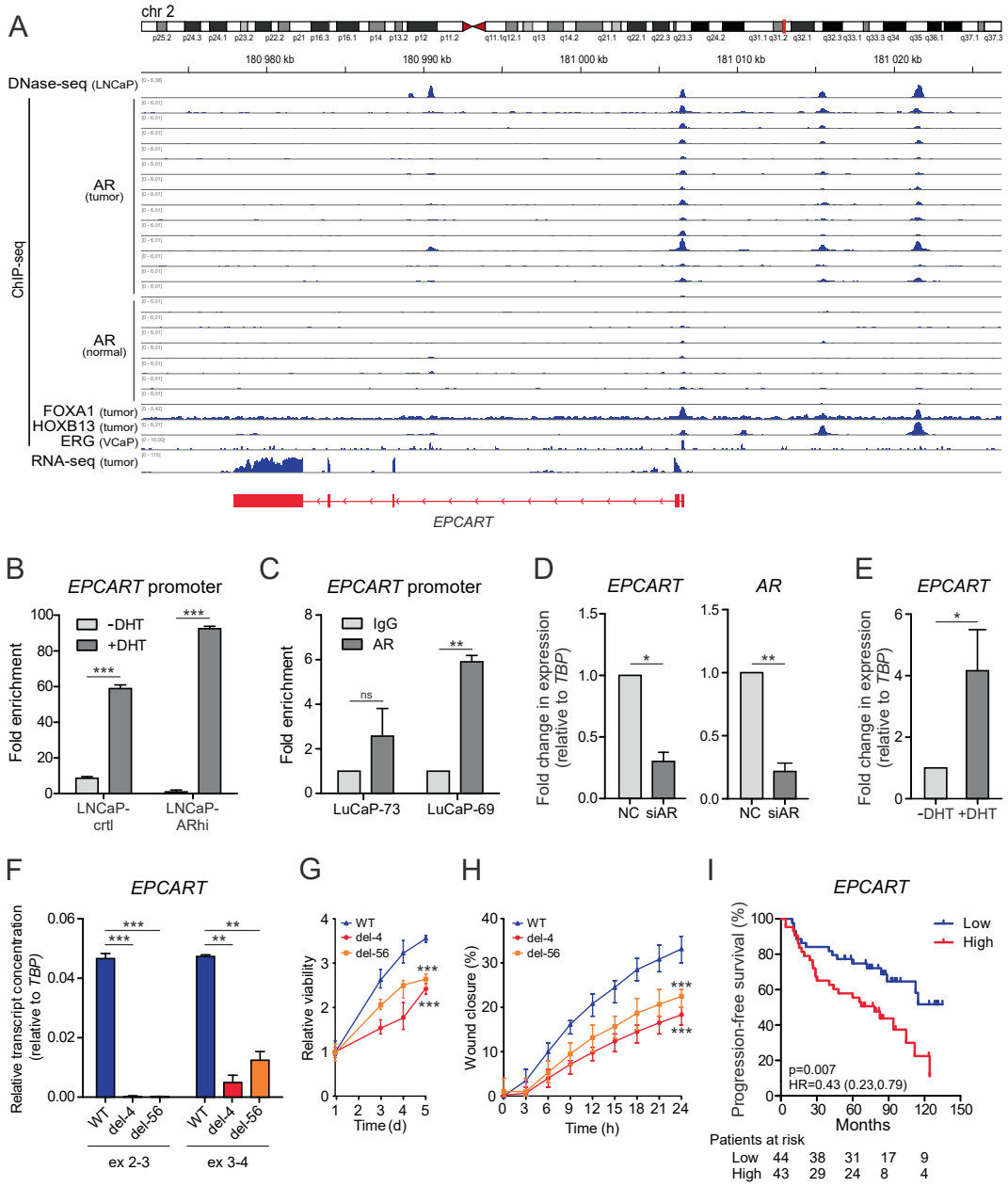
From our experiments, it became evident that *TPCAT-2-180961*, officially termed ERG-positive PC-associated androgen responsive transcript (EPCART), was highly expressed in PCs overexpressing ERG (Fig. 1**;** Supplementary Fig. S1a, b), and data suggested that it was regulated by both AR and ERG (Fig. 2). According to our previously generated RNA-seq data, *EPCART* is located in chromosome 2 and has five exons (Fig. 4a). Publicly available DNase-seq data in LNCaP cells [27] showed chromatin to be open where there were three ARBS located in the regulatory region of *EPCART* (Fig. 4a). These ARBS were also highly PC-associated and were co-occupied by FOXA1 and/or HOXB13 (Fig. 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 with that of the parental LNCaP cells (Fig. 4b). In addition, LuCaP69 xenograft containing AR gene amplification [28] showed more AR binding to *EPCART* compared with what was observed in the LuCaP73 xenograft without amplification (Fig. 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 droplet digital PCR (ddPCR). In these experiments, the expression of *EPCART* was significantly downregulated after AR knockdown (Fig. 4d), while DHT induced the expression of *EPCART* (Fig. 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* from LNCaP cells (*EPCART*-del) using CRISPR/Cas9. Two sgRNAs were designed to target the area covering the promoter, the 1st exon, and the 2nd exon of *EPCART* (Supplementary Fig. 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 Fig. 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 Fig. S4a). We detected a considerable reduction, although not a full abolition, of the *EPCART* transcript in both *EPCART*-del clones when compare with the WT clone (Fig. 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 (Fig. 4g) and migration (Fig. 4h, Supplementary Fig. S4c) were significantly reduced in both *EPCART*-del clones as compared with 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 (Fig. 4i). Furthermore, multivariate Cox regression analysis showed that the expression of *EPCART* had independent prognostic value (other parameters included were age, Gleason score, diagnostic PSA, and pathological T stage (pT)) (Table 1). Prompted by this, we further investigated whether the expression of other TPCATs was associated with PC progression. We found that *TPCAT-3-174133* and *TPCAT-18-31849* were also associated with a short biochemical progression-free survival in PC patients (Supplementary Fig. S5). Both of these lncRNAs also had 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, 29–31].



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.

◀ **Fig. 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 [28]. 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 were 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 with 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 were 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 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 were 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 were assessed with an unpaired two-tailed *t*-test. **g–h** Proliferation (**g**) and migration (**h**) was decreased in *EPCART*-del cells when compared with WT LNCaP cells. Cell viability was measured by alamarBlue over 5 days, and wound healing was analyzed by Cell-IQ time-lapse imaging over 24 h. Error bars, range; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; data were 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.

TMPRSS2-ERG gene fusion has previously been associated with early-onset PC and high-risk tumors as a result of ERG overexpression [9, 32–34], 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 [35].

Previous studies have shown several lncRNAs to be associated with AR signaling in PC [23–26], 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

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.

development of clinical metastasis and PC-related death [35]. 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. In addition, 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.

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 PC 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 [36]. 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 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 (−15 kb/+2 kb 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 [37] (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 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. [38]. 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 h before RNA extraction and 72 h 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 4 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 PC cell line, the area covering the promoter and the 1st and 2nd 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 <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 ≤10), intermediate (PSA from 10 to 19.9), and high (PSA >20). Cox regression analysis was performed using coxph function from the survival package version 2.41-3 in R.

Acknowledgements This study was supported by grants from the Academy of Finland (TV 317755, MN 310829, and LL 317871), Sigrid Juselius Foundation (TV and 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.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

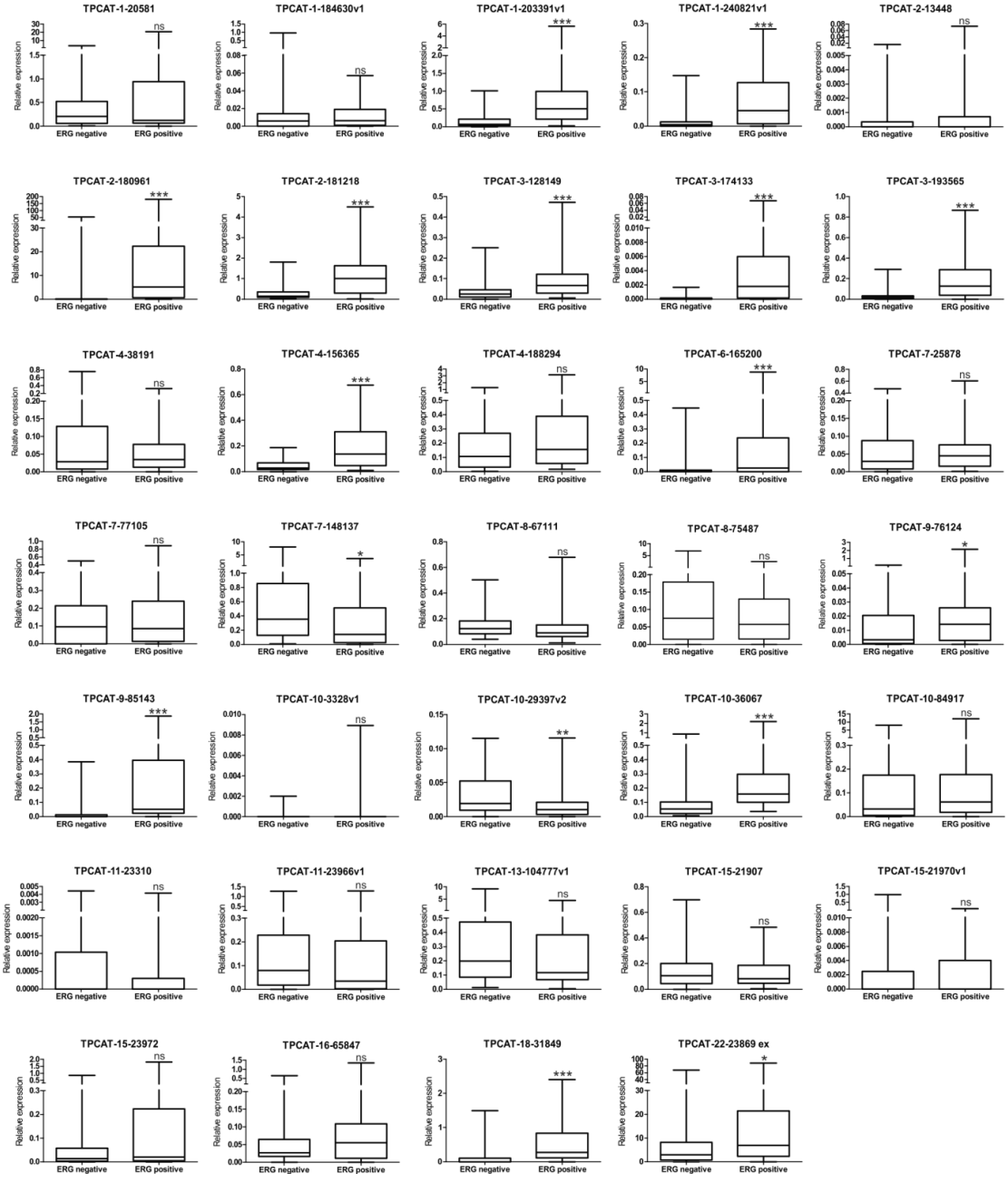
References

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;17:443–54.
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;19:1023–9.
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. Nat 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. Nat Genet. 2017;49:1336–45.
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;310:644–8.
7. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. Cell. 2015;163:1011–25.
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;8:826–32.
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;31:882–8.
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;10:177–88.
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;41:619–24.
12. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. Cancer Cell. 2016;29:452–63.
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;65:1140–51.
14. Sahu A, Singhal U, Chinnaiyan AM. Long noncoding RNAs in cancer: from function to translation. Trends Cancer. 2015;1:93–109.
15. Hessels D, Klein Gunnewiek JMT, 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;15:1469–80.
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;18:489–99.
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;71:257–66.
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;29:742–9.
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;75:4026–31.
21. 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;60:183–4.
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;29:2519–20.
23. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-dependent mechanisms of androgen-receptor- regulated gene activation programs. Nature. 2013;500:598.
24. 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;32:1665–80.
25. 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;13:209–21.
26. 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;50:814–24.
27. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
28. 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;61:3550–5.
29. 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.
30. 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;5:9864–76.
31. 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;28:529–40.
32. 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;66:8337–41.
33. 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 Biol Ther. 2007;6:40–45.
34. 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.
35. 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;6:4036–50.
36. 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;69:8141–9.
37. 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:D794–801.
38. 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;31:2153–63.

Supplementary Figures

A



B

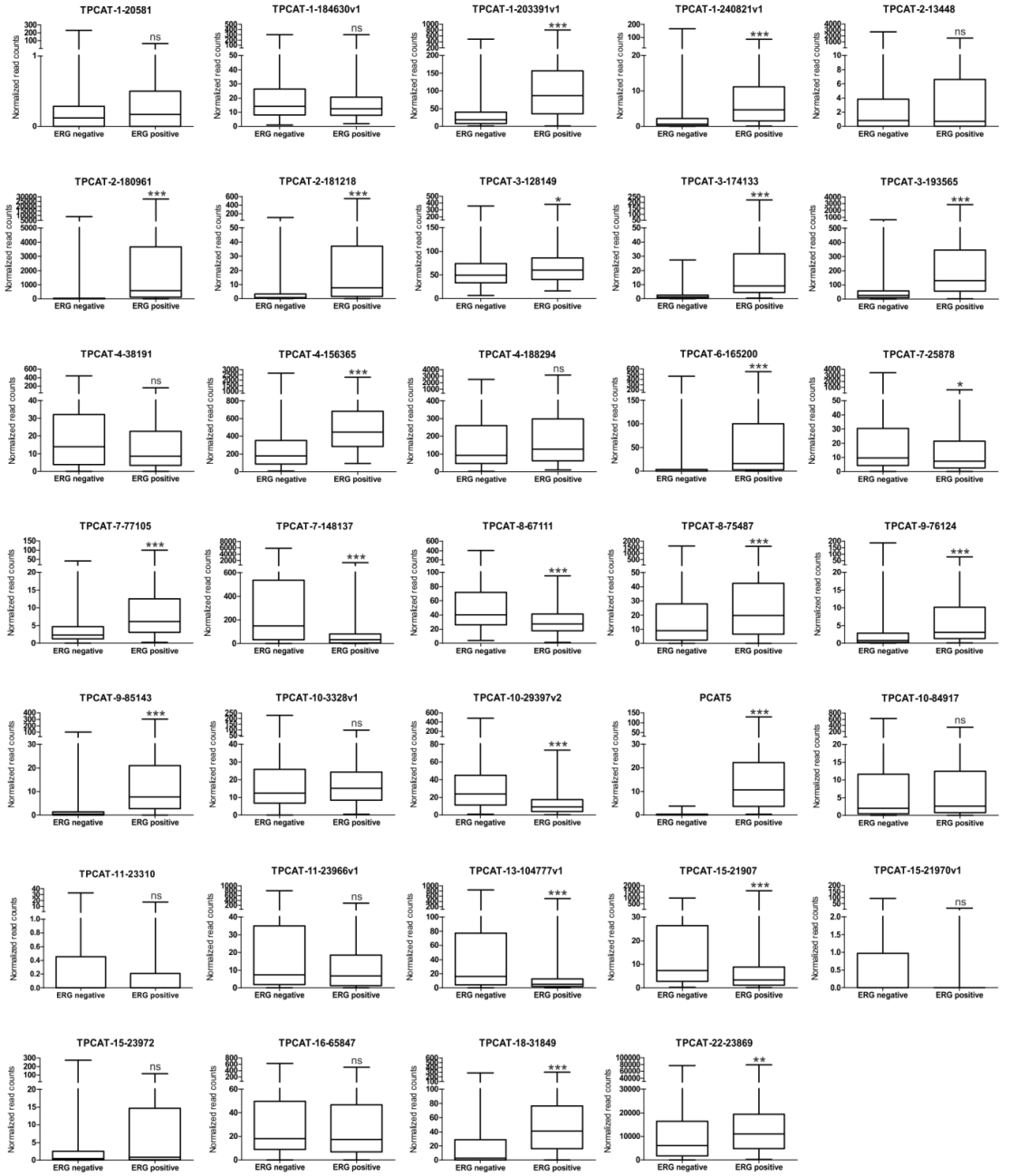


Figure S1. Expression of TPCATs in ERG-positive and ERG-negative samples in **a**, Tampere prostatectomy cohort (50 ERG negative and 34 ERG positive) and **b**, TCGA-PRAD data set (145 ERG negative and 121 ERG positive). P values were determined by using two-tailed Mann Whitney test; *, p<0.05; **, p<0.01; ***, p<0.001. The bar indicates the median expression, and whiskers minimum and maximum values.

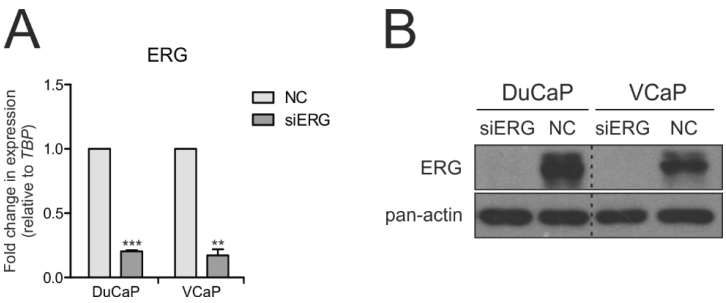


Figure S2. Expression of ERG after ERG siRNA knockdown. **a**, mRNA levels of *ERG* after siRNA transfection (25nM, 48h) were measured in three biological replicates by qRT-PCR and normalized against *TBP*. The expression was. Error bars, SD; *, p<0.05; **, p<0.01; ***, p<0.001; unpaired two-tailed t test. **b**, Protein levels of ERG were analyzed by Western blot after siRNA transfection (25nM, 72h). Pan actin was used as a loading control. Dashed line indicates a break in the blot. siERG, ERG siRNA; NC, negative siRNA control.

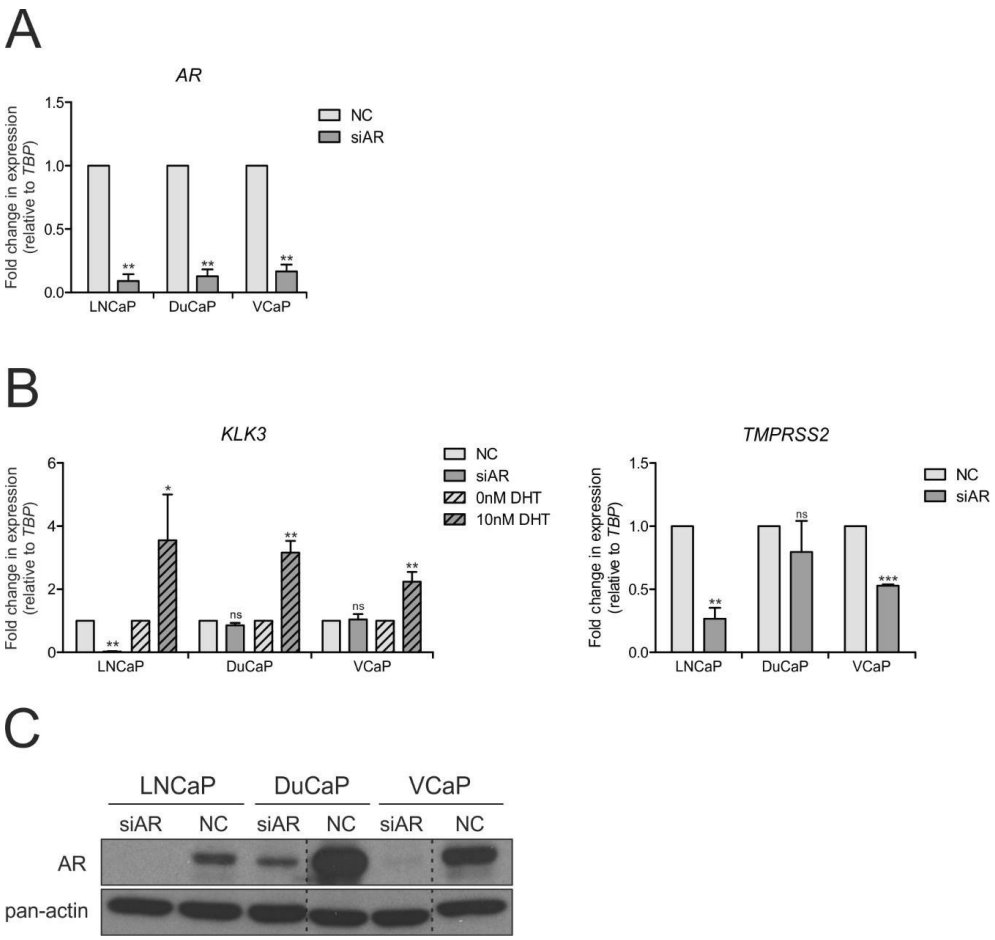


Figure S3. Expression of AR and its target genes. **a**, mRNA levels of *AR* were analyzed after AR siRNA knockdown (25nM, 48h). The expression was measured in three biological replicates by qRT-PCR and normalized against *TBP*. Error bars, SD; *, p<0.05; **, p<0.01; ***, p<0.001; unpaired two-tailed t test. **b**, mRNA levels of an AR-target gene *KLK3* were analyzed after AR siRNA knockdown (25nM, 48h) and DHT induction (10nM, 24h). In addition, mRNA levels of another AR-target gene, *TMPRSS2* were analyzed in the siAR knockdown samples. The expression was measured in three biological replicates by qRT-PCR and normalized against *TBP*. Error bars, SD; *, p<0.05; **, p<0.01; ***, p<0.001; unpaired two-tailed t test. **c**, Protein levels of AR were analyzed by Western blot after siRNA transfection (25nM, 72h). Pan-actin was used as a loading control. Dashed line indicates a break in the blot. siAR, AR siRNA; NC, negative siRNA control.

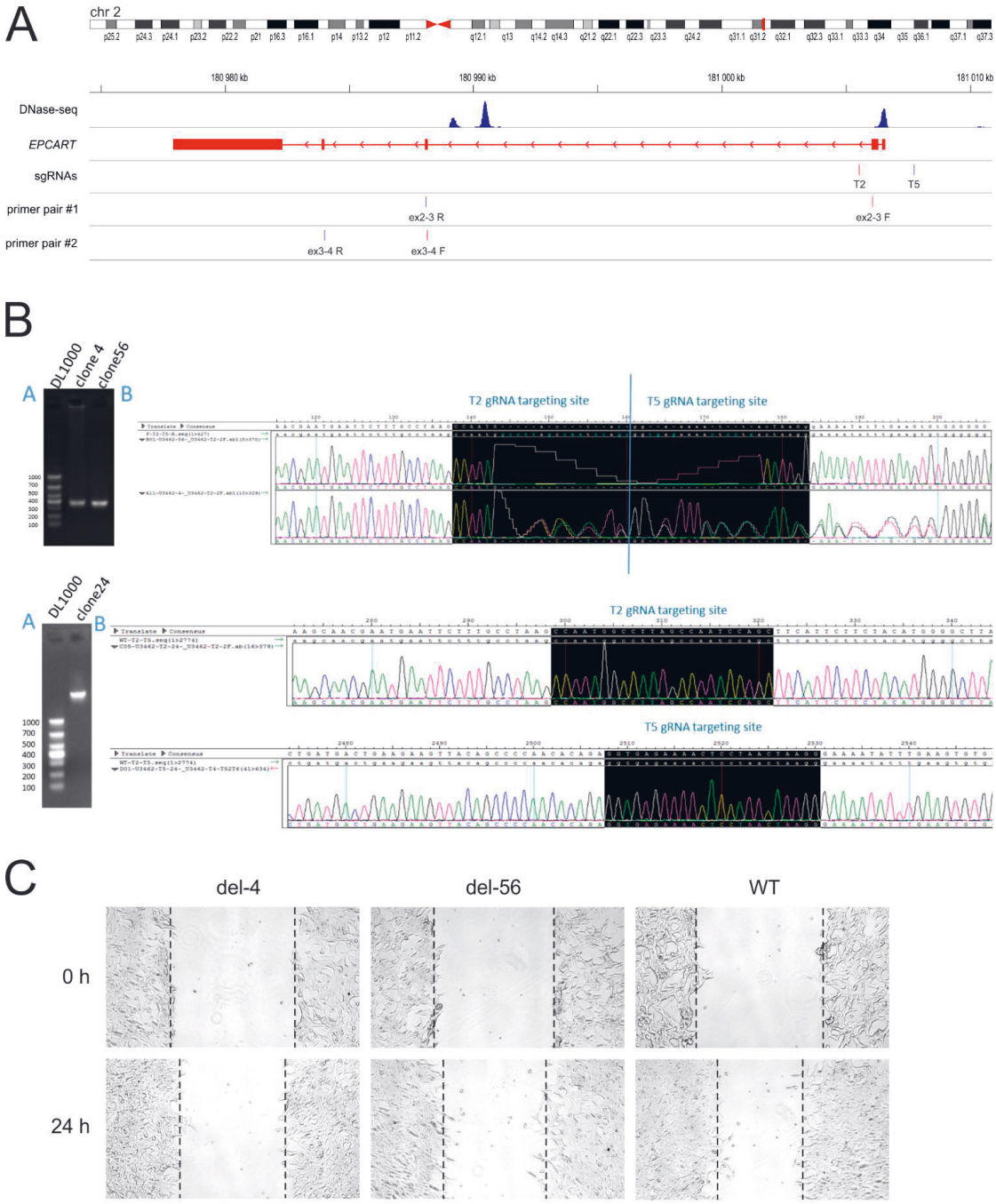


Figure S4. *EPCART* knockout by CRISPR/Cas9. **a**, Knockout of *EPCART* was accomplished by two sgRNAs (T2 and T5) targeting ~1.7 kb region that covers the promoter and the first two exons of *EPCART* and validated by two PCR primer pairs overlapping exon junctions. **b**, Deletion of the region was confirmed by PCR (A) and Sanger sequencing (B) in two LNCaP clones (clone 4 and clone 56), but not in one (clone 24) that still had the WT sequence. **c**, Wound healing assay for the *EPCART* deletion clones (del-4 and del-56) and the WT clone revealed increased migration of the deletion clones compared to the WT after 24h. Assay was performed by Cell-IQ automated imaging system for 500 000 cells on a well of a 24 well plate grown for 2 days before the experiment.

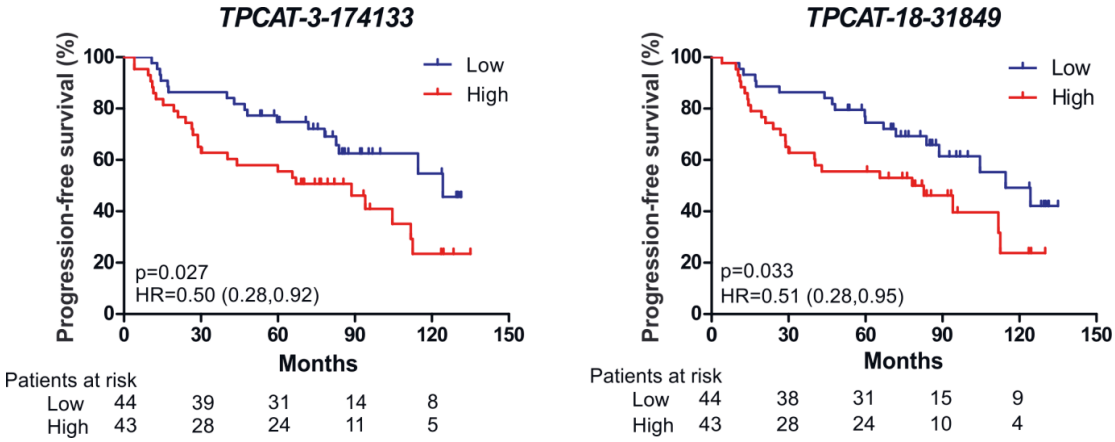


Figure S5. Kaplan-Meier analysis for progression-free survival of PC patients divided by median expression of TPCATs. P-values were calculated by log-rank test. HR = hazard ratio.

PUBLICATION III

Long noncoding RNA *EPCART* regulates translation through PI3K/AKT/mTOR pathway and PDCD4 in prostate cancer.

Annika Kohvakka, Mina Sattari, Janika Nättinen, Ulla Aapola, Pavlína Gregorová, Teuvo L.J. Tammela, Hannu Uusitalo, Peter Sarin, Tapio Visakorpi*, and Leena Latonen*

Manuscript

*These authors contributed equally to the work